Challenges and Advances
in Computational Chemistry and Physics 30
Series Editor: Jerzy Leszczynski

Huixiao Hong *Editor*

Advances in Computational Toxicology

Methodologies and Applications in Regulatory Science

Springer

Challenges and Advances in Computational Chemistry and Physics

Volume 30

Series Editor

Jerzy Leszczynski
Department of Chemistry and Biochemistry
Jackson State University, Jackson, MS, USA

This book series provides reviews on the most recent developments in computational chemistry and physics. It covers both the method developments and their applications. Each volume consists of chapters devoted to the one research area. The series highlights the most notable advances in applications of the computational methods. The volumes include nanotechnology, material sciences, molecular biology, structures and bonding in molecular complexes, and atmospheric chemistry. The authors are recruited from among the most prominent researchers in their research areas. As computational chemistry and physics is one of the most rapidly advancing scientific areas such timely overviews are desired by chemists, physicists, molecular biologists and material scientists. The books are intended for graduate students and researchers.

All contributions to edited volumes should undergo standard peer review to ensure high scientific quality, while monographs should be reviewed by at least two experts in the field. Submitted manuscripts will be reviewed and decided by the series editor, Prof. Jerzy Leszczynski.

More information about this series at http://www.springer.com/series/6918

Huixiao Hong
Editor

Advances in Computational Toxicology

Methodologies and Applications in Regulatory Science

Springer

Editor
Huixiao Hong
National Center for Toxicological Research
US Food and Drug Administration
Jefferson, AR, USA

ISSN 2542-4491 ISSN 2542-4483 (electronic)
Challenges and Advances in Computational Chemistry and Physics
ISBN 978-3-030-16442-3 ISBN 978-3-030-16443-0 (eBook)
https://doi.org/10.1007/978-3-030-16443-0

Library of Congress Control Number: 2019935555

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Given the ever-increasing panoply of human and animal drugs, food products, and environmental chemicals, the need for science-based risk/safety regulation is greater than ever. From a pharmaceutical perspective, accurate and effective toxicity evaluation is critical in several areas such as dose–response characteristics including organ exposure and first-in-human dosages, reproductive and carcinogenicity toxicity, exposure assessment, and biological pathway characterization. Food product assessment requires understanding of gastrointestinal delivery, metabolic breakdown into metabolites, context of use, and dietary exposures. Lastly, environmental toxicity necessitates system-level approaches considering chemical mixtures and chemical transport into target organs in multiple species. Despite these extensive efforts, idiosyncratic toxicities can occur, suggesting the need for personalized toxicity approaches.

Conventional approaches along with some new methodologies like "organ-or-a-chip" have been developed to address key questions in this area. Many of these approaches are limited in cost, time, translational accuracy, and scalability. Consequently, scientific endeavors in the computational space have inspired new and powerful tools, ushering in the era of computational toxicology. This exciting field facilitates the paradigm shift from bench-based toxicology to the computational assessment and will provide regulators globally with the benefit of fast, accurate, and low-cost methods to supplement conventional toxicity assessment. Moreover, integrative predictive approaches may enhance personalized toxicological prediction to prevent idiosyncratic events.

To inform not only regulators around the world but also key stakeholders, industry, and academic trainees, this textbook provides a deep dive into computational toxicological approaches needed to advance toxicological regulation through research. It includes sections outlining theory, methods, applications, as well as tangible examples and covers development through implementation. Information in this book will apprise the reader with a greater understanding of computer-based toxicological predictive capabilities. Information in this book will also enable the reader to develop their own cutting-edge computational strategy to address a toxicological question of interest. The provided information may also

foster collaboration by providing inspiration for scientific discourse among readers with diverse training backgrounds.

Currently, computational toxicology has gained acceptance as an "alternative" testing method compared to traditional approaches for rapid toxicity assessment. The toxicology community of scientists and regulators look forward to the validation of computational methods that may supplement and, in some cases, replace traditional assays. The contents of this textbook, inspired by new computational methods and approaches, provides a comprehensive overview of the representative methodologies in the land of computational toxicology with an emphasis on regulatory science research.

Jefferson, AR, USA

William Slikker Jr. Ph.D.
Director, National Center for Toxicological Research
US Food and Drug Administration

Preface

In comparison with the field of toxicology which dates back to ancient civilizations, the field of computer science is just coming of age. After many years of focused development, computational tools, methods, hardware, and knowledge have advanced enough to be of utility in other fields. Computational techniques are now employed in toxicology for regulatory, research, and development purposes. We compiled this book with the ambition to capture the latest advancements at the intersection of computer science and toxicology together in one location.

Our motivation in creating this book was twofold. First of all, the number of new chemical entities being developed by the ever-growing pharmaceutical, biotechnology, and food industry is dizzying. Absolutely all of these products must be screened for safety at various stages of development. Secondly, the ecosystem of our planet is increasingly bathed in an assortment of molecules, many of which nature has never experienced before. The surge of new molecules and entities entering the human body and the environment presents an insurmountable challenge to traditional toxicology. Impressively, the novel computational toxicology methods described herein are rising to meet the challenge.

Machine learning, artificial intelligence, quantitative structure–activity relationship (QSAR), bioinformatics, genomics, proteomics, molecular dynamics, and more are described via examples of applications to toxicology. Both safety evaluation and risk assessment are topics of consideration across multiple chapters. A background introduction followed by details is provided for computational toxicology methods, as well as applications. Toxicology from the perspective of medicines, food products, and the environment is described in multiple chapters.

This book is intended as a text for established computer scientists looking to enter the toxicology field, experienced toxicologists seeking to enable research through computational methods, or students and trainees curious about stepping into the field. Thus, this book includes not only introductory sections to help readers become familiar with new concepts but also detailed actionable methods which can be deployed by the reader. Each chapter of this book can stand alone to update the reader on a specific topic of interest. Alternatively, this textbook can be read in sections as chapters that are roughly organized in topical order. For a graduate

course on computational toxicology, this book would provide excellent reading material. Studying the entire book will provide the reader with not only a broad but also a deep understanding of the field.

While this book does focus on computational toxicology, it does not contain any learning exercises, quizzes, or snippets of computer code for the reader to advance and test knowledge. In the case of the classroom setting, these materials would be left to the course instructor. Moreover, herein we do not review basic toxicology concepts as they are effectively covered by other established bodies of work. We would be appreciative of any corrections, feedback, comments, or criticism from readers on how to improve for a future body of work.

Inspired by how computational toxicology is rising to meet the challenges currently facing traditional toxicology methods, we put forth this book for the community as an educational tool. The broad scope and deep depth of this textbook would not be possible without the herculean efforts of and tremendous cooperation from the authors, for which we are tremendously appreciative. This book also would not have been possible without the support and vision of Springer, who we acknowledge for having a visionary understanding of the importance of the topic at hand.

This preface reflects the views of the authors and should not be construed to represent the FDA's views or policies.

Cambridge, MA, USA

Rebecca Kusko, Ph.D.
Senior Vice President and Head of R&D
Immuneering Corporation

Jefferson, AR, USA

Huixiao Hong, Ph.D.
Chief, Bioinformatics Branch
Division of Bioinformatics and Biostatistics
National Center for Toxicological Research
US Food and Drug Administration

Contents

Editor and Contributors

About the Editor

Huixiao Hong is Chief of Bioinformatics Branch, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research (NCTR), US Food and Drug Administration (FDA), working on the scientific bases for regulatory applications of bioinformtics. Before joining the FDA, he was Manager of Bioinformatics Division of Z-Tech, an ICFI company. He was Research Scientist at Sumitomo Chemical Company in Japan and Visiting Scientist at National Cancer Institute at National Institutes of Health. He was also Associate Professor and Director of Laboratory of Computational Chemistry at Nanjing University in China. He is Member of steering committee of OpenTox, Member of board directors of US MidSouth Computational Biology and Bioinformatics Society, and in the leadership circle of US FDA modeling and simulation working group. He published more than 180 scientific papers and served as Editor-in-Chief and Editorial Board member for multiple peer-reviewed journals. He received his Ph.D. from Nanjing University in China in 1990 and conducted research in University of Leeds in England in 1990–1992. e-mail: Huixiao.Hong@fda.hhs.gov

Contributors

Tero Aittokallio Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland;
Department of Mathematics and Statistics, University of Turku, Turku, Finland

Kristin Ashby Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Scott S. Auerbach Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Durham, NC, USA

Daniel Bachler Edelweiss Connect GmbH, Technology Park Basel, Basel, Switzerland

Chris Barber Lhasa Limited, Leeds, UK

Emilio Benfenati Istituto di Ricerche Farmacologiche Mario Negri – IRCCS, Milano, Italy

Richard Bergström Bergstroem Consulting GmbH, Zug, Switzerland

Jürgen Borlak Hannover Medical School, Center of Pharmacology and Toxicology, Hannover, Germany

Daniel Burgwinkel Guardtime AS, Basel, Switzerland

Alex Cayley Lhasa Limited, Leeds, UK

Jingwen Chen Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology, Dalian University of Technology, Dalian, China

Minjun Chen Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Feixiong Cheng Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA;
Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH, USA;
Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA

Martyn Chilton Lhasa Limited, Leeds, UK

Brian Delavan Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA;
Department of Biosciences, University of Little Rock, Little Rock, AR, USA

Joh Dokler Edelweiss Connect GmbH, Technology Park Basel, Basel, Switzerland

Thomas Exner Edelweiss Connect GmbH, Technology Park Basel, Basel, Switzerland

Lilia Fisk Lhasa Limited, Leeds, UK

Zhiqiang Fu Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology, Dalian University of Technology, Dalian, China

Ping Gong Environmental Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS, USA

Roland Grafström Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden;
Predictomics AB, Stockholm, Sweden;
Misvik Biology Oy, Turku, Finland

Sébastien Guesné Lhasa Limited, Leeds, UK

Thierry Hanser Lhasa Limited, Leeds, UK

Barry Hardy Edelweiss Connect GmbH, Technology Park Basel, Basel, Switzerland

Connor Hardy Edelweiss Connect GmbH, Technology Park Basel, Basel, Switzerland

Huixiao Hong Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Ruili Huang National Center for Advancing Translational Sciences, NIH, Rockville, MD, USA

Gabriel Idakwo School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS, USA

Shruti V. Kabadi FDA/CFSAN/OFAS, College Park, MD, USA

Suleiman A. Khan Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

Pekka Kohonen Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden;
Predictomics AB, Stockholm, Sweden;
Misvik Biology Oy, Turku, Finland

Rebecca Kusko Immuneering Corporation, Cambridge, MA, USA

Yan Li Bennett Aerospace, Inc., Cary, NC, USA

Huihui Liu Jiangsu Key Laboratory of Chemical Pollution Control and Resources Reuse, School of Environmental and Biological Engineering, Nanjing University of Science and Technology, Nanjing, China

Zhichao Liu Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Anna Lombardo Istituto di Ricerche Farmacologiche Mario Negri – IRCCS, Milano, Italy

Joseph Luttrell IV School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS, USA

Donna Macmillan Lhasa Limited, Leeds, UK

Alberto Manganaro KODE, Pisa, Italy

Jean François Marchaland Lhasa Limited, Leeds, UK

Mukesh Patel Lhasa Limited, Leeds, UK

Ruth Robert ApconiX, BioHub at Alderley Park, Alderley Edge, UK; Department of Biosciences, University of Birmingham, Birmingham, UK

Alessandra Roncaglioni Istituto di Ricerche Farmacologiche Mario Negri – IRCCS, Milano, Italy

Trey Saddler Kelly Government Solutions, Durham, NC, USA

Sugunadevi Sakkiah National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA

Andreas Scherer Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

Daniel L. Svoboda Sciome LLC, Durham, NC, USA

Sundar Thangapandian Environmental Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS, USA

Weida Tong Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Patra Volarath FDA/CFSAN/OFAS, College Park, MD, USA

Dong Wang Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Zhongyu Wang Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology, Dalian University of Technology, Dalian, China

Stéphane Werner Lhasa Limited, Leeds, UK

Richard Williams Lhasa Limited, Leeds, UK

Leihong Wu Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Xianhai Yang Jiangsu Key Laboratory of Chemical Pollution Control and Resources Reuse, School of Environmental and Biological Engineering, Nanjing University of Science and Technology, Nanjing, China

Yu (Janet) Zang FDA/CFSAN/OFAS, College Park, MD, USA

Chaoyang Zhang School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS, USA

Chaoyang (Joe) Zhang School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS, USA

Jieqiang Zhu Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Liyuan Zhu Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA

Chapter 1
Computational Toxicology Promotes Regulatory Science

Rebecca Kusko and Huixiao Hong

Abstract New tools have become available to researchers and regulators including genomics, transcriptomics, proteomics, machine learning, artificial intelligence, molecular dynamics, bioinformatics, systems biology, and other advanced techniques. These new advanced approaches originated elsewhere but over time have perfused into the toxicology field, enabling more efficient risk assessment and safety evaluation. While traditional toxicological methods remain in full swing, the continuing increase in the number of chemicals introduced into the environment requires new toxicological methods for regulatory science that can overcome the shortcoming of traditional toxicological methods. Computational toxicology is a new toxicological method which is much faster and cheaper than traditional methods. A variety of methods have been developed in computational toxicology and some have been adopted in regulatory science. This book summarizes some methods in computational toxicology and reviews multiple applications in regulatory science, indicating that computational toxicology promotes regulatory science.

Keywords Computational toxicology · Regulatory science · Risk assessment · Safety evaluation · Chemicals

Abbreviations

3D	Three dimensional
AI	Artificial intelligence

R. Kusko
Immuneering Corporation, Cambridge, MA, USA
e-mail: bkusko@immuneering.com

H. Hong (✉)
National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA
e-mail: Huixiao.Hong@fda.hhs.gov

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_1

CYP	Cytochrome P450 enzyme
DA	Department of Agriculture
DILI	Drug-induced liver injury
EDC	Endocrine disrupting chemical
EDSP	Endocrine disruptor screening program
EPA	Environment Protection Agency
FDA	Food and Drug Administration
ML	Machine learning
MD	Molecular dynamics
MDDT	Medical Device Development Tools
MOA	Mechanism of Action
MoA	Mode of Action
NCATS	National Center for Advancing Translational Sciences
NIEHS	National Institute of Environmental Health Sciences
NN	Neural Networks
POD	Point of Departure
QSAR	Quantitative structure–activity relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
TsTKB	Target-specific Toxicity Knowledgebase
US	United States

1.1 Computational Toxicology

Toxicology as a broad field seeks to predict and eliminate substances which may cause a living body harm, including pharmaceuticals, natural products, food products, and environmental substances. Toxicology has been performed since the ancient Greeks and Chinese [1]. It is currently a major field of study around the world. The study of toxicology is of importance not only to governmental regulatory agencies, but also to the pharmaceutical/biotech industry, the veterinary industry, food manufacturers, and academics. Toxicology also spans many sub-disciplines as it must consider the entire path of a potential toxicant, including exposure, absorption, distribution, metabolism, excretion, as well as interactions with cellular machinery throughout this entire pathway (Fig. 1.1). Pinpointing the exact mechanism or mode of toxicity as a potential toxicant interacts with a living organism is paramount. Adding to an already complex system, nearly any known substance can be toxic at a high enough exposure. Moreover, toxicity is dependent on an array of other factors including organism size, species, age, sex, genetics, diet, combination with other chemicals, overall health, and/or environmental context.

Toxicological methods can be classed into experimental and computational [2]. Experimental methods consist of two types: in vivo and in vitro experiments indicated by the blue arrows in Fig. 1.1. Traditional experiments in toxicology are conducted on non-human animals such and mice and rats [3]. Though in vivo experiments are generally treated as the gold standard method in toxicological studies and remain as

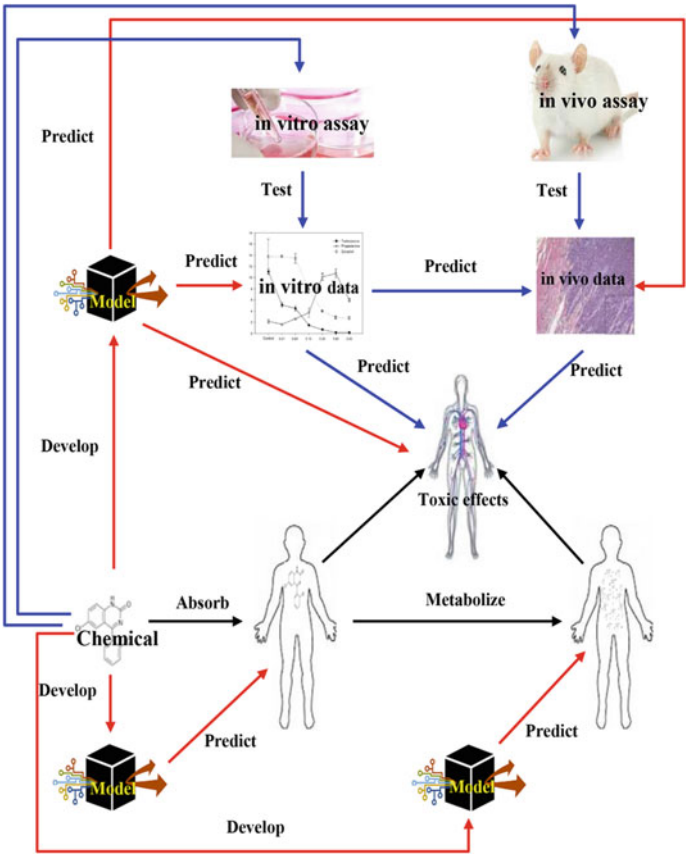

Fig. 1.1 Predictive toxicology schematic. A potential route of a toxicant is represented as black arrows. In vivo and in vitro toxicology as well as connections between components are represented as blue arrows. Computational toxicology's key role is represented as red arrows. Arrows are labeled by actions that achieve the arrow targets

a major approach of estimating human effects of substances, ethical, economic, and technical concerns on use of animals make toxicologists search for alternative testing methods [4–6]. Computational toxicology, an emerging component in toxicology, is an alternative method where computational methods are used to understand and predict toxicological effects of substances in the environment such as drugs, food, and environmental chemicals [7–9]. Many computational methods have been developed for predicting absorption, metabolism of chemicals, estimating in vitro and in vivo experimental data, and assessing human risk solely based on chemical structures as illustrated by the red arrows in Fig. 1.1 [10–17].

Given impressive advances recently in computing power as well as developments in advanced computational algorithms, many toxicologists are now reaching out to computer science to enable prediction of toxicological effects or outcomes. Thus,

the intersection of computer science and toxicology is what we here call "Computational Toxicology." Computational toxicology integrates both the long-standing computational methods and the newer approaches including neural networks (NNs) and artificial intelligence (AI). Rather than individual scientists and researchers trying to understand multiple complex phenomena via bench experiments, these complex biological systems can now be modeled and predicted in the computational space. Issues which may have previously seemed impossible or intractable are increasingly becoming solvable due to the scalability of computational toxicology.

1.2 Domain of Computational Toxicology

Safe drugs, safe food products, and a safe environment for living organisms are of concern in all countries around the world. Toxicology leadership usually stems from governmental regulatory agencies, in the USA including the FDA (Food and Drug Administration), EPA (Environment Protection Agency), DA (Department of Agriculture), NCATS (National Center for Advancing Translational Sciences), NIEHS (National Institute of Environmental Health Sciences) and others. These regulatory agencies are responsible for maintaining the health and well-being of a population and actively seek to prevent any exposures to toxic chemicals. Additionally, the pharmaceutical and biotechnology industry strive to improve patient lives by bringing both new and generic medicines to the market and must do so while minimizing harm to human life. Safety and toxicity screening is critical throughout the steps of any drug development program, starting from the preclinical stage, during clinical trials, and even in post-market surveillance. For the food and agriculture industry, safety screening is also a key step in establishing safe exposure levels to new additives or pesticides. Academics, while rarely developing a product for commercial purposes, do seek to create and test toxicity screens and also assess toxicity mechanism of action (MOA) or mode of action (MoA). This textbook emphasizes the methods of computational toxicology and their potential applications in regulatory science, but the topic is clearly relevant across sectors and around the world.

1.3 Need for Computational Toxicology

The field of computational toxicology has been blooming due to the fundamental limitations of experimental toxicology. While a dizzying array of novel chemical matter is being created every day, traditional experiments are bottlenecked by throughput and cost. In other words, the need for fast toxicity screening and prediction is ever increasing and traditional in vivo and in vitro approaches cannot keep pace. Moreover, there is a global push to avoid the use of animals for experimental testing. Traditional approaches are also limited in the number of doses, time points, organ systems, and combinations that can possibly be tested sanely in one experiment or laboratory.

When using research to guide toxicology regulation, reproducible and rigorous analysis are absolutely required. Bench experiments often have many variables which are difficult to control, including variations in technician, machinery, laboratories, reagent lots, reagent age, or other protocol subtleties. Advances in computer science offer not only faster experiments, but also more reproducible ones. For example, a computational analysis can be exactly repeated by an independent scientist provided that the raw data is available, code is captured in a publicly available source such as GitHub and the compute environment is dockerized. The ease of sharing experiments not only allows computational toxicology to be more rigorous and reproducible, but fosters collaboration between researchers as protocols are readily shared.

1.4 Methods in Computational Toxicology

Many computational techniques, including the ones originated from other fields such as computational chemistry and pure computer science, have been developed and applied in toxicology. To summarize, this book solicited chapters to review some popular methods in computational toxicology that can be used to assess risk, evaluate safety, and/or predict toxicology of a drug or other substance.

Chapter 2 introduced the modeling framework of computational toxicology, defined its scope, listed the major tasks, reviewed the methods, and discussed the challenges in computational toxicology.

Structural alerts and quantitative structure–activity relationship (QSAR) models are two of the most popular methods for predicting toxicological activity of chemicals, especially for the simple toxicological endpoints [18, 19]. Chapter 3 reviewed the applications of structural alerts and QSAR models in computational toxicology and summarized some lessons learned from some successful models. It also discussed some challenges such as making negative predictions, moving to quantitative predictions and weight of evidence approaches.

Emerging technologies such as next-generation sequencing enable fast generation of huge amounts of data. Computational analysis is challenging and crucial to extract knowledge from such big data [20]. Machine learning algorithms have been developed and applied in computational toxicology for prediction of unexpected, toxic effects of chemicals. Moreover, computer science has enabled computational prediction to scale to supermassive sizes. For example, the field of machine learning has birthed matrix and tensor factorization. These two approaches have been used to analyze $>2.5 \times 108$ data points spanning 1300 compounds. It would be absolutely impossible to analyze such a dataset in a simple traditional program such as Microsoft Excel! Chapter 4 reviewed the recent progresses in machine learning-based computational methods and tools and further detailed matrix and tensor factorization approaches.

One feature of modern science is diverse data for a specific scientific question such as specific risk of chemicals to humans and the environment. Thus, integrating diverse data sources from toxicological research to extract more consistent and

reliable knowledge than that provided by any individual data source for risk assessment of chemicals attracts attention of computational toxicologists [21]. Network analysis-based algorithms have been developed for analyzing such large, diverse, and sparse data in computational toxicology [22, 23]. To shed insight into this new method, Chap. 5 presented a network-based systems pharmacology approach that integrates the networks of proteins, genes, drug target, and the human protein–protein interactome for assessing the risk of drug-induced cardiotoxicity in humans.

MoA is the functional or anatomical change caused by chemicals, at the cellular level or at the molecular level that is often used as mechanism of action [24]. It is important knowledge for understanding toxicology of chemicals when the molecular target of chemicals has not yet been determined. It can be used to guide development of predictive models in computational toxicology. Chapter 6 introduced a MoA-guided novel computational toxicology approach that is based on molecular modeling and is implemented in the target-specific toxicity knowledgebase (TsTKb) that contains a pre-categorized database of MoA for chemicals and provides pre-built and category-specific predictive models.

Predictive models in computational toxicology are often developed based on many molecular descriptors using different machine algorithms [25]. One of the key steps in development is to select important descriptors. Chapter 7 discussed different methods for removal of redundant and irrelevant molecular descriptors to improve the performance and interpretability of the model. The strengths and shortcomings of some feature selection and extraction methods in current computational toxicology practices were summarized.

Genomics is the study of genomes, including all molecules such as DNA and RNA and their structures and functions. Adverse effect of a chemical could be caused by the interactions between the chemical and the target genome such as human genome, such is the scope of toxicogenomics [26]. Toxicogenomics has been widely applied in current toxicology practices. A database spanning disciplines of toxicogenomics is the DrugMatrix, which includes gene expression of some 600 therapeutics at multiple doses and 96 signatures relating to phenotypes. Chapter 8 gave a comprehensive description of a legacy resource of toxicogenomics, DrugMatrix and its automated toxicogenomics reporting system, the largest molecular toxicology reference database and informatics systems, which contains thousands of gene expression datasets generated using different microarray platforms.

Given the increasing prevalence of toxicogenomics resources such as the DrugMatrix database, a methodology known as pair ranking was developed to compare transferability between the systems used for testing. Chapter 9 introduced the pair ranking (PRank) method that is developed for quantitative evaluation of assay transferability between the different toxicogenomics platforms.

Several computational toxicology approaches have emerged as a hybrid with computational chemistry. For example, molecular dynamics (MD) simulation was originally used in chemistry to detail interactions between chemicals and biological molecules (including DNA and proteins). For computational toxicologists, MD simulation allows for surveillance of potential fluctuations or conformational changes that a chemical might induce on a biomolecule [27]. Chapter 10 reviewed available

software tools for MD simulations and the challenges to apply these software tools to computational toxicology and summarized key protocols to run MD simulations.

The applicability domain of a prediction model is defined as the structural space that is covered by the chemicals of the training set. It is expected that the predictions from the model for new compounds within the structural space are more accurate than the predictions of chemicals out of the space. Analysis of applicability domains in computational toxicology is important for assessing QSAR models [28]. Chapter 11 reviewed different perspectives of the applicability domain and the existing methods for analysis of applicability domain. It also formalized a holistic approach for utilization of the applicability domain in computational toxicology.

1.5 Potential Applications of Computational Toxicology in Regulatory Science

Computational toxicology has been accepted in the regulation of products. One of the examples is the International Council for Harmonisation M7 (ICH M7) guideline that describes the assessment of carcinogenic risk of mutagenic impurities in drug products [29]. This indicates the state of the art of a computational toxicology method and is the milestone for regulatory acceptance of computational toxicology for pharmaceutical products [30, 31]. In the USA, the FDA accepted QSAR modeling results for impurities in applications of drug products. The FDA developed the Medical Device Development Tools (MDDT) program to qualify tools that can be used in evaluation of medical devices [32]. In the newly released FDA' predictive toxicology roadmap, computational toxicology is listed as one of the new technologies might be able to address some of the needs in regulatory science [33]. The EPA's Endocrine Disruptor Screening Program (EDSP) in the twenty-first century is using computational toxicology, coupling with in vitro methodologies, to prioritize and identify EDSP Tier 1 information needs for pesticide active ingredients that will be included in the registration review program [34]. In Europe, read-across, a commonly used computational toxicology method, is adopted for data gap filling in registrations submitted under the REACH regulation [35]. Computational toxicology is gaining attention in chemical risk assessment and management in China [36]. This book's solicited chapters shed lights on examples of potential applications of computational toxicology in regulatory science in USA, Europe, and China.

In terms of consumer food safety, toxicokinetics, QSAR modeling, and bioinformatics approaches are currently in use. Over time, certainly many more approaches will be added to screen for toxic food products. Chapter 12 reviewed quantitative structure–activity relationships, toxicokinetic modeling and simulation, and bioinformatics in the FDA's Center for Food Safety and Applied Nutrition in-house food ingredient knowledgebase to show the scientific utility of computational toxicology for improving regulatory review efficiency.

In the space of drug development, log regression analysis has predicted drug-induced liver injury, which has proven challenging for both the pharmaceutical industry and regulators. Chapter 13 briefed the drug-induced liver injury (DILI) research efforts at the National Center for Toxicological Research (NCTR), FDA, including drug-label-based-approach to annotate the DILI risk associated with individual drugs including a series of models developed to assess the potential of DILI risk.

Alternative methods including computational toxicology have been considered to inform regulation of drugs, foods, and environmental chemicals. Spanning all three of these fields, a collaborative project across US governmental agencies known as Tox21 screened 10 k chemicals against a large panel of cell-based assays in a quantitative high-throughput screen [37]. Chapter 14 described the efforts to build in vivo toxicity prediction models based on the Tox21 in vitro activity profiles of compounds and discussed the limitations of the current data and strategies for selection of optimal assays to improve the performance of the developed models. The Tox21 project served as powerful fuel for computational predictive modeling across many projects and institutions including predicting point of departure (POD). Chapter 15 reviewed common data modeling approaches that use gene expression profiles to estimate the PODs and compared with the PODs determined using Tox21 data.

From an environmental perspective, endocrine disrupting chemicals (EDCs) are of grave concern and the MOA has been effectively detailed by target-based molecular modeling methods. Computational toxicology methods are an essential and powerful tool to elucidate the MOA of endocrine disruptors. Chapter 16 reviewed the critical processes to perform the molecular modeling of EDCs, including preparation of three-dimensional (3D) structures of the biomacromolecules and EDCs, generation and optimization of the structures of EDC–biomacromolecule complexes, and investigation of the underlying interaction mechanism.

The metabolism of xenobiotics by cytochrome P450 enzymes (CYPs) represents an important mechanism for in vivo compound processing via environmental exposure. Density functional theory (DFT) calculations have been used to highlight the underpinnings of the mechanisms of various environmental toxicants by CYPs including brominated flame retardants. Chapter 17 reviewed the recent progress in molecular simulations of xenobiotic metabolism catalyzed by the typical phase I enzyme CYPs.

Computational toxicology methods including QSAR and read-across are gaining acceptance in regulatory science in the USA, Europe, and Japan [38]. To facilitate the applications of computational toxicology in regulatory science, tools for utilization of QSAR models and read-across have been developed. Chapter 18 introduced a tool (VEGA) that was designed to reduce the barriers between the different read-across and QSAR models for the evaluation of specific chemicals for the assessment of populations of substances. VEGA provides multiple tools for different purposes.

Rigorous and reproducible in silico workflows are needed for toxicological databases and analysis to be successful. OpenTox is stepping in to fill this gap. OpenTox advocates the establishment of good practice and guidance for tracking computational toxicology models to enhance reproducibility, a very important parameter for acceptance of the computational models in regulatory science. Chapter 19 dis-

cussed the implementations of workflows for assessing trusted reproducible in silico evidence supported within OpenTox and OpenRiskNet.

1.6 Conclusions

Translating computational approaches into the complex and intricate field of toxicity is not a simple task. Working together as a community across countries, disciplines, and organizations, computational toxicology was born and has been taking flight. As the field advances, open communication must continue across international borders as well as between regulators and researchers to continually move toward more efficient and effective toxicology.

Disclaimer This article reflects the views of the authors and should not be construed to represent the FDA's views or policies.

References

1. Wennig R (2009) Back to the roots of modern analytical toxicology: Jean Servais Stas and the Bocarmé murder case. Drug Test Anal 1(4):153–155
2. de Bruin YB, Eskes C, Langezaal I, Coecke S, Kinsner-Ovaskainen A, Hakkinen PJ (2009) In: Wexler P, Hakkinen PJ, Mohapatra A, and Gilbert SG (eds) Testing methods and toxicity assessment (including alternatives), in information resources in toxicology, 4th edn. Elsevier, pp 497–513
3. Kolar R (1998) OECD pushes antiquated animal test for endocrine disrupting activity. Altex 15(4):215–217
4. Bruner LH (1992) Alternatives to the use of animals in household product and cosmetic testing. J Am Vet Med Assoc 200(5):669–673
5. Indans I (2002) The use and interpretation of in vitro data in regulatory toxicology: cosmetics, toiletries and household products. Toxicol Lett 127(1–3):177–182
6. McNamee P1, Hibatallah J, Costabel-Farkas M, Goebel C, Araki D, Dufour E, Hewitt NJ, Jones P, Kirst A, Le Varlet B, Macfarlane M, Marrec-Fairley M, Rowland J, Schellauf F, Scheel J. (2009) A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: eye irritation. Regul Toxicol Pharmacol 54(2):197–209
7. Reisfeld B1, Mayeno AN (2012) What is computational toxicology? Methods Mol Biol 929:3–7
8. Kavlock RJ, Ankley G, Blancato J, Breen M, Conolly R, Dix D, Houck K, Hubal E, Judson R, Rabinowitz J, Richard A, Setzer RW, Shah I, Villeneuve D, Weber E (2008) Computational toxicology—a state of the science mini review. Toxicol Sci 103:14–27
9. Benfenati E (2007) Predicting toxicity through computers: a changing world. Chem Central J 1:32
10. Hong H, Neamati N, Winslow HE, Christensen JL, Orr A, Pommier Y, Milne GWA (1998) Identification of HIV-1 integrase inhibitors based on a four-point pharmacophore. Antiviral Chem Chemother 9:461–472
11. Shen J, Zhang W, Fang H, Perkins R, Tong W, Hong H (2013) Homology modeling, molecular docking, and molecular dynamics simulations elucidated α-fetoprotein binding modes. BMC Bioinform 14:S6

12. Ng HW, Doughty SW, Luo H, Ye H, Ge W, Tong W, Hong H (2015) Development and validation of decision forest model for estrogen receptor binding prediction of chemicals using large data sets. Chem Res Toxicol 28:2343–2351
13. Sakkiah S, Ng HW, Tong W, Hong H (2016) Structures of androgen receptor bound with ligands: advancing understanding of biological functions and drug discovery. Expert Opin Ther Targets 20(10):1267–1282
14. Hong H, Shen J, Ng HW, Sakkiah S, Ye H, Ge W, Gong P, Xiao W (2016) A rat α-fetoprotein binding activity prediction model to facilitate assessment of the endocrine disruption potential of environmental chemicals. Int J Environ Res Public Health 13(4):372
15. Luo H, Ye H, Ng HW, Sakkiah S, Mendrick DL, Hong H (2016) sNebula, a network-based algorithm to predict binding between human leukocyte antigens and peptides. Sci Rep 6:32115
16. Hong H, Thakkar T, Chen M, Tong W (2017) Development of decision forest models for prediction of drug-induced liver injury in humans using a large set of FDA-approved drugs. Sci Rep 7(1):17311
17. Selvaraj C, Sakkiah S, Tong W, Hong H (2018) Molecular dynamics simulations and applications in computational toxicology and nanotoxicology. Food Chem Toxicol 112:495–506
18. Ashby J, Tennant RW (1988) Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. Mutat Res 204:17–115
19. Hong H, Tong W, Fang H, Shi L, Xie Q, Wu J, Perkins R, Walker JD, Branham W, Sheehan DM (2002) Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts. Environ Health Perspect 110(1):29–36
20. Hong H, Zhang W, Shen J, Su Z, Ning B, Han T, Perkins R, Shi L, Tong W (2013) Critical role of bioinformatics in translating huge amounts of next-generation sequencing data into personalized medicine. Sci China Life Sci 56(2):110–118
21. Haghighat M, Abdel-Mottaleb M, Alhalabi W (2016) Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans Inf Forensics Secur 11(9):1984–1996
22. Ye H, Luo H, Ng HW, Meehan J, Ge W, Tong W, Hong H (2016) Applying network analysis and Nebula (neighbor-edges based and unbiased leverage algorithm) to ToxCast data. Environ Int 89–90:81–92
23. Luo H, Ye H, Ng H, Shi L, Tong W, Mattes W, Mendrick D, Hong H (2015) Understanding and predicting binding between human leukocyte antigens (HLAs) and peptides by network analysis. BMC Bioinform 16(Suppl 13):S9
24. US EPA. Mechanisms and mode of dioxin action. Retrieved from https://web.archive.org/web/20150226023800/, http://www.epa.gov/ncea/pdfs/dioxin/nas-review/pdfs/part3/dioxin_pt3_ch03_oct2004.pdf. Accessed on 24 Nov 2018
25. Hong H, Xie Q, Ge W, Qian F, Fang H, Shi L, Su Z, Perkins R, Tong W (2008) Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. J Chem Inf Model 48(7):1337–1344
26. Wang C, Gong B, Bushel PR, Thierry-Mieg J et al (2014) The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol 32(9):926–932
27. Sakkiah S, Kusko R, Pan B, Guo W, Ge W, Tong W, Hong H (2018) Structural changes due to antagonist binding in ligand binding pocket of androgen receptor elucidated through molecular dynamics simulations. Front Pharmacol 9:492
28. Tetko IV, Sushko I, Pandey AK, Zhu H, Tropsha A, Papa E, Oberg T, Todeschini R, Fourches D, Varnek A (2008) Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection. J Chem Inf Model 48(9):1733–1746
29. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human use (ICH). Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. Retrieved from https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_R1_Addendum_Step_4_2017_0331.pdf. Accessed on Nov 26 2018

30. Amberg A, Beilke L, Bercu J et al (2016) Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses. Regul Toxicol Pharmacol 77:13–24
31. Barber C, Amberg A, Custer L et al (2015) Establishing best practice in the application of expert review of mutagenicity under ICH M7. Regul Toxicol Pharmacol 73:367–377
32. FDA. Qualification of medical device development tools- guidance for industry, tool developers, and food and drug administration staff. Retrieved from https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM374432.pdf. Accessed on 26 Nov 2018
33. FDA. FDA' predictive toxicology roadmap. Retrieved from https://www.fda.gov/downloads/scienceresearch/specialtopics/regulatoryscience/ucm587831.pdf. Accessed on 26 Nov 2018
34. EPA. Endocrine disruptor screening program (EDSP) in the 21st century. Retrieved from https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-21st-century. Accessed on 26 Nov 2018
35. European Chemical Agency. Read-across assessment framework (RAAF). Retrieved from https://echa.europa.eu/documents/10162/13628/raaf_en.pdf. Accessed on 26 Nov 2018
36. Solid Waste and Chemicals Management Center, MET, China. Latest progress of chemicals management in China. Retrieved from http://chemical-net.env.go.jp/pdf/20180123_Seminar1.pdf. Accessed on 26 Nov 2018
37. Huang R, Xia M, Sakamuru S, Zhao J, Shahane SA, Attene-Ramos M, Zhao T, Austin CP, Simeonov A (2016) Modelling the Tox21 10K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat Commun 7:10425
38. Chesnut M, Yamada T, Adams T, Knight D, Kleinstreuer N, Kass G, Luechtefeld T, Hartung T (2018) Regulatory acceptance of read-across. Altex 35(3):413–419

Rebecca Kusko is the SVP and Head of R&D at Immuneering Corporation. She completed her undergraduate degree at MIT in Biological Engineering and earned a Ph.D. in Genetics and Genomics from BU before joining Immuneering in 2014. At Immuneering, she has been accelerating pharma R&D using advanced algorithms. Additionally, she is a board member of the MAQC Society, an initiative around best practices for big data analysis.

Huixiao Hong is Chief of Bioinformatics Branch, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research (NCTR), US Food and Drug Administration (FDA), working on the scientific bases for regulatory applications of bioinformatics. Before joining the FDA, he was Manager of Bioinformatics Division of Z-Tech, an ICFI company. He was a research scientist position at Sumitomo Chemical Company in Japan and a visiting scientist at National Cancer Institute at National Institutes of Health. He was also Associate Professor and the Director of Laboratory of Computational Chemistry at Nanjing University in China. Dr. Hong is the member of steering committee of OpenTox, the member of board directors of US MidSouth Computational Biology and Bioinformatics Society, and in the leadership circle of US FDA modeling and simulation working group. He published more than 180 scientific papers and served as Editor-in-Chief and Editorial Board member for multiple peer-reviewed journals. He received his Ph.D. from Nanjing University in China in 1990 and conducted research in Leeds University in England in 1990–1992.

Part I
Methods in Computational Toxicology

Chapter 2
Background, Tasks, Modeling Methods, and Challenges for Computational Toxicology

Zhongyu Wang and Jingwen Chen

Abstract Sound chemicals management requires scientific risk assessment schemes capable of predicting physical–chemical properties, environmental behavior, and toxicological effects of vast number of chemicals. However, the current experimental system cannot meet the need for risk assessment of the large and ever-increasing number of chemicals. Meanwhile, current experimental approaches are not sufficient for toxicology to thrive in the era of information. Thus, an auxiliary yet critical field for complementing the experimental sector of chemicals risk assessment has emerged: computational toxicology. Computational toxicology is an interdisciplinary field based especially on environmental chemistry, computational chemistry, chemo-bioinformatics, and systems biology, etc., and it aims at facilitating efficient simulation and prediction of environmental exposure, hazard, and risk of chemicals through various in silico models. Computational toxicology has profoundly changed the way people view and interpret basic concepts of toxicology. Meanwhile, this field is continuously borrowing ideas from exterior fields, which greatly promotes innovative development of toxicology. In this chapter, backgrounds and tasks of computational toxicology are firstly introduced. Then, a variety of in silico models linking key information of chemicals involved in the continuum of source to adverse outcome, such as source emission, concentrations in environmental compartments, exposure concentrations at biological target sites, and adverse efficacy or thresholds are described and discussed. Finally, challenges in computational toxicology such as parameterization for the proposed models, representation of complexity of living systems, and modeling of interlinked chemicals as mixtures are also discussed.

Z. Wang · J. Chen (✉)
Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology, Dalian University of Technology, Dalian, China
e-mail: jwchen@dlut.edu.cn

Z. Wang
e-mail: wzy1989@mail.dlut.edu.cn

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_2

Keywords Computational toxicology · Chemicals risk assessment · Digitization · Fugacity models · PBTK · Systems toxicology · Theoretical and computational chemistry · Molecular model · QSAR

Abbreviations

3R	Replacement, reduction, and refinement
ABMs	Agent/individual-based models
AM1	Austin model 1
AMBER	Assisted Model Building with Energy Refinement
AO	Adverse outcome
AOP	Adverse outcome pathway
CAS	Chemical abstract service
CC	Coupled-cluster
CGenFF	CHARMM General Force Field
CHARMM	Chemistry at HARvard Molecular Mechanics
CI	Configuration interaction
CSBP	Computational systems biology pathway
DFT	Density functional theory
DNA	Deoxyribonucleic acid
EPA	Environmental Protection Agency
ESD	Emission scenario documents
EU	European Union
FF	Force field
GAFF	General AMBER Force Field
HF	Hartree–Fock
HTS	High-throughput screening
IVIVE	In vitro–in vivo extrapolation
KE	Key event
MD	Molecular dynamics
MIE	Molecular initiating event
MM	Molecular mechanics
MNDO	Modified neglect of diatomic overlap
MP	Many-body perturbation
Nrf2	Nuclear factor erythroid 2-related factor 2
OECD	Organization of Economic Cooperation and Development
OSIRIS	Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-test and Test Information
PBDE	Polybrominated diphenyl ester
PBTK	Physiologically based toxicokinetics
QM	Quantum mechanics

QSAR	Quantitative structure–activity relationship
REACH	Registration, evaluation, authorization and restriction of chemicals
RNA	Ribonucleic acid
SE	Semi-empirical
SEURAT	Safety Evaluation Ultimately Replacing Animal Testing
SMILES	Simplified molecular-input line-entry system
SOC	Semi-volatile organic compound
US	United States

2.1 Background for Computational Toxicology

Environmental chemicals (e.g., industrial chemicals, pesticides, pharmaceuticals, personal care products, flame retardants, etc.) constitute a major risk of affecting human and ecological health [1, 2]. To solve problems caused by the pollution of chemicals, it is necessary to assess/predict the exposure, hazard, and risk of chemicals before their entrance into the market, and to restrict the use of chemicals that are of high concern [3]. Such a preventive perception of risk management has brought about the most rigorous regulation in the history of humanity—the registration, evaluation, authorization, and restriction of chemicals (REACH) by the European Union (EU) [4], which has profoundly reshaped the global chemicals management system.

Chemicals risk assessment itself, however, has encountered a bottleneck. Since the twentieth century, the core discipline for assessing toxic effects of chemicals, toxicology, has barely developed. The discipline has relied heavily on in vivo animal tests for a long time, which may violate the replacement, reduction, and refinement principles (3R principles) for animal tests. Besides, conventional in vivo tests have to deal with uncertainties from conservative extrapolations between distinct dose levels or different species. Based only on traditional in vivo tests, it will definitely be difficult to accurately predict the toxicological effect of chemicals on human and ecological health. On the other hand, an unprecedented number of chemicals are awaiting comprehensive assessment. According to the statistics by REACH, there are more than 140,000 chemicals registered in the European markets, among which 80% lack the safety data required [5, 6]. It is estimated that around 500–1000 new chemicals are introduced into the market per year, which is much faster than the speed of traditional chemicals risk assessment (ca. 2–3 years per chemical). Therefore, if the assessment is to be merely based on conventional experiments, application of novel chemicals as well as alternatives of legacy chemicals could be severely impeded due to the inefficiency of the traditional risk assessment process.

In the recent decade, with the acknowledgement of the above-mentioned challenges, chemical toxicity test methodology has been going through a radical revolution [7]. In the report "Toxicity test in 21st century: A vision and strategy" published in 2007 by the United States (US) National Research Council [8], toxicity pathway was emphasized and a paradigm shift from traditionally descriptive toxicology

toward a predictive science that would increasingly rely on in vitro tests based on human tissues and cells was advocated. in silico/computational models were suggested to characterize the toxicity pathways and to assess the exposure, hazard, and risk of chemicals, in order to reduce the time and expense spent, and the number of animals sacrificed and to extend knowledge on the mechanisms of toxic effects.

Toxicity pathways generally refer to biochemical/cellular signaling pathways that if improperly perturbed would eventually lead to adverse health effects. Traditionally, descriptive toxicity end points or apical end points such as individual death or abnormal behavior could be thus reduced or attributed to multiple nodes or key events anchored along the toxicity pathways or networks. The nodes could then be examined with specifically designed in vitro tests. Furthermore, efficiency of in vitro biomacromolecule or cell-based assays has been significantly enhanced by automatic technology, which resulted in so-called high-throughput screening (HTS) technology capable of performing over 1500-well-plate level operations and readouts with just a single run [9, 10]. HTS technology paved the way for projects such as ToxCast [11] and Tox21 [12] that aim at screening relatively large libraries of chemicals and generate the so-called toxicological big data [13].

Novel/alternative experimental methods such as HTS indeed promote the development of toxicology. However, they have met their own problems [14, 15]. Historically, in vivo end points have formed the basis for chemicals risk assessment and are deemed so-called "golden standards" by some toxicologists and regulators. In fact, most cell-based in vitro end points cannot be simply mapped onto the traditional in vivo end points [16]. False positive hits resulting from in vitro tests that are inconsistent with available in vivo evidences hence become a nuisance. It is thus necessary to clearly delineate the relationship between in vitro and in vivo end points, which emphasizes in vitro–in vivo extrapolation (IVIVE) that aims at applying the in vitro results for evaluating in vivo effects [17, 18]. Optimistically, adverse outcome pathways (AOPs) coined by Ankley et al. [19] appear to be a straightforward conceptual framework to explain the specifically designed in vitro end points and apical in vivo end points as molecular initiating events (MIEs)/key events (KEs) and adverse outcomes (AOs), respectively. However, a quantitative AOP that is applicable to IVIVE is still far from real practice. Besides, the types of in vitro end points that can be implemented on an HTS platform are still limited [12]. The implementation of HTS that relies on specific apparatus cannot be adequately cost-effective. Moreover, the number of chemicals screened by Tox21 during 2008–2013 is ca. 10,000 [14], which is de facto much less than the number (>15,000) of newly registered chemicals in a single day on the chemical abstract service (CAS, www.cas.org) system in 2014. In conclusion, the current experimental system for toxicology can meet neither the need for chemicals risk assessment nor the requirement for development of toxicology. Thus, an auxiliary yet critical field for complementing the experimental sector of chemicals risk assessment has emerged: computational toxicology [20].

2.2 Tasks for Computational Toxicology

Computational toxicology is a typical interdisciplinary field based on environmental chemistry, computational chemistry, chemo-bioinformatics, and systems biology, etc., and it aims at facilitating prediction of environmental exposure, hazard, and risk of chemicals by various in silico models. There are two major tasks for computational toxicology: to facilitate sound chemicals management and to shape digitized predictive toxicology.

2.2.1 Facilitating Sound Chemicals Management

One of the urgent needs for sound chemicals management is to build capability to assess virtually all the existing chemicals in markets, i.e., to build a scientific high-throughput system for chemicals risk assessment [1]. HTS technology has made a steady contribution to this envisioned system, but it is still not sufficient.

In 2005, the US Environmental Protection Agency (EPA) founded the National Center for Computational Toxicology to lead and implement research on computational toxicology [21]. Meanwhile, the Joint Research Center of the EU along with many research groups have also carried out projects around core topics of computational toxicology under the 6th and 7th Framework Programs, such as OSIRIS (Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-test and Test Information) [22] and SEURAT (Safety Evaluation Ultimately Replacing Animal Testing) [23]. In this sense, computational toxicology or in silico toxicology serves as one of the tools to meet the requirement of certain countries, regional organizations [e.g., Organization of Economic Cooperation and Development (OECD)] and/or regulatory laws (e.g., REACH) for chemicals risk assessment.

A classic risk assessment scheme includes hazard identification, exposure assessment, effect assessment (dose–response relationships), and risk characterization [24]. Risk characterization is always represented as a mathematical function (e.g., risk quotient) of exposure levels and effect thresholds. Nowadays, a framework of in silico models has emerged, linking key values such as source emission, concentrations in environmental compartments, exposed concentrations at biological target sites, and adverse efficacy or thresholds involved in the continuum of source to adverse outcome of one queried chemical (Fig. 2.1) [20]. The framework needs parameter modifications to be ready for application to other chemicals. For a large number of concerning chemicals, the parameters might be virtually generated by quantitative structure–activity relationship (QSAR) models in a high-throughput manner. Ultimately, the giant gap of safety data required by chemicals risk assessment could be hopefully filled with this framework of computational toxicology in a truly pragmatic sense.

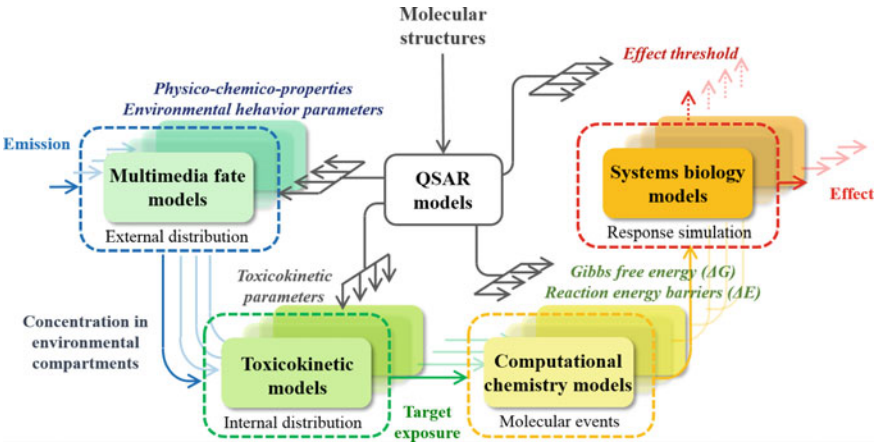

Fig. 2.1 Framework of in silico models of computational toxicology [20]

2.2.2 Shaping Digitized Predictive Toxicology

As mentioned above, traditionally descriptive toxicology is being transformed into a predictive discipline, which certainly requires a sound understanding of the mechanisms underlying the toxicological phenomena of chemicals [7]. The reductionist perception of toxicity pathways or AOPs that breaks apical end points or AOs into MIEs/KEs can feasibly handle complex toxicological phenomena as relatively simple pieces. However, there are two things that wet experiments can hardly achieve.

First, wet experiments do not endeavor to establish systems where metadata of chemicals can be stored and processed conveniently by computer or artificial intelligence. Digitization is a necessary trend for toxicology to thrive in this information era. Obviously, in the very nature of computational toxicology, all simulated objects respond to reasonably digitized counterparts from the real world. For example, in typical QSAR studies, chemicals have to be neatly pretreated as machine-readable formats, such as canonical simplified molecular-input line-entry system (SMILES) codes [25]. To promote sharing and exchanging toxicological big data, public-domain web servers and/or databases have been established, which provides valuable experience for regulators to digitize bioassays in the toxicological field [26, 27]. Digitization of toxicology would significantly decrease laborious work load for toxicologists and allow much more sophisticated studies on complicated systems.

Second, wet experiments do not have resolution high enough to directly observe atom-level behavior of chemical molecules. It is the atom-level behavior that forms the molecular basis for explaining all environmental or toxicological phenomena, and provides molecular mechanisms that can be taken advantage of to predict behavior of newly designed molecules. Nowadays, only molecular simulation based on theoretical and computational chemistry can provide almost infinite resolution and freedom for studying atom-level behavior of chemicals in various situations [28].

Solid theoretical grounds such as quantum mechanics (QM) and density functional theory (DFT) [29] also make it reliable to predict behavior or properties of arbitrary chemical molecules created as in silico objects.

Computational toxicology as an interdisciplinary field has profoundly changed the way people view and interpret the basic concepts of toxicology, and meanwhile it is continuously borrowing ideas from exterior fields, which greatly promotes the innovative development of toxicology.

2.3 Modeling Methods for Computational Toxicology

In practice, computational toxicology creates in silico objects that could properly characterize their real-world counterparts at diverse spatial levels. First of all, each chemical substance (maybe a compound that is queried by regulators or toxicologists) would be digitized as an in silico object. This queried chemical object may interact with or be changed by different in silico situational objects for various research purposes. Useful information could be generated by running simulations with these in silico objects, which results in models with controllable parameters that bear sound or at least reasonable physical meanings. Or otherwise, merely the attributes of a series queried chemicals could be studied for an underlying pattern that has some statistical significance, which would typically result in fast predictive but less mechanistic QSAR models.

2.3.1 Environmental Multimedia Fate Models

The concentrations of chemicals in environmental compartments are the basis for exposure assessment of chemicals [24]. Regulatory laws or policies about chemicals require a standardized documentation of the information on the release of chemicals. For example, OECD declares that the emission scenario documents (ESDs, http://www.oecd.org/env/exposure/esd) which describe source, production, and use of chemicals should be compiled. ESDs can be used to determine emissions of chemicals into the environmental compartments, e.g., air, water, and soil etc., which provide interfaces between human activity and the environmental system.

If the situational objects make a macroscale environmental system and the influence of the queried chemical object on the situational objects is assumed to be negligible, then the simulated model shall resemble a fugacity model proposed by Mackay [30]. The fugacity model in brief can describe behavior/fates of a queried chemical in an idealized environmental system. Typically, a fugacity model consists of:

(1) several mathematical equations describing the mass balance of the chemical among several predefined environmental compartments;

(2) parameters that characterize the environmental media and the behavior of the chemical in certain environmental process. The principles and application of fugacity models have been described in specialized textbooks [31], and fugacity models have been widely employed for predicting environmental fates of concerning chemicals [32, 33]. Furthermore, geographical information systems may improve the fugacity model upon its spatial resolution and visualization [34, 35].

2.3.2 Physiologically-Based Toxicokinetics Models

Exposure can be further divided into external exposure and internal exposure, depending on whether a chemical is located outside or inside of an organism [36]. Concentrations in environmental compartments predicted by fugacity models only quantify the external exposure. The response of different target sites inside the body to the same chemical would be distinct. For example, liver of mammals could metabolize xenobiotics, but fat tissue would typically store hydrophobic chemicals. Therefore, the internal distribution of chemicals actually provides useful information for interpreting mechanisms of toxic effects [36].

Analogous to fugacity models, if the situational objects make an individual organism with a negligible counter-influence of the queried chemical object, then the simulation model would more or less resemble a physiologically based toxicokinetics (PBTK) model [37], which shares similar schemes of parameters and equations with those of the fugacity model. In brief, a PBTK model employs in silico objects representing brain, lung, liver, kidney, fat tissues, venous blood, arterial blood or generally poorly perfused and richly perfused tissues or organs, typically termed as boxes/compartments [38]. According to the flow of chemicals along the blood vessels, ordinary differential equations can be developed to solve the concentrations as a function of time in various boxes [37]. The concentrations in urine or blood can be reversely extrapolated to total intake doses [39], enabling a direct comparison between data from biomonitoring and exposure scenarios derived from ESDs and fugacity models.

Fugacity models or PBTK models have structures that are straightforward to understand and parameters that can be tuned in the simulation, which permit speculation on the mechanism of associated macroscale systems. With the neglected counter-influence of the queried chemical objects on the situational objects, simulations of fugacity models or PBTK models are generally fast and always have deterministic results. Furthermore, with techniques such as Monte Carlo simulation, diagnosis on the sensitivity, and uncertainties are also available making the macroscale empirical equations-based models more robust [32].

Traditionally, exposure of chemicals is viewed as a linear model from sources of emission to targets. However, exposure and effects are not mutually independent. For example, long-term exposure of toxic chemicals would definitely change the parameters of physiological tissues or organs, which is a topic of toxicodynamics.

As for individual behavior, when exposed to irritating chemicals, a person would certainly response to or evade from those harmful stressors. In 2016, a new field termed "computational exposure science" has been coined to specifically simulate the exposure of all types of stressors of concern to risk assessment regulators, where nonlinear relationships between all objects under a context of exposome [40] and exposure ontology have been emphasized [41]. Further incorporation of this computational exposure science might help computational toxicologists to have a more realistic framework for modeling the exposure of chemicals.

2.3.3 Systems Toxicology Models

As mentioned above, during the absorption, distribution, metabolism, and excretion, i.e., toxicokinetics by the organism, xenobiotic chemicals would also exert their influence on the organism after they reach certain targets, which results in toxicodynamics of chemicals. In this sense, the target sites can be regarded as interfaces linking the xenobiotic chemicals and the physiological functions of life systems.

Homeostasis, proliferation, differentiation, and apoptosis of cells, the basic functional units for life systems, are regulated by cellular signaling pathways/network. Nowadays, bioinformatic and systems toxicology employ network models to map the cellular biochemical components such as an upstream DNA sequences and its downstream mRNA, by analyzing data from molecular biology technologies, especially genomics, transcriptomics, proteomics, etc. [42]. When xenobiotic chemicals are tested within these -omics assays and certain toxic effects are focused on, the xenobiotics could be thus anchored onto the biological network and linked with phenotypical in vitro or in vivo end points or diseases [43, 44]. The chemo-bioinformatics and -omics assays would be an efficient strategy for selecting more relevant targets/marker interfaces for xenobiotics to exert their influences from a large in vitro test battery [45, 46].

Qualitative signaling networks delineated by chemo-bioinformatics are indeed informative. However, they still cannot quantitatively describe the toxicodynamics or the dose–response curve of tested chemicals. Borrowing concepts from cybernetics, general network motifs, e.g., negative/positive feedback loop, feed-forward loop, etc., have been extracted from the biochemical components of cells [47]. Furthermore, these motifs can be composed as functional modules, e.g., hypersensitivity, periodical oscillation, cellular memory, etc., which serve as in silico situational objects for simulating the dynamics of the cellular signaling networks. These models termed as computational systems biology pathway (CSBP) models, could give results that can be compared with those of associated in vitro tests, becoming a promising tool for chemicals risk assessment [47]. For example, a feed-forward loop can explain the hormesis-shape dose–response curves of the phase I and phase II metabolism with relatively low exposed doses of xenobiotic chemicals [48]. A CSBP model was also applied successfully to describe the anti-oxidative stressors responses regulated by nuclear factor erythroid 2-related factor 2 (Nrf2) [49].

CSBP models are based purely on the topology of signaling networks, not considering the heterogeneous cellular surroundings. If the crowded space of real cells is needed, then agent/individual-based models (ABMs) might be considered [50–53]. But generally ABMs would have single cells as their agents, which are typically used to simulate the cellular or tissue-level behavior/effects such as tumorigenesis or vasculogenesis [54]. The agents or cells would act according to a predefined set of rules. Specialized software and platform tools such as CompuCell3D [55] and NetLogo (http://ccl.northwestern.edu/netlogo/) could be employed to perform these simulations. With a similar strategy, tissues, and organs such as hepatic lobule can also be simulated if certain functional and/or survival dose–response curve data is integrated into a model, which enables one straightforward form of so-called virtual tissues [56, 57].

In conclusion, the Systems toxicology models introduced in this section are at the cutting edge of computational toxicology, most of which are explanatory, tentative, and not ready for prediction. Nonetheless, in the future context of toxicology and chemicals risk assessment, these models would promisingly serve as better alternatives than the current non-testing systems.

2.3.4 Molecular Models

As suggested by AOPs, toxic effects of chemicals originate from MIEs, i.e., the interaction between chemical molecules and biomacromolecules [19, 58]. In a generalized sense, partition/adsorption and transformation of chemicals in either an inorganic or biological environment can also originate from molecular behavior/events. As previously discussed, the experimental sector of toxicology generally cannot observe atom-level behavior of the molecules. With the advent of theoretical and computational chemistry, computational toxicologists are now able to establish so-called molecular models that include the queried chemical molecules and their situational objects, i.e., their surrounding molecules to be interacted with. In an MIE, the situational objects are typically functional proteins [58]. Meanwhile, in a simulation for the gaseous transformation of a chemical, the situational objects might be airborne reactive species such as hydroxyl radical or chlorine radical.

QM methods from computational chemistry can be adopted to calculate electronic structures of molecules and elementary steps along a chemical reaction pathway. Ab initio QM methods, including Hartree–Fock (HF) [59], configuration interaction (CI) [60], many-body perturbation (MP) theory, and coupled-cluster (CC) theory [61], etc., are based only on fundamental assumptions, such as Born–Oppenheimer approximation which assumes that the motion of atomic nuclei and electrons in a molecule can be separated. Among ab initio QM methods, post-HF methods, e.g., CI, MP, and CC are proved to be able to give very accurate results that are very consistent with experimental observations. However, ab initio QM methods, especially the post-HF methods require a large amount of computer resources. For larger systems with up to a hundred atoms, these methods would not be feasible currently.

On the other hand, semi-empirical (SE) methods have been created that adopt experimental values or empirical parameters as expedient replacement for certain complicated time-consuming integral items of ab initio QM methods. Therefore, the SE methods are much faster than the ab initio QM methods, but less accurate especially when the molecule or system has non-typical conformations (such as those of transition states). Well-known SE methods include MNDO (modified neglect of diatomic overlap) [62], AM1 (Austin model 1) [63] etc.

Besides, DFT is a promising method based on slightly different assumptions from that of QM theories, which can simulate larger systems, with accuracy comparable to ab initio QM methods. Therefore, DFT has become a widely applied method for computational toxicology. For example, the DFT method has been employed to probe the hydrolysis pathways of antibiotics [64] and sites of metabolism of brominated flame retardants by reactive oxo-heme of P450 enzymes [65].

For biomacromolecules such as proteins with thousands of atoms or inorganic systems with explicit solvent molecules in condensed state, the DFT method is generally not feasible. For these systems, the theory must be further simplified. An empirical force field (FF) that describes the interaction between atoms from a classical mechanics perspective provides a feasible route for the large-scale simulations. In some widely employed academic force fields such as CHARMM (Chemistry at HARvard Molecular Mechanics) [66] and AMBER (Assisted Model Building with Energy Refinement) [67] force fields, potential energy functions are used to describe bonding lengths, angles, dihedrals, electrostatic/Coulomb interactions, and Van der Waals interactions. The parameters of FF-based methods/software are more complex than those of QM or DFT methods/software. Therefore, it is also relatively complicated to establish model systems with FF-based methods. CHARMM and AMBER force fields have provided compatible topologies and high-quality FF parameters for common biomolecules such as proteins, DNAs, RNAs, lipids, and carbohydrates, making it very convenient to simulate these biological systems. Besides, CHARMM and AMBER also provide CGenFF (CHARMM General Force Field) [68] and GAFF (General AMBER Force Field) [69] for small molecules, which are useful for toxicological systems involving xenobiotic molecules. FF-based methods typically include molecular mechanics (MM) that minimize energy of a conformation to obtain its best geometry, and molecular dynamics (MD) or Monte Carlo simulations that sample the ensemble space of the simulated systems and generate trajectories for real-time/post-treatments to obtain useful physical quantities [28].

Unlike QM or DFT methods, typical FF-based methods do not consider the forming and breaking of covalent bonds. However, it is of interests to toxicologists to simulate chemical reactions taking place in either an inorganic environment or in an enzymatic environment [70]. Thus, schemes of interfaces between QM and MM methods have been developed to simulate these special cases, with the reactive site handled by QM methods and the rest of the system handled by MM methods. QM/MM method is now mostly utilized for enzymatic systems and for explanatory purposes.

Molecular models are different from the previously introduced macroscopic models in many aspects. First of all, when a chemical substance is queried in a macroscopic model, it always refers to a population of a huge number of same-type molecules. In chemistry or physics, Avogadro's constant is employed to translate the microscopically huge number of molecules into macroscopic amount of substances with mole as the unit. Noticeably, even a nanomole of substance corresponds to 6.02×10^{14} molecules, which is definitely an astronomical figure far beyond the capacity of molecular simulations. In a molecular simulation, a queried chemical is typically represented by a single (in some cases a few) molecule(s) of its type. For example, when a MD simulation is carried out to calculate binding free energy of a ligand to a receptor, there is typically just one ligand molecule in the simulated system. A macroscopic binding affinity bioassay usually gives an inhibitory concentration–response curve, from which a half-inhibition concentration (IC_{50}) or inhibition constant (K_i) can be calculated as quantitative features for the binding affinity. However, molecular simulation, with only one ligand molecule, would only give the free energy difference between or the probability distribution of the binding state and the free state of the ligand, which could be later translated into equilibrium constants such as K_i. In conclusion, molecular simulation checks the relative potential energies of important microstates, and then uses the obtained potential energy surface to explain the macroscopic phenomena with principles of statistical mechanics or statistical thermodynamics. Currently, molecular simulation mainly aims at explaining experimental phenomena by revealing molecular mechanisms, and qualitatively foreseeing tendencies of some properties for a series of congeners. But, as computational power is increasing and novel efficient algorithms are being implemented, larger systems with thousands of atoms would ultimately be simulated with more accurate and sophisticated empirical FF or even ab initio QM/DFT methods, which should give much better predictions to convincingly fill the data gap required by chemicals risk assessment.

2.3.5 QSAR Models

With the so-called toxicological big data [13], machine learning algorithms have also proven their usefulness in predicting parameters required by the macroscale models or data required by chemicals risk assessment, which typically refer to QSAR modeling or supervised learning in the field of computational toxicology or machine learning, respectively. QSARs are based on the linear free energy relationship theory suggested by Hammett [71], Hansch et al. [72] or even on the chemistry intuition that "chemicals with similar structures have similar properties."

Distinct from all previously introduced models, QSARs do not simulate a particular chemical object in a physical process/event. In a QSAR study, features are firstly extracted from a series of chemical objects (so-called training and testing sets) or a series of molecular scenarios where each queried chemical interacts with certain situational objects, and then these features are employed to predict certain properties

of the chemicals. It deserves mentioning that it is correlation rather than causality between the features and targeted properties that QSARs eventually obtain. To learn more about the history and/or future perspectives of QSAR, some literature reviews can be consulted [25, 73, 74].

From a mathematical point of view, a QSAR model has three basic elements: the features ($\mathbf{X}$), the properties to be predicted ($\mathbf{Y}$), and the algorithm adopted to map the features onto the properties (f). In brief, a QSAR model can always be expressed as $\mathbf{Y} = f(\mathbf{X})$. As for $\mathbf{X}$, if features are extracted only from chemical molecules themselves, then these features are typically called molecular (structural) descriptors. Details about molecular descriptors can be found in Todeschini and Consonni's *Molecular Descriptor for Chemoinformatics* [75]. Besides, the application domain beyond which QSAR models could not be reliable also depends strongly on the feature space of chemicals in the training set. As for $\mathbf{Y}$, if the properties have discrete categories/levels, then it refers to a classification task, resulting in a predictive model termed as classifier. If the properties have continuous values it refers to a regression task, resulting in a regressor. The quality of QSAR models heavily relies on the quality of the input data sets. Therefore, collection and curation of data need particular patience and caution. As for f, nowadays, a large number of algorithms for classification or regression are available, e.g., multivariate linear regression, partial least regression, support vector machine, decision tree, naïve Bayes, artificial neural network, etc., as well as ensemble algorithms such as random forest [76], among which computational toxicologists may find some suitable for their specialized cases.

It is also notable that molecular models could inspire novel descriptors that can better characterize the underlying pattern for certain toxicological phenomena. For example, simulation of the interaction between halogenated compounds and human transthyretin protein with QM/MM methods indicated specific descriptors for QSAR modeling [77]. Similarly, QSARs can translate the time-consuming computational chemistry models into empirical rules and mathematics, thus resulting in more efficient predictive models. For example, Rydberg and Olsen et al. performed a series of QM/DFT simulations on the active sites of P450 enzymes with various types of small molecules [78–80], based on which a web server named SMARTCyp has been developed that is capable of rapidly predicting sites of metabolism and associated reaction energy barriers [81].

Currently, the rapid development of modern machine learning algorithms could most likely promote a renaissance of the field of QSAR modeling. Previously unnoticed details might also be recaptured by innovative feature extraction and well-established machine learning algorithms, making scientists' conclusions or predictions less arbitrary and more robust.

2.4 Challenges for Computational Toxicology

Although modeling frameworks seem to have been nicely established, challenges in computational toxicology still remain for many aspects.

2.4.1 Nuisance in Parameters for In Silico Models

One vital issue for all computational models is that their simulation can never begin until the in silico objects of the models are completely and properly parameterized. This issue is especially typical for macroscale models, i.e., the fugacity model, PBTK model, or systems toxicology models [37]. For example, the fugacity model requires partition coefficients between each pair of environmental compartments and rate constants of certain types of transforming reactions. Obtaining values of these parameters for certain models is not a trivial job. Generally speaking, the macroscale models based on empirical equations are parameter-intensive models in their very nature. Although experiments can be performed under the OECD or EPA test guidelines to determine a part of the parameters, sometimes wet experiments are time-consuming and even impossible due to extremely inert reactivity, low solubility, etc., of the queried chemicals. It is also not pragmatic to conduct wet experiments for all the queried chemicals. Hopefully, if adequate experimental data around certain parameters have been elaborately collected, QSAR tools could be built to quickly predict those parameters. However, QSAR models still suffer from limited application domains and weak mechanistic interpretation.

Computational chemistry models can directly simulate the microscopic process of certain partitions or reactions of chemical substances and calculate the properties that can be used as parameters for the macroscale models. For example, advanced quantum chemistry modeling techniques have been proven to be able to calculate the gaseous reaction rate constants of semi-volatile organic compounds (SOCs) with airborne radicals very accurately [82, 83]. Meanwhile, molecular dynamics simulation with enhanced sampling strategy could re-establish free energy curves of water–air partition of certain chemicals with properly developed force field parameters [84], then the hydration free energy can be readily translated into water–air partition coefficients needed for the macroscale models. A warning on these models is that establishment of the modeling systems is sometimes truly sophisticated and simulation for the modeling systems requires a huge amount of computational resources and time. Of note, the parameterization of empirical force fields for xenobiotic small molecules could be a very serious issue when practicing associated simulations. Note that even with CGenFF or GAFF, the generated FF parameters for complex xenobiotic molecules may still bear very high penalties indicating that these generated FF parameters are not suitable for the simulation. Closer cooperation between computational toxicologists and theoretical chemists, thus, should be encouraged in order to overcome this issue. Nevertheless, computational chemistry models based on relatively rigorous theories could provide us unprecedented details on molecular mechanisms of atom-level behavior of chemicals, which seems to envision an ultimate solution for predicting the macroscale parameters.

2.4.2 *In the Face of Complex Living Systems*

Although the fugacity models possess simplified structures under the principle of Ockham's Razor, they are indeed good at describing the fate of chemicals within an environmental system if properly parameterized [31]. The same goes for PBTK models. As mentioned previously, these models neglect the counter-influence of chemicals on respective situational objects, simulating merely the fate of chemicals rather than the fate of the whole systems, and thus are relatively straightforward to build and to simulate.

However, at cellular or subcellular levels, things are significantly different. Because it is the fates of the biological systems that are focused on in the simulations, the counter-influence of chemicals on the cellular components and function cannot be ignored any more. To enable the simulation of perturbation on biological systems from chemicals, mechanisms and functional structures of biological systems themselves and interfaces between the biological systems and the xenobiotic chemicals must be firstly known. It is well acknowledged that modern toxicology studies based on in vitro assays have successfully parsed particular interfaces. However, decoding the mystery of life is never an easy job. The functional structures of most biological systems are still not clear. Moreover, there is a lack of knowledge on relevant gene polymorphisms of human populations. Computational toxicologists obviously cannot digitize so many objects within these unknown fields. This state will hopefully be improved with systems biology and "-omics" technologies that are emphasized to piece the whole picture of life together [46, 47].

Furthermore, the ultimate goal of chemicals risk assessment is to protect not only human beings, but also the diverse species that inhabit the earth's ecosystems, which is also a topic of ecotoxicology [85]. In fact, the sensitivity of different species to the same chemicals could vary distinctly. However, it is neither pragmatic nor necessary to evaluate the toxic effects of chemicals on more than a million species via wet experiments. A key point is to extrapolate toxicities of chemicals cross different species, which also requires a sound understanding of mechanisms and functional structures of different concerned species [86]. Computational toxicology has provided some strategies to address the problem of cross-species extrapolation. For example, by modifying structures and parameters of PBTK models, internal distribution or dynamic bioaccumulation of chemicals across different species can be evaluated [87]. To simulate biomacromolecules from different species, homology modeling [88] could provide atom-level structures that have not been determined by X-ray diffraction or nuclear magnetic resonance approaches. However, there is still very little knowledge of non-human species, except for a very limited number of model species [89], which can be employed for establishing in silico models. Therefore, feasible and convincing models for cross-species extrapolation are still far from real practice.

After all, computational toxicology is only an emerging interdisciplinary field. It is not in its nature obliged to decode the mystery of life. Therefore, development of computational toxicology would definitely rely on the advances of exterior fields, especially on the thrilling breakthrough of life sciences.

2.4.3 The Everlasting List of Interlinked Chemicals as Mixtures

It should be noted that some chemicals will always be interlinked with certain other chemicals, such as impurities, plasticizers, cosolvents, etc., particularly in industrial products or in pharmaceutical or cosmetic formulations. It should also be noticed that a CAS- or REACH-registered chemical could have multiple reaction products that could be generated during major environmental processes such as photolysis, hydrolysis, and transformation by reactive oxygen species or by biochemical metabolism. These latent products are not necessarily registered chemicals, and they could be more hazardous to human and ecological health than their parent compounds. For example, polybrominated diphenyl esters (PBDEs) can possibly be transformed into HO-PBDEs or even notorious dioxins with the help of P450 enzymes [90, 91]. Evidences from in vitro assay have proved that the HO-PBDEs are more efficacious on disrupting thyroid receptor β than the parent PBDEs [92].

No matter how interlinked the chemicals are, they always appear as mixtures in the real environment. In an inorganic environment, mixtures of chemicals could result in phenomena such as catalysis, inter-reaction, adsorption, etc. Experimentalists who study the environmental fate of a queried chemical, usually adopt so-called central composite design to probe the influence of environmental factors on the chemical [93, 94]. Note that some of the environmental factors such as anions and dissolved organic matters are also components of the mixture. If the queried chemical is readily reactive, then spectra of downstream chemicals of associated reactions would also contribute to the composition of the mixture. Theoretically, these central composite designs or full-factorial experiments might characterize a response surface or hypersurface that quantitatively reflects influences from components of the mixture. In biological systems, similar experiments could be conducted for mixtures in order to determine so-called joint toxic effects or combined effects of chemicals [95]. The most vital defect of the response hypersurface obtained via wet experiments is that it just describes the phenotypic response or apparent phenomena and cannot reveal underlying mechanisms. Therefore, in its nature, the response hypersurface can never be used to predict quantitative impact of a chemical that is not among the original components/factors during the experimental determination of the response hypersurface. In addition, scale of these joint effect-determining experiments would grow exponentially as the number of factors increases. It seems that, mechanism-based models might be the only promising tools to predict the joint toxic effects of mixtures with arbitrary composition. Nonetheless, the response hypersurface indeed

provides valuable information as experimental evidences, which could serve as validating benchmarks for in silico models.

Current explanatory or prediction models/schemes for joint toxic effects are coarse, and predicted toxic thresholds can only be adopted in a conservative manner over a large safety factor [24]. Systems toxicology might shed light on the joint toxic effects [96]. Except for molecular models, almost all the macroscale models are designed for just one chemical. If reliance on so-called safety factors is to be diminished and models based on transparent mechanisms are to be emphasized in the future toxicology, novel models with associated in silico objects have to be developed, which shall allow reasonable characterization of the dynamic interaction network of multiple queried chemicals.

2.5 Conclusions and Perspectives

Currently, except for QSAR models, computational toxicology models are rarely employed in real practice for chemicals risk assessment. Nevertheless, the modeling framework of computational toxicology has envisioned an attractive paradigm for future toxicity testing and toxicological studies. With previously described reality-mirroring in silico models, general rules that are transferable among similar cases at the same spatial level can be modeled by well-understood mathematics or logic rather than obscurely descriptive paragraphs. Although challenges remain for computational toxicology, the endeavor to overcome these challenges will definitely result in continuous innovation and prosperous development for the field of both chemicals risk assessment and toxicology.

References

1. UNEP (2013) Global chemicals outlook—towards sound management of chemicals. United Nations Environment Programme, Nairobi
2. Rappaport SM, Smith MT (2010) Environment and disease risks. Science 330(6003):460–461
3. Schwarzman MR, Wilson MP (2009) New science for chemicals policy. Science 326(5956):1065–1066
4. EU (2006) Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006, concerning the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH). Official Journal of the EU, EU, Brussels
5. Hartung T (2009) Toxicology for the twenty-first century. Nature 460(7252):208–212
6. Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E (2009) The toxicity data landscape for environmental chemicals. Environ Health Perspect 117(5):685–695
7. Collins FS, Gray GM, Bucher JR (2008) Toxicology—transforming environmental health protection. Science 319(5865):906–907
8. NRC (2007) Toxicity testing in the 21st century: a vision and a strategy. National Research Council, Washington, DC

9. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP (2006) Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A 103(31):11473–11478
10. Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, Auld DS (2007) High-throughput screening assays for the identification of chemical probes. Nat Chem Biol 3(8):466–479
11. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5–12
12. Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, Austin CP, Shinn P, Simeonov A, Tice RR, Xia M (2013) The Tox21 robotic platform for the assessment of environmental chemicals—from vision to reality. Drug Discov Today 18(15–16):716–723
13. Zhu H, Zhang J, Kim MT, Boison A, Sedykh A, Moran K (2014) Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. Chem Res Toxicol 27(10):1643–1651
14. Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121(7):756–765
15. Judson R, Kavlock R, Martin M, Reif D, Houck K, Knudsen T, Richard A, Tice RR, Whelan M, Xia M, Huang R, Austin C, Daston G, Hartung T, Fowle JR III, Wooge W, Tong W, Dix D (2013) Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. Altex 30(1):51–66
16. Thomas RS, Black MB, Li L, Healy E, Chu T-M, Bao W, Andersen ME, Wolfinger RD (2012) A comprehensive statistical analysis of predicting *in vivo* hazard using high-throughput *in vitro* screening. Toxicol Sci 128(2):398–417
17. Yoon M, Campbell JL, Andersen ME, Clewell HJ (2012) Quantitative *in vitro* to *in vivo* extrapolation of cell-based toxicity assay results. Crit Rev Toxicol 42(8):633–652
18. Yoon M, Efremenko A, Blaauboer BJ, Clewell HJ (2014) Evaluation of simple in vitro to in vivo extrapolation approaches for environmental compounds. Toxicol Vitro 28(2):164–170
19. Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741
20. Wang ZY, Chen JW, Qiao XL, Li XH, Xie HB, Cai XY (2016) Computational toxicology: oriented for chemicals risk assessment. Scientia Sinica Chimica 46(2):222–240 (in Chinese with English abstract)
21. Kavlock R, Dix D (2010) Computational toxicology as implemented by the US EPA: providing high throughput decision support tools for screening and assessment chemical exposure, hazard and risk. J Toxicol Environ Health-Pt b-Crit Rev 13(2–4):197–217
22. Vermeire T, Aldenberg T, Buist H, Escher S, Mangelsdorf I, Paune E, Rorije E, Kroese D (2013) OSIRIS, a quest for proof of principle for integrated testing strategies of chemicals for four human health endpoints. Regul Toxicol Pharmacol 67(2):136–145
23. Daston G, Knight DJ, Schwarz M, Gocht T, Thomas RS, Mahony C, Whelan M (2015) SEURAT: Safety Evaluation Ultimately Replacing Animal Testing—recommendations for future research in the field of predictive toxicology. Arch Toxicol 89(1):15–23
24. Van Leeuwen CJ, Vermeire TG (2007) Risk assessment of chemicals: an introduction, 2nd edn. Springer, Dordrecht
25. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010
26. Judson RS, Martin MT, Egeghy P, Gangwal S, Reif DM, Kothiya P, Wolf M, Cathey T, Transue T, Smith D, Vail J, Frame A, Mosher S, Hubal EAC, Richard AM (2012) Aggregating data for computational toxicology applications: The US Environmental Protection Agency (EPA) aggregated computational toxicology resource (ACToR) system. Int J Mol Sci 13(2):1805–1831

27. Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO, Wang J, Xiao J, Zhang J, Bryant SH (2010) An overview of the PubChem BioAssay resource. Nucleic Acids Res 38:D255–D266
28. Hinchliffe A (2008) Molecular modelling for beginners, 2nd edn. Wiley, UK
29. Lewars EG (2011) Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics. Springer, Netherlands
30. Mackay D (1979) Finding fugacity feasible. Environ Sci Technol 13(10):1218–1223
31. Mackay D (2001) Multimedia environmental models: the fugacity approach, 2nd edn. CRC Press, Boca Raton
32. Ao J, Chen J, Tian F, Cai X (2009) Application of a level IV fugacity model to simulate the long-term fate of hexachlorocyclohexane isomers in the lower reach of Yellow River basin, China. Chemosphere 74(3):370–376
33. Arnot JA, Mackay D, Webster E, Southwood JM (2006) Screening level risk assessment model for chemical fate and effects in the environment. Environ Sci Technol 40(7):2316–2323
34. Coulibaly L, Labib ME, Hazen R (2004) A GIS-based multimedia watershed model: development and application. Chemosphere 55(7):1067–1080
35. Song HM, Xu LY (2011) A method of urban ecological risk assessment: combining the multimedia fugacity model and GIS. Stoch Environ Res Risk Assess 25(5):713–719
36. Escher BI, Hermens JLM (2004) Internal exposure: linking bioavailability to effects. Environ Sci Technol 38(23):455A–462A
37. Bessems JG, Loizou G, Krishnan K, Clewell HJ III, Bernasconi C, Bois F, Coecke S, Collnot E-M, Diembeck W, Farcal LR, Geraets L, Gundert-Remy U, Kramer N, Kuesters G, Leite SB, Pelkonen OR, Schroeder K, Testai E, Wilk-Zasadna I, Zaldivar-Comenges J-M (2014) PBTK modelling platforms and parameter estimation tools to enable animal-free risk assessment Recommendations from a joint EPAA—EURL ECVAM ADME workshop. Regul Toxicol Pharmacol 68(1):119–139
38. Blaauboer BJ (2010) Biokinetic modeling and *in vitro-in vivo* extrapolations. J Toxicol Environ Health-Pt b-Crit Rev 13(2–4):242–252
39. Tan Y-M, Liao KH, Clewell HJ (2007) Reverse dosimetry: interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. J Expo Sci Environ Epidemiol 17(7):591–603
40. Cui YX, Balshaw DM, Kwok RK, Thompson CL, Collman GW, Birnbaum LS (2016) The exposome: embracing the complexity for discovery in environmental health. Environ Health Perspect 124(8):A137–A140
41. Egeghy PP, Sheldon LS, Isaacs KK, Ozkaynak H, Goldsmith M-R, Wambaugh JF, Judson RS, Buckley TJ (2016) Computational exposure science: an emerging discipline to support 21st-century risk assessment. Environ Health Perspect 124(6):697–702
42. Waters MD, Fostel JM (2004) Toxicogenomics and systems toxicology: aims and prospects. Nat Rev Genet 5(12):936–948
43. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kuijer MB, Matos RC, Tran TB, Whaley R, Glennon RA, Hert J, Thomas KLH, Edwards DD, Shoichet BK, Roth BL (2009) Predicting new molecular targets for known drugs. Nature 462(7270):175–181
44. Li J, Wu Z, Cheng F, Li W, Liu G, Tang Y (2014) Computational prediction of microRNA networks incorporating environmental toxicity and disease etiology. Sci Rep 4:5576
45. Hong H, Zhang W, Shen J, Su Z, Ning B, Han T, Perkins R, Shi L, Tong W (2013) Critical role of bioinformatics in translating huge amounts of next-generation sequencing data into personalized medicine. Sci China-Life Sci 56(2):110–118
46. Zhang X, Xia P, Wang P, Yang J, Baird DJ (2018) Omics advances in ecotoxicology. Environ Sci Technol 52(7):3842–3851
47. Zhang Q, Bhattacharya S, Andersen ME, Conolly RB (2010) Computational systems biology and dose-response modeling in relation to new directions in toxicity testing. J Toxicol Environ Health-Pt b-Crit Rev 13(2–4):253–276
48. Zhang Q, Bhattacharya S, Conolly RB, Clewell HJ, Kaminski NE, Andersen ME (2014) Molecular signaling network motifs provide a mechanistic basis for cellular threshold responses. Environ Health Perspect 122(12):1261–1270

49. Zhang Q, Pi J, Woods CG, Andersen ME (2010) A systems biology perspective on Nrf2-mediated antioxidant response. Toxicol Appl Pharmacol 244(1):84–97
50. Birbaumer M, Schweitzer F (2011) Agent-based modeling of intracellular transport. Eur Phys J B 82(3–4):245–255
51. Dalmasso G, Zapata PAM, Brady NR, Hamacher-Brady A (2017) Agent-based modeling of mitochondria links sub-cellular dynamics to cellular homeostasis and heterogeneity. PLoS ONE 12(1):e0168198
52. Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 99:7280–7287
53. Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T, DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310(5750):987–991
54. Kleinstreuer N, Dix D, Rountree M, Baker N, Sipes N, Reif D, Spencer R, Knudsen T (2013) A computational model predicting disruption of blood vessel development. PLoS Comput Biol 9(4):e1002996
55. Swat MH, Thomas GL, Belmonte JM, Shirinifard A, Hmeljak D, Glazier JA (2012) Multi-scale modeling of tissues using CompuCell 3D. In: Asthagiri AR, Arkin AP (eds) Computational methods in cell biology, vol 110. Academic Press, Waltham, pp 325–366
56. Wambaugh J, Shah I (2010) Simulating microdosimetry in a virtual hepatic lobule. PLoS Comput Biol 6(4):e1000756
57. Shah I, Wambaugh J (2010) Virtual tissues in toxicology. J Toxicol Environ Health-Pt b-Crit Rev 13(2–4):314–328
58. Allen TEH, Goodman JM, Gutsell S, Russell PJ (2014) Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem Res Toxicol 27(12):2100–2112
59. Bobrowicz FW, Goddard WA III (1977) The self-consistent field equations for generalized valence bond and open-shell hartree-fock wave functions. Plenum, New York
60. Pople JA, Headgordon M, Raghavachari K (1987) Quadratic configuration-interaction—a general technique for determining electron correlation energies. J Chem Phys 87(10):5968–5975
61. Bartlett RJ (1981) Many-body perturbation-theory and coupled cluster theory for electron correlation in molecules. Annu Rev Phys Chem 32:359–401
62. Dewar MJS, Thiel W (1977) Ground-states of molecules. 38. MNDO method—approximations and parameters. J Am Chem Soc 99(15):4899–4907
63. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) The development and use of quantum-mechanical molecular-models. 76. AM1—a new general-purpose quantum-mechanical molecular-model. J Am Chem Soc 107(13):3902–3909
64. Zhang HQ, Xie HB, Chen JW, Zhang SS (2015) Prediction of hydrolysis pathways and kinetics for antibiotics under environmental pH conditions: a quantum chemical study on cephradine. Environ Sci Technol 49(3):1552–1558
65. Shaik S, Kumar D, de Visser SP, Altun A, Thiel W (2005) Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. Chem Rev 105(6):2279–2328
66. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmueller H, MacKerell AD Jr (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14(1):71–73
67. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the AMBER biomolecular simulation package. WIREs Comput Mol Sci 3(2):198–210
68. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD Jr (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31(4):671–690
69. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174
70. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. Angew Chem Int Ed 48(7):1198–1229

71. Hammett LP (1936) The effect of structure upon the reactions of organic compounds. Benzene derivatives. J Am Chem Soc 59(1):96–103
72. Hansch C, Maloney PP, Fujita T (1962) Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature 194(4824):178
73. Cronin MTD, Schultz TW (2003) Pitfalls in QSAR. J Mol Struct-Theochem 622(1–2):39–51
74. Ekins S (2014) Progress in computational toxicology. J Pharmacol Toxicol Methods 69(2):115–140
75. Todeschini R, Consonni V (2009) Molecular descriptors for chemoinformatics, vol I & II. Wiley, Weinheim
76. Flach P (2012) Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, Cambridge
77. Yang X, Xie H, Chen J, Li X (2013) Anionic phenolic compounds bind stronger with transthyretin than their neutral forms: nonnegligible mechanisms in virtual screening of endocrine disrupting chemicals. Chem Res Toxicol 26(9):1340–1347
78. Olsen L, Rydberg P, Rod TH, Ryde U (2006) Prediction of activation energies for hydrogen abstraction by cytochrome P450. J Med Chem 49(22):6489–6499
79. Rydberg P, Ryde U, Olsen L (2008) Prediction of activation energies for aromatic oxidation by cytochrome P450. J Phys Chem A 112(50):13058–13065
80. Rydberg P, Ryde U, Olsen L (2008) Sulfoxide, sulfur, and nitrogen oxidation and dealkylation by cytochrome P450. J Chem Theory Comput 4(8):1369–1377
81. Rydberg P, Gloriam DE, Zaretzki J, Breneman C, Olsen L (2010) SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. ACS Med Chem Lett 1(3):96–100
82. Li C, Chen J, Xie H-B, Zhao Y, Xia D, Xu T, Li X, Qiao X (2017) Effects of atmospheric water on •OH-initiated oxidation of organophosphate flame retardants: a DFT investigation on TCPP. Environ Sci Technol 51(9):5043–5051
83. Li C, Xie HB, Chen JW, Yang XH, Zhang YF, Qiao XL (2014) Predicting gaseous reaction rates of short chain chlorinated paraffins with •OH: overcoming the difficulty in experimental determination. Environ Sci Technol 48(23):13808–13816
84. Fu Z, He N, Zhou P, Liu J, Xie H-B, Yu Q, Ma F, Fu Z, Wang Z, Chen J (2017) Adsorption of nitrobenzene on the surface of ice: a grand canonical monte carlo simulation study. J Phys Chem C 121(29):15746–15755
85. Van Straalen N (2003) Ecotoxicology becomes stress ecology. Environ Sci Technol 37(17):324A–330A
86. Celander MC, Goldstone JV, Denslow ND, Iguchi T, Kille P, Meyerhoff RD, Smith BA, Hutchinson TH, Wheeler JR (2011) Species extrapolation for the 21st century. Environ Toxicol Chem 30(1):52–63
87. Brinkmann M, Schlechtriem C, Reininghaus M, Eichbaum K, Buchinger S, Reifferscheid G, Hollert H, Preuss TG (2016) Cross-species extrapolation of uptake and disposition of neutral organic chemicals in fish using a multispecies physiologically-based toxicokinetic model framework. Environ Sci Technol 50(4):1914–1923
88. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 4(1):1–13
89. Doering JA, Wiseman S, Giesy JP, Hecker M (2018) A cross-species quantitative adverse outcome pathway for activation of the aryl hydrocarbon receptor leading to early life stage mortality in birds and fishes. Environ Sci Technol 52(13):7524–7533
90. Fu Z, Wang Y, Chen J, Wang Z, Wang X (2016) How PBDEs are transformed into dihydroxylated and dioxin metabolites catalyzed by the active center of cytochrome P450s: a DFT study. Environ Sci Technol 50(15):8155–8163
91. Wang X, Wang Y, Chen J, Ma Y, Zhou J, Fu Z (2012) Computational toxicological investigation on the mechanism and pathways of xenobiotics metabolized by cytochrome P450: a case of BDE-47. Environ Sci Technol 46(9):5126–5133
92. Li F, Xie Q, Li X, Li N, Chi P, Chen J, Wang Z, Hao C (2010) Hormone activity of hydroxylated polybrominated diphenyl ethers on human thyroid receptor-β: *in vitro* and in silico investigations. Environ Health Perspect 118(5):602–606

93. Ge L, Chen J, Lin J, Cai X (2009) Light-source-dependent effects of main water constituents on photodegradation of phenicol antibiotics: mechanism and kinetics. Environ Sci Technol 43(9):3101–3107
94. Ge L, Chen J, Wei X, Zhang S, Qiao X, Cai X, Xie Q (2010) Aquatic photochemistry of fluoroquinolone antibiotics: kinetics, pathways, and multivariate effects of main water constituents. Environ Sci Technol 44(7):2400–2405
95. Altenburger R, Walter H, Grote M (2004) What contributes to the combined effect of a complex mixture? Environ Sci Technol 38(23):6353–6362
96. Spurgeon DJ, Jones OAH, Dorne J-LCM, Svendsen C, Swain S, Stuerzenbaum SR (2010) Systems toxicology approaches for understanding the joint effects of environmental chemical mixtures. Sci Total Environ 408(18):3725–3734

Zhongyu Wang received his B.S. degree at Dalian University of Technology in 2012. Then, he decided to develop his vision and techniques for scientific research and became a Ph.D. student of Prof. Dr. Jingwen Chen. His major interest is to review available ideas and innovate new ideas, tools, and schemes for modeling phenomena and solving problems in the field of environmental health science. In 2016, under the supervision of Prof. Chen, Wang published a review article titled "Computational toxicology: Oriented for chemicals risk assessment" in Scientia Sinica Chimica, which helped in introducing then relatively unfamiliar concept of computational toxicology to Chinese researchers. In 2018, Wang published a monograph titled "Environmental computational chemistry and toxicology" together with Prof. Chen, and Dr. Zhiqiang Fu. Fellows may find him staring at the screen of his laptop, working on some Python scripts for bizarre ideas. Besides, he also enjoys listening to music of various genres, appreciating and creating digital arts, etc.

Jingwen Chen is Professor (since 2001) of environmental chemistry and dean for the School of Environmental Science and Technology at Dalian University of Technology. He graduated from the Northeast Normal University, received his B.S. in 1991 and M.S. in 1994. After getting Ph.D. from the Nanjing University in 1997, he performed post-doctoral studies at the German Research Center for Environmental Health as an Alexander von Humboldtian. His research interests concern ecological chemistry and computational toxicology on organic chemicals. He has published over 260 peer-reviewed papers, reviews, book chapters, and three books (two monographs and one textbook). He served as a member of editorial boards of several international (Ecotox. Environ. Safe., J. Enviorn. Sci.) and domestic (e.g., Environ. Chem., Asian J. Ecotox., Mar. Environ. Sci.) journals. In 2013, he won the National Science Fund for Distinguished Young Scholars and was appointed as chair professor of "Cheung Kong Scholars Program" by China Ministry of Education.

Chapter 3
Modelling Simple Toxicity Endpoints: Alerts, (Q)SARs and Beyond

Richard Williams, Martyn Chilton, Donna Macmillan, Alex Cayley, Lilia Fisk and Mukesh Patel

Abstract The correlation of chemical structure with physicochemical and biological data to assess a desired or undesired biological outcome now utilises both qualitative and quantitative structure–activity relationships ((Q)SARs) and advanced computational methods. The adoption of in silico methodologies for predicting toxicity, as decision support tools, is now a common practice in both developmental and regulatory contexts for certain toxicity endpoints. The relative success of these tools has unveiled further challenges relating to interpreting and applying the results of models. These include the concept of what makes a negative prediction and exploring the use of test data to make quantitative predictions. Due to several factors, including the lack of understanding of mechanistic pathways in biological systems, modelling complex endpoints such as organ toxicity brings new challenges. The use of the adverse outcome pathway (AOP) framework as a construct to arrange models and data, to tackle such challenges, is reviewed.

Keywords QSAR · Expert systems · Mutagenicity · Skin sensitisation · Negative predictions · Defined approach · Hepatotoxicity · AOP · MIE

Abbreviations

(Q)SAR	(Quantitative) structure–activity relationship
ADME	Adsorption, distribution, metabolism, and excretion
AOP	Adverse outcome pathway
BSEP	Bile salt export pump
DA	Defined approach
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DPRA	Direct peptide reactivity assay

R. Williams (✉) · M. Chilton · D. Macmillan · A. Cayley · L. Fisk · M. Patel
Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS, UK
e-mail: Richard.Williams@lhasalimited.org

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_3

EC3	Effective concentration to cause a threefold increase in T-cell proliferation
GHS	Globally harmonised system
GPMT	Guinea pig maximisation test
h-CLAT	Human cell line activation test
IATA	Integrated approach to testing and assessment
KE	Key event
kNN	k-Nearest neighbours
LLNA	Local lymph node assay
MIE	Molecular initiating event
MW	Molecular weight
OATP	Organic anion transporting polypeptide
OECD	Organisation for Economic Co-operation and Development
PPAR	Peroxisome proliferator-activated receptor

3.1 Introduction

Structural alerts and quantitative structure–activity relationships (QSARs) have a long history of utility for the qualitative prediction of toxicity. One of the earliest examples is the Ashby–Tennant superstructure, a chemical concatenation of toxicophores associated with (and causal for) mutagenic and carcinogenic activity [1]. Where such toxicophores were found within a chemical, then that chemical could be predicted to be mutagenic and carcinogenic. The success of that early *in papyro/in cerebro* model led to the development of increasingly sophisticated computational systems for predicting toxicity in silico which have sought to replicate human-like reasoning [2, 3] or utilise advanced machine-learning techniques [4, 5].

3.2 Lessons Learnt from Successful Models

Retrospectively, it has become clear that predictions of toxicity from in silico models have often shown the highest levels of acceptance and application where the endpoints being modelled are governed by one (or few) molecular initiating events (MIE) and there are limited absorption, distribution, metabolism, and excretion (ADME) considerations. An MIE describes the interaction between a chemical and a biological target and is linked to a toxicity endpoint via an adverse outcome pathway (AOP) [6]. For endpoints such as skin sensitisation and mutagenicity, toxicity is largely driven by chemical reactivity, leading to the formation of protein or DNA adducts which are the respective MIEs. Although metabolic activation and deactivation can be a precursor to these MIEs, there are relatively few ADME factors to consider. Thus, in these cases, a model where the activity of a wide range of chemicals can be assessed (i.e. a global model) can be generated from a single descriptor set (often chemical fragments) which provides an acceptable simulation of biological reality.

Additionally, data for training mutagenicity and skin sensitisation models comes from standardised assays such as the Ames test and murine local lymph node assay (LLNA) which have OECD test guidelines and thus high levels of reproducibility [7].

As a result, there has been increasing regulatory acceptance of models for simple endpoints. Mutagenicity predictions are now accepted in lieu of data from an Ames test for certain pharmaceutical impurities or plant protection product metabolites [8, 9]. In silico skin sensitisation predictions, especially when used as part of a defined approach (DA), can now be used in place of traditional in vivo LLNA or Guinea pig maximisation tests (GPMT) [10–12]. Predictions from such models are also straightforward to validate, e.g. by running an in vitro assay to confirm or refute the in silico result. This data can then be supplied to the developers to incrementally improve model performance [13].

It is now straightforward (and accepted) to produce qualitative models for simple toxicity endpoints. However, the relative success of these tools has unveiled further challenges relating to interpreting and applying the results of models. These include: how to make negative predictions, moving from qualitative to quantitative predictions, the necessity for expert review of predictions and how to model complex endpoints where these methods are not suitable.

3.3 Making Negative Predictions

The first challenge to be addressed is whether the lack of toxicity can be predicted using structural alerts. These define chemical fragments that are causative of an adverse outcome. When no alerts are found in a compound, is it reasonable to assume that the lack of a positive prediction is enough evidence to make an explicit negative prediction? In this context, two questions need further consideration:

1. Do the existing structural alerts in the appropriate chemical space cover the known mechanisms of toxicity well?
2. Is the adverse outcome driven by a single MIE?

If the answer to both questions is yes, then it can be expected that an absence of structural alerts will be indicative of a lack of toxicity. This has been demonstrated for some endpoints including bacterial in vitro mutagenicity and in vivo skin sensitisation [14]. Both endpoints have been studied for numerous years, resulting in many structural alerts, and both are largely dependent on a reactivity-driven MIE. However, methodologies are also required to assess the reliability of individual predictions, to identify (and justify) whether these can be treated with a higher or lower level of confidence.

Two differing methodologies for assessing the reliability of alert-based negative predictions have previously been investigated for bacterial mutagenicity: an expert knowledge-based approach considered whether a non-alerting chemical had been purposely excluded from the scope of a structural alert, and a data-driven approach

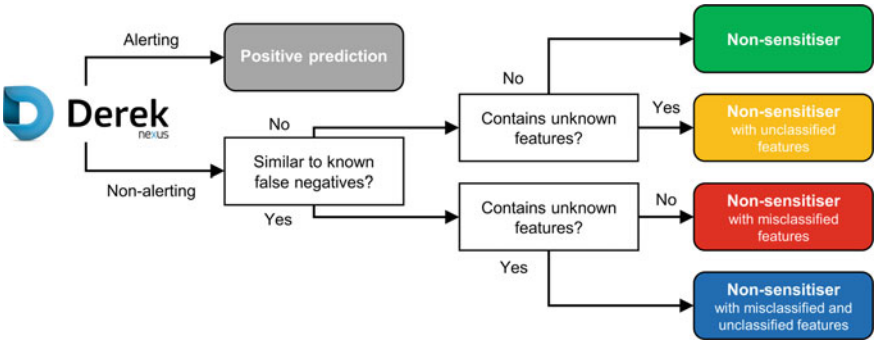

Fig. 3.1 Outline of the negative prediction methodology employed by an expert knowledge-based system

that involved comparing the fragments present in a non-alerting chemical to a large dataset of other chemicals with known activities [15].

The first method was based on the expectation that chemicals which have purposely been excluded from a structural alert based on a deactivating feature were more likely to be inactive than those that simply contain no substructures that match an alert, as more is known about the former than the latter. In practice, this method was a poor indicator of the reliability of a negative prediction. The border between active and inactive compounds was rarely unambiguous due to limitations in both the quality and quantity of the available data. In contrast, the second method was successful in highlighting cases where the negative prediction could be treated with either a higher or lower degree of confidence. This was achieved by answering two questions of a non-alerting chemical that an expert user might also ask (Fig. 3.1):

1. Is the chemical similar to known active chemicals that the expert system predicts incorrectly (i.e. false negatives)?
2. Does the chemical contain any fragments that the expert system has not seen before?

Both questions are answered by comparing the non-alerting chemical to a large, curated dataset of chemicals with known activity data collected from the public domain. When a chemical contains a fragment that is found exclusively in false negative compounds in the dataset, this is flagged to the user as a misclassified feature. Likewise, where a chemical contains a fragment that is not present at all in the dataset, it is highlighted as an unclassified feature. The presence of either type of feature is likely to reduce the confidence a user has in the negative in silico outcome. Misclassified features are expected to decrease the accuracy of the prediction as the non-alerting chemical is similar to other active compounds that the expert system predicts poorly. Unclassified features are expected to increase the uncertainty around the prediction as the non-alerting chemical resides in an unstudied area of chemical space.

These expectations were borne out by the data when the methodology (as implemented in Derek Nexus [16]) was tested against several proprietary datasets, for both the mutagenicity and skin sensitisation endpoints [15, 17]. The general trend observed was that misclassified and unclassified features occurred relatively infrequently. When misclassified features did occur, they tended to result in a drop in negative predictivity compared to the entire dataset. This suggests that the identification of chemical fragments that were found elsewhere in known false negatives was a useful similarity metric and worth highlighting. The presence of unclassified features, however, did not show the same trend in that the negative predictivity tended to remain high. However, the observed negative predictivity values across several proprietary mutagenicity datasets displayed a large interquartile range, indicating the greater variability in the accuracy of predictions for chemicals containing novel fragments.

Across the proprietary test sets the negative predictivity for chemicals containing either misclassified or unclassified features typically remained higher than the prevalence of inactive chemicals in the datasets, indicating that their presence was not a definitive flag for activity. Rather, they should be interpreted as weak arguments against the negative prediction, indicating greater uncertainty in the predictions. It is possible that additional expert review of the in silico negative predictions could resolve some of this additional uncertainty. As such, the identification of misclassified and/or unclassified features serves a secondary purpose, by providing the user with fragments within the chemical structure to focus on as a starting point for any further assessment.

The endpoints that have been described so far are driven by a single reactivity-based MIE, reflecting the fact that to make negative predictions, expert systems have leveraged the power of fragment-based approaches to model chemical reactivity. The challenges which are expected when making negative predictions for more complex endpoints (e.g. carcinogenicity or hepatotoxicity) are: the need to use more relevant descriptors to model non-reactivity-based MIEs and creation of multiple models for each individual MIE, to ensure that a chemical is not expected to initiate any of the multiple pathways that could lead to the adverse outcome in question.

3.4 Moving to Quantitative Predictions and Weight of Evidence Approaches

Another challenge that structural alerts do not address is the need to make quantitative toxicity predictions. However, they can be used to group chemicals into categories which react through the same toxicity mechanism, which can then provide a starting point for making quantitative read across predictions within these categories. In computational terminology these predictions can be described as k-nearest neighbour (kNN) models, although in practice this is an example of how SARs and read across can be used together to make interpretable quantitative toxicity predictions.

Canipa et al. [18] described a kNN model which was developed to make quantitative skin sensitisation predictions. Initially, the query chemical fires a skin sensitisation structural alert, and chemicals in the dataset firing the same alert are considered as nearest neighbours. These nearest neighbours are then arranged by Tanimoto similarity generated by a radial fingerprinting method. The minimum number of neighbours required is 3; otherwise, no prediction is given, and the most similar neighbours, up to tenth place, are considered. The predicted EC3 value is the weighted average of all the valid neighbours (Eq. 3.1).

Equation 3.1: EC3 prediction equation using weighted average of nearest neighbours as calculated by Tanimoto scoring.

$$\frac{\mathrm{MW}}{\mathrm{EC3}_{\mathrm{prediction}}} = \frac{\sum \left(\mathrm{Tanimoto} * \frac{\mathrm{MW}}{\mathrm{EC3}_{\mathrm{Nearest\ Neighbour}}} \right)}{\sum \mathrm{Tanimoto}} \tag{3.1}$$

The model was assessed using a public test set of 45 chemicals, as well as a proprietary test set containing 103 chemicals donated by Lhasa Limited members (Table 3.1). The model predicts relatively well for both, although the inherent variability of the LLNA limits the predictive capacity of this model. As more LLNA data is added to the prediction dataset, the accuracy is expected to improve. When the prediction is incorrect, it tends towards conservatism, overpredicting rather than underpredicting which is more protective of human health.

Another key consideration for in silico predictions is how best to combine their output with data generated from *in chemico* and in vitro tests, especially for those endpoints where there has been a considerable drive to use non-animal alternatives such as skin sensitisation. Although several *in chemico* and in vitro assays have been developed to measure individual key events (KEs) in the skin sensitisation AOP and are accepted by the OECD [19–21], it is generally recognised that a single assay is not an adequate replacement for the in vivo assays. Instead, it has been suggested that using multiple information sources (e.g. physicochemical properties, read across, in silico expert/(Q)SAR predictions, *in chemico* and in vitro tests, historical in vivo data) in combination, in either a DA or an integrated approach to testing and assessment (IATA), is a more reliable way to predict skin sensitisation potential [22].

For example, using structural alerts, combined with negative predictions and potency predictions, alongside *in chemico*/in vitro assays (DPRA, KeratinoSens™, LuSens, h-CLAT, U-SENS™) a DA was developed, built on previous work [23]. Known limitations and applicability domain knowledge were used to de-prioritise less applicable assay(s)/in silico outcome(s) and prioritise more appropriate information sources. The types of information that are available in an expert system alongside the simple presence/absence of structural alerts include: the likelihood of any structural alert-based positive prediction; the uncertainty around any negative prediction based on the lack of structural alerts; whether the chemical is likely to require metabolism to show sensitisation potential; how lipophilic the chemical is; and information on the exact nature of the biological nucleophile that causes the MIE between the chemical and the human body. All these considerations can help a user

Table 3.1 Performance of EC3 prediction model when compared against an internal test set of 45 chemicals, and an external proprietary test set containing 103 chemicals donated by Lhasa Limited members

	ECETOC classification			GHS classification		
	Overpredicted (%)	Correctly predicted (%)	Underpredicted (%)	Overpredicted (%)	Correctly predicted (%)	Underpredicted (%)
Public test set	33.3	51.1	15.6	24.4	64.4	11.1
Proprietary test set	45.6	36.0	18.4	25.2	64.1	10.7

Table 3.2 Summary of exclusion criteria for the defined approach

Exclusion criteria used in the defined approach		Information source applicable for defined approach?			
		Derek	MIE/KE1	KE2	KE3
Derek likelihood	Certain, probable, plausible, doubted, improbable	✓			
	Equivocal	✗			
Derek negative prediction	Non-sensitiser	✓			
	Non-sensitiser containing misclassified feature(s)	✗			
	Non-sensitiser containing unclassified feature(s)	✗			
Metabolic activation	Prohapten	✓	✗	✓	✓
Lipophilicity	>3.5	✓	✓	✓	✗
	>5	✓	✓	✗	✗
Lysine-reactivity	Exclusive	✓	✓	✗	✓

MIE/KE1 = molecular initiating event/key event 1 (hapten binding). KE2 = key event 2 (keratinocytes activation). KE3 = key event 3 (activation of dendritic cells). ✓ = information source is prioritised in defined approach. ✗ = information source is de-prioritised in defined approach

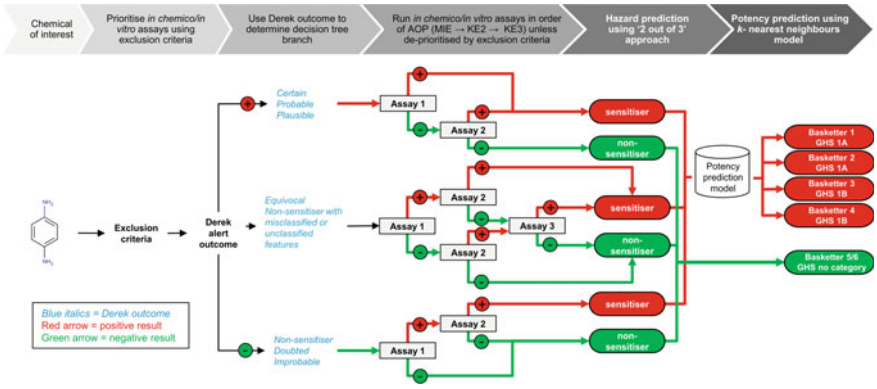

Fig. 3.2 Defined approach decision tree uses exclusion criteria to de-prioritise in chemico/in vitro assays and then uses the Derek outcome to determine which branch of the tree to follow. Between 1 and 3 in chemico/in vitro assays are then run in order of the AOP (MIE → KE2 → KE3) unless de-prioritised in the previous step and the outcome(s) used to assign a hazard classification (sensitiser/non-sensitiser) and predict the potency category (Basketter 1-4, 5/6) and GHS classification

to prioritise the information sources that are believed to be most informative for the chemical in question (Table 3.2).

After a chemical has been assessed and any relevant exclusion criteria considered, the results from prioritised assays were used in a 2 out of 3 approach (run in order of

KEs in the AOP unless they were de-prioritised by the exclusion criteria) to predict the hazard (sensitiser/non-sensitiser). The potency of chemicals assigned as sensitisers is then predicted using a similar mechanistic read across process within a structural alert as described previously, except now considering both known human skin and mouse skin sensitisation data for analogue chemicals (Fig. 3.2).

The DA accurately predicted skin sensitisation hazard when compared against both LLNA data and human data (85% DA vs. LLNA; 86% DA vs. human) indicating that the DA can predict both species equally well. Furthermore, the DA is more predictive of human skin sensitisation (86% accuracy—DA vs. human) than the LLNA (81% accuracy—LLNA vs. human) for the dataset analysed (Table 3.3).

The DA correctly predicted the Basketter potency category (5 categories: 1–4, 5/6) for 59% and the GHS classification (three categories: 1A, 1B, no category) for 73% of the evaluation dataset when compared to LLNA data. When compared against human data, the DA correctly predicted the Basketter potency category for 68% and the GHS classification for 76% of the dataset, respectively (Table 3.4). The DA was also shown to predict both classification methods for human data more accurately than the LLNA (DA vs. human, 68% Basketter and 76% GHS; LLNA vs. human, 54% Basketter and 65% GHS).

3.5 Enabling Expert Review

Whilst in silico tools can provide accurate predictions in isolation [24, 25], expert review adds significant value to the accuracy of the conclusions drawn [26–29]. There has been considerable effort to delineate the types of information that should be considered when carrying out expert review of in silico predictions, particularly for the endpoint of mutagenicity [24, 30–33].

Knowledge of inadequacies in the test system being modelled as well as the arguments associated with the data used to generate the prediction should be considered when carrying out expert review. Analysis of the results presented, bearing in mind strengths and limitations associated with different modelling techniques, is also an important factor in the assessment of the predictions generated.

Therefore, it is important that any prediction systems employed in this context provide enough detailed information about how the prediction was derived so that they can be probed by the expert user to support the overall decision they make. Some general considerations which should be made during the expert review process and the information that should be provided by the predictive system to allow for this analysis are:

1. Limitations of the test being modelled
2. Relevance and adequacy of the data used to make a prediction
3. Similarity of the query compound to the compounds used to make the prediction
4. Coverage of any potential toxicophores
5. Causality of any toxicophore identified

Table 3.3 Skin sensitisation hazard predictivity of individual test methods and the defined approach compared to both LLNA and human reference data

Test method	LLNA				Human			
	n	Accuracy (%)	Sensitivity (%)	Specificity (%)	*n*	Accuracy (%)	Sensitivity (%)	Specificity (%)
LLNA	–	–	–	–	108	80.6	88.0	63.6
DPRA	192	75.5	77.4	71.2	100	80.6	93.3	51.5
KeratinoSens™	184	73.9	77.6	64.0	99	84.0	84.3	83.3
LuSens	77	75.3	77.1	72.4	60	82.8	81.7	85.7
h-CLAT	163	74.2	80.3	58.7	95	81.7	78.4	87.0
U-SENS™	123	82.9	92.3	56.3	91	81.1	88.6	60.0
Derek	210	83.3	92.6	61.3	108	80.6	88.0	63.6
Defined approach	194	84.5	88.1	76.7	102	86.3	88.7	80.6

Table 3.4 Skin sensitisation potency predictivity of the defined approach compared to both LLNA and human reference data

Benchmark data	Basketter potency category				GHS classification			
	n	Overpredicted (%)	Correctly predicted (%)	Underpredicted (%)	*n*	Overpredicted (%)	Correctly predicted (%)	Underpredicted (%)
DA vs. LLNA	174	20.1	58.6	21.3	174	12.1	73.0	14.9
DA vs. human	79	13.9	68.4	17.7	79	10.1	75.9	13.9
LLNA vs. human	89	24.7	53.9	21.3	79	20.2	65.2	14.6

6. Generality of any toxicophore identified
7. Limitations of the software being used.

If a model is being generated based on data coming from an in vitro or in vivo assay linked to an ultimate toxicity endpoint, the limitations of the assay should be considered either as part of the model building process or, more likely, during expert review [32, 34, 35]. For example, there are a minority of compound classes for which the (in vitro) Ames mutagenicity assay may not reflect the true DNA reactivity and consequently the hazard may be over- or underestimated. Models produced from these data will not, therefore, reflect the true (in vivo) carcinogenicity hazard caused by the compounds. Amberg et al. have highlighted the fact that the Ames test may over- or underestimate the mutagenicity of compounds containing the acid halide group depending on the solvent chosen to carry out the test [32, 36]. Water would hydrolyse the acid halide deactivating it to a carboxylic acid, whereas dimethyl sulphoxide (DMSO) would react directly with the acid halide producing halodimethyl sulphides which then act as the DNA-reactive species. To get accurate results for this compound class, the compounds should be tested in an inert solvent such as acetonitrile.

The in silico prediction should be transparent enough that the user can interrogate the data on which the prediction is based. It is important that the data supporting each prediction can be probed to the lowest captured level, to allow for adequate assessment of the weighting of this evidence. Any inadequacies in the testing protocol of the data being used to make the prediction should be noted during expert review [33]. This is particularly important for data points which are pivotal for the prediction being made, and as a consequence the in silico prediction system should identify those key compounds used to make the prediction and their relative weighting.

Many in silico prediction systems account for the ability of the model to make accurate predictions based on the similarity of the predicted structure to the training set chemicals, by employing an applicability domain and/or confidence metrics. In some cases, these machine-generated boundaries may not be as good as a human in assessing the relevance of the data used, and the adequacy of the extrapolation should always be assessed by the expert as part of the review process [33]. Again, this requires that the data supporting the derivation of the model and the training set compounds should be provided, with the prediction, for assessment.

Where a potential toxicophore has been identified and a negative prediction has been made, it is important to determine whether the toxicophore has been assessed adequately by the model and that the factors negating it are appropriate [30]. If the toxicophore has been considered and the negating factors are acceptable, then the prediction can be upheld. However, if the toxicophore was not considered or the negating factors are not relevant, then the prediction may have to be overturned. As a consequence, the in silico prediction should provide some information relating to the chemical space covered and when a potential toxicophore may not be adequately covered (e.g. the misclassified and unclassified features in Derek Nexus mutagenicity and skin sensitisation predictions, and the overturned hypotheses in Sarah Nexus [37]). Where structural alerts are being used to make a prediction, it is also useful to

provide information and a rationale behind any exclusions along with the associated alert.

There are some common limitations with statistical approaches which should be considered during expert review where such a system is being used. These predictions usually deal with association and do not assess causation. As a result, any positive prediction made by the system should be checked with a view towards assessing the likelihood that the feature identified in the query compound is the one likely to be causing the positive result in the compounds used to make the prediction and that activity of these compounds cannot be attributed to other features not present in the predicted structure [24, 30, 33]. The in silico prediction should, therefore, provide information about the toxicophore identified in the query structure as well as the training set examples which also possess this feature in order to allow for analysis of whether the feature attributed to toxicity is likely to be causative.

Many in silico prediction systems employ SARs based on structural alerts. It is important to note that these alerts can sometimes be very general and, as a result, specific substitution patterns representing steric or electronic factors which may negate the hazard caused by the toxicophore may be missed [24]. For example, structural alerts for mutagenicity based solely on broad toxicophores, such as those proposed by Ashby and Tennant [1], will fail to take into consideration mitigating factors that mean many compounds belonging to these structural classes do not show mutagenic activity [38]. For these types of alerts, expert review should be carried out to assess steric or electronic factors which may influence the activity of the training set compounds, and the query compound. The scope of any alert activated should be provided to allow the user to assess how broad the coverage of the alert is, along with a rationale for the scope of the alert.

Some consideration should also be given to limitations of the software being used to make the predictions. For example, in order that an accurate prediction be made for a query compound, it is important that the structure is represented in the same way as those used to build the model. Standardisation of the structures should, therefore, be carried out in the same way in both cases to produce accurate predictions. If this standardisation is not carried out automatically by the system, then it must be done manually before making the prediction.

3.6 Modelling Complex Endpoints

The prediction of outcomes for complex endpoints, following oral exposure to a chemical, presents new and different challenges. Additionally, the influence of ADME factors requires a deeper consideration, which can be illustrated by several questions, e.g.

1. Is the chemical stable in the gastrointestinal tract?
2. Will the chemical be absorbed?
3. Will first-pass metabolism detoxify the chemical?
4. Will the chemical be metabolised to a reactive species?

5. Will the chemical (or any metabolite) generate a molecular initiating event?
6. Will the chemical show high levels of protein plasma binding?
7. Will the chemical be ‘quickly’ cleared from the body?

Each of these processes will have a unique structure–activity relationship (SAR—which may be increasingly granular, e.g. interactions with specific transporters), which is masked by the apical endpoint data. For example, if a chemical is classified as ‘not hepatotoxic’ based on experimental data—is this because the chemical does not ‘activate’ an MIE, or because it is not absorbed or quickly excreted? From this, it becomes obvious that summary data from apical endpoints is an oversimplified summary of the biological reality. Additionally, factors that promote activity in some of the component models could curtail activity in another. For example, the addition of a methyl group to an arene-containing chemical creates a potential ‘metabolic hook’ which can lead to the generation of reactive metabolites, whereas the same methyl group attached close to a reactive moiety can deactivate this by preventing access to a biological nucleophile, such as DNA, through steric blocking. Moreover, to model all these processes to the fine detail required to make accurate predictions requires much data. If in silico models of mutagenicity are now based on greater than 10,000 chemicals, how much data would be required to create broadly applicable models of more complex apical endpoints to a similar level of precision? This is without considering the increasing heterogeneity within the data when moving away from reproducible standardised assays.

Hence, to accurately model multifactorial endpoints such as hepatotoxicity, ideally the multiple SARs that relate to the multiple processes leading to the high-level outcomes to activity or inactivity need to be accounted for. This would enable persuasive arguments to be made; e.g., a chemical is predicted to be hepatotoxic (due to cholestasis) because it blocks the bile salt export pump (BSEP) channel as well as being straightforward to validate (e.g. in this case by running a BSEP assay).

Some of the complication inherent in complex global models can be rectified by using local models. However, local models are designed to only predict the activity of a narrow series of congeneric chemicals and lack general applicability. A knowledge base containing a series of structural alerts is, in effect, a collection of local models which obviates the problem of applicability. Structural alerts can be designed to be activated by a given chemical class operating through a single MIE, taking account of heterogeneity in the training data [39]. Thus, one method which has been employed to model in vivo endpoints is through extrapolation of alerts from a related in vitro endpoint that share a common MIE. This has been demonstrated with some success for the chromosome damage endpoint [40], though the resulting alerts still carry the complication of additional ADME factors.

With respect to creating SARs for complex endpoints, there are several models available which use available in vivo data to predict potential liabilities for new chemicals. In many of these cases, the toxicophores were identified either statistically [41] or by an expert [42]. In other cases, automated clustering followed by human evaluation where the relevance of statistically determined toxicophores was assessed by an expert and/or by searching public literature for mechanistic evidence has been

presented [43–45]. Despite the relativity high sensitivity of these models for the training sets, their performance for independent test sets was often low.

QSAR models using histopathology data have been published for liver and nephrotoxicity [46]. However, the performance of these models (with initially high sensitivity and specificity against external test sets) dropped when subsequently validated with new data [46]. This and earlier examples indicate that currently global QSAR models for complex endpoints have limited coverage of relevant chemical space.

As knowledge of different mechanisms leading to organ toxicity increases, it is increasingly being presented in the form of AOPs [47]. There are several in vitro assays that measure MIEs or KEs for some endpoints (e.g. BSEP inhibition, mitochondrial toxicity, etc.). The output of these assays in the form of activity towards a specific target can be used as indicators of whether events on an AOP are likely or not. A battery of the assays describing a selection of the key mechanisms in combination with daily dose (and covalent binding to proteins) has been shown to be useful in identification of potential hepatotoxic liabilities of compounds [48]. However, this approach somewhat lacks comprehensive coverage for an AOP, as it is limited to a selection of mechanisms. As the array of mechanistically derived tests develops, it should lead to a wider coverage of MIEs and KEs for each AOP.

Several publications describe approaches for modelling various MIE models, where the physicochemical properties of chemicals [49] or structural information in combination with multitarget bioactivity [50] are used. The bioactivity data for the modelling is obtained from in vitro assays, that can be used to derive SARs or for machine-learning algorithms. Different approaches for using this data have been adopted: whilst in some cases all available data was utilised, in other cases only a subset of assays that were assumed to be relevant for the adverse outcome were taken into consideration. For example, from the available in vitro high-throughput data generated in the ToxCast project, data from only nine MIE endpoints (peroxisome proliferator-activated receptor (PPAR) alpha, PPAR beta, PPAR gamma, constitutive androstane receptor, pregnane X receptor, aryl hydrocarbon receptor, liver X receptor, nuclear factor (erythroid-derived 2)-like 2, farnesoid X receptor) was used to develop random forest models for MIEs relevant to hepatic steatosis [51]. Similarly, when creating a liver cholestasis model, the authors utilised existing knowledge and selected data for the inhibition of hepatic transporters (BSEP, breast cancer resistance protein, P-glycoprotein, organic anion transporting polypeptide(OATP)1B1 and OATP1B3) previously shown to disrupt the bile flow [52]. In other cases, when data from a high number of in vitro assays was applied, the predictive capacity of almost all models (regardless of the algorithm) was improved when combined with structural information about the chemical itself [53]. This was exemplified for the hepatotoxicity endpoint [54], with a similar outcome for other organ toxicities [55].

Other types of data, such as those from gene expression studies, are also becoming more accessible and have been used for predicting organ toxicity in a systems biology approach [56], or to generate machine-learning models [57].

Utilising in vitro assays data for building (Q)SAR models enables the prediction of possible liabilities for a new chemical; these global models for individual MIE/KEs or apical endpoints are likely to ‘catch’ potential intrinsic toxicities of compounds and

can be used for prioritisation during drug development. It is important, however, to understand the coverage of the chemical space of each model, to ensure that generated predictions are reliable. To enable the prediction of idiosyncratic toxic effects, the variability amongst individual protein/enzyme levels, expression and activity in the target population should be considered [58].

3.7 Conclusion and Future Directions

Structural alerts provide much more than a simple binary prediction of toxicity hazard. Firstly, they can also be used to form mechanistic chemical categories within which quantitative read across predictions can be made. Secondly, the additional expert knowledge housed within an expert system can be used to prioritise *in chemico/*in vitro testing by considering the applicability of each individual information source. Thirdly, the generated data can then be combined with all of the in silico information in a weight of evidence approach to arrive at a final conclusion about the hazard potential and/or potency of a chemical of interest.

The use of in silico methodologies as decision support tools is now common practice for certain toxicity endpoints. The relative success of these tools has unveiled further challenges relating to interpreting and applying the results of models. The key issues going forward for all these models, with respect to the regulatory context, are how to guide the appropriate use of these techniques, and to provide an appropriate level of interpretation for the results they produce to instil familiarity and confidence in their use. The use of the AOP framework as a construct to arrange models and data provides one way to tackle such challenges, whilst helping to focus the development of new tests to support the mechanistic requirements.

References

1. Ashby J, Tennant RW (1988) Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. Mutat Res 204:17–115
2. Marchant C et al (2008) In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic. Toxicol Mech Methods 18:177–187
3. Barber C et al (2015) Establishing best practise in the application of expert review of mutagenicity under ICH M7. Regul Toxicol Pharmacol 73:367–377
4. Hanser T et al (2014) Self organising hypothesis networks: a new approach for representing and structuring SAR knowledge. J Chemoinf 6:21
5. Barber C et al (2017) Distinguishing between expert and statistical systems for application under ICH M7. Regul Toxicol Pharmacol 84:124–130
6. Ankley GT et al (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730–741
7. OECD (2010) Test No. 429: skin sensitisation: local lymph node assay. https://doi.org/10.1787/9789264071100-en. Accessed 28 Aug 2018

8. ICH (2017) Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. M7(R1). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_R1_Addendum_Step_4_31Mar2017.pdf. Accessed 9 Sept 2018
9. EFSA (2016) Guidance on the establishment of the residue definition for dietary risk assessment. EFSA J 14: 180. https://doi.org/10.2903/j.efsa.2016.4549. Accessed 9 Sept 2018
10. OECD (2017) Guidance document on the reporting of defined approaches and individual information sources to be used within Integrated Approaches to Testing and Assessment (IATA) for skin sensitisation. In: OECD series on testing and assessment, No. 256. OECD Publishing, Paris. https://doi.org/10.1787/9789264279285-en. Accessed 28 Aug 2018
11. European Union (2009) Regulation (EC) No 1223/2009 of the European Parliament and of the council of 30 November 2009 on cosmetic products. http://data.europa.eu/eli/reg/2009/1223/oj. Accessed 28 Aug 2018
12. European Union (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. http://data.europa.eu/eli/reg/2006/1907/2018-05-09. Accessed 28 Aug 2018
13. Elder DP et al (2015) Mutagenic impurities: precompetitive/competitive collaborative and data sharing initiatives. Org Process Res Dev 19:1486–1494
14. Judson PN et al (2013) Assessing confidence in predictions made by knowledge-based systems. Toxicol Res 2:70–79
15. Williams RV et al (2016) It's difficult, but important, to make negative predictions. Regul Toxicol Pharmacol 76:79–86
16. Derek Nexus v6.0 (Lhasa Limited). https://www.lhasalimited.org/products/derek-nexus.htm. Accessed 28 Aug 2018
17. Chilton ML et al (2018) Making reliable negative predictions of human skin sensitisation using an in silico fragmentation approach. Regul Toxicol Pharmacol 95:227–235
18. Canipa SJ et al (2016) A quantitative in silico model for predicting skin sensitization using a nearest neighbours approach within expert-derived structure-activity alert spaces. J Appl Toxicol 37:985–995
19. OECD (2015) Test No. 442C: In chemico skin sensitisation: Direct Peptide Reactivity Assay (DPRA). https://doi.org/10.1787/9789264229709-en. Accessed 28 Aug 2018
20. OECD (2018) Key event based test guideline 442D: in vitro skin sensitisation assays addressing the AOP key event on keratinocyte activation. https://doi.org/10.1787/9789264229822-en. Accessed 28 Aug 2018
21. OECD (2018) Key event based test guideline 442E: In vitro skin sensitisation assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitisation. https://doi.org/10.1787/9789264264359-en. Accessed 28 Aug 2018
22. Kleinstreuer NC et al (2018) Non-animal methods to predict skin sensitization (II): an assessment of defined approaches. Crit Rev Toxicol 48:359–374
23. Macmillan DS et al (2016) Predicting skin sensitisation using a decision tree integrated testing strategy with an in silico model and in chemico/in vitro assays. Regul Toxicol Pharmacol 76:30–38
24. Barber C et al (2016) Establishing best practise in the application of expert review of mutagenicity under ICH M7. Regul Toxicol Pharmacol 73:367–377
25. Verheyen GR et al (2017) Evaluation of in silico tools to predict the skin sensitization potential of chemicals. SAR QSAR Environ Res 28:59–73
26. Dobo KL et al (2012) In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry survey. Regul Toxicol Pharmacol 62:449–455

27. Kruhlak NL et al (2012) (Q)SAR modeling and safety assessment in regulatory review. Clin Pharmacol Ther 91:529–534
28. Naven RT et al (2012) Latest advances in computational genotoxicity prediction. Expert Opin Drug Metab Toxicol 8:1579–1587
29. Sutter A et al (2013) Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regul Toxicol Pharmacol 67:39–52
30. Powley MW (2015) (Q)SAR assessments of potentially mutagenic impurities: a regulatory perspective on the utility of expert knowledge and data submission. Regul Toxicol Pharmacol 71:295–300
31. Greene N et al (2015) A practical application of two in silico systems for identification of potentially mutagenic impurities. Regul Toxicol Pharmacol 72:335–349
32. Amberg A et al (2016) Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses. Regul Toxicol Pharmacol 77:13–24
33. Myatt GJ et al (2018) In silico toxicology protocols. Regul Toxicol Pharmacol 96:1–17
34. Roberts DW et al (2016) Chemical applicability domain of the Local Lymph Node Assay (LLNA) for skin sensitisation potency. Part 3. Apparent discrepancies between LLNA and GPMT sensitisation potential: false positives or differences in sensitivity? Regul Toxicol Pharmacol 80:260–267
35. Honda H et al (2016) Modified Ames test using a strain expressing human sulfotransferase 1C2 to assess the mutagenicity of methyleugenol. Genes Environ. https://doi.org/10.1186/s41021-016-0028-x
36. Amberg A et al (2015) Do carboxylic/sulfonic acid halides really present a mutagenic and carcinogenic risk as impurities in final drug products? Org Process Res Dev 19:1495–1506
37. Sarah Nexus v3.0 (Lhasa Limited). https://www.lhasalimited.org/products/sarah-nexus.htm. Accessed 28 Aug 2018
38. Myden A et al (2017) Utility of published DNA reactivity alerts. Regul Toxicol Pharmacol 88:77–86
39. Faulkner D et al (2017) Tools for green molecular design to reduce toxicological risk. In: Johnson DE, Richardson RJ (eds) Computational systems pharmacology and toxicology, Royal Society of Chemistry, London, Chapter 3, p 36–59
40. Canipa S et al (2015) Using in vitro structural alerts for chromosome damage to predict in vivo activity and direct future testing. Mutagenesis 31:17–25
41. Egan WJ et al (2004) In silico prediction of drug safety: despite progress there is abundant room for improvement. Drug Discov Today Technol 1:381–387
42. Greene N et al (2010) Developing structure-activity relationships for the prediction of hepatotoxicity. Chem Res Toxicol 23:1215–1222
43. Hewitt M et al (2013) Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action. Crit Rev Toxicol 43:537–558
44. Pizzo F et al (2016) A new structure-activity relationship (SAR) model for predicting drug-induced liver injury, based on statistical and expert-based structural alerts. Front Pharmacol. https://doi.org/10.3389/fphar.2016.00442
45. Liu R et al (2015) Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries. J Cheminf. https://doi.org/10.1186/s13321-015-0053-y
46. Myshkin E et al (2012) Prediction of organ toxicity endpoints by QSAR modeling based on precise chemical-histopathology annotations. Chem Biol Drug Des 80:406–416
47. OECD (2012) AOP knowledgebase. https://aopkb.oecd.org/. Accessed 28 Aug 2018
48. Thompson RA et al (2016) Reactive metabolites: current and emerging risk and hazard assessments. Chem Res Toxicol 29:505–533
49. Warner DJ et al (2012) Mitigating the inhibition of human bile salt export pump by drugs: opportunities provided by physicochemical property modulation, in silico modeling, and structural modification. Drug Metab Dispos 40:2332–2341. https://doi.org/10.1124/dmd.112.047068
50. Qiu T et al (2018) Finding the molecular scaffold of nuclear receptor inhibitors through high-throughput screening based on proteochemometric modelling. J Cheminf. https://doi.org/10.1186/s13321-018-0275-x

51. Gadaleta D et al (2018) QSAR modeling of ToxCast assays relevant to the molecular initiating events of AOPs leading to hepatic steatosis. J Chem Inf Model 58:1501–1517
52. Kotsampasakou E, Ecker GF (2017) Predicting drug-induced cholestasis with the help of hepatic transporters-an in silico modeling approach. J Chem Inf Model 57:608–615
53. Altamira LLC, Molecular Networks GmbH (2013) ChemoTyper. https://chemotyper.org/. Accessed 28 Aug 2018
54. Liu J et al (2015) Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem Res Toxicol 28:738–751
55. Liu J et al (2017) Predicting organ toxicity using in vitro bioactivity data and chemical structure. Chem Res Toxicol 30:2046–2059
56. Carbonell P et al (2017) Hepatotoxicity prediction by systems biology modeling of disturbed metabolic pathways using gene expression data. Altex 34:219–234
57. Kim J, Shin M (2014) An integrative model of multi-organ drug-induced toxicity prediction using gene-expression data. BMC Bioinform. https://doi.org/10.1186/1471-2105-15-s16-s2
58. Alyass A et al (2015) From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics. https://doi.org/10.1186/s12920-015-0108-y

Richard Williams is Chief Scientific Officer at Lhasa Limited. He is an organic chemist by training, having graduated with a Ph.D. in organic chemistry from Swansea University in 2000, and is a chemical toxicologist through practice. He joined Lhasa Limited in 2004, and his early work focused on the prediction of genotoxicity (mutagenicity, chromosome damage) through the derivation of structural alerts, in collaboration with world leading genetic toxicologists (Drs. Hayashi and Honma). Subsequent work led to a broadening of his endpoint knowledge (e.g. carcinogenicity, skin sensitisation, phototoxicity), and he has also led development work around negative predictions and ICH M7 workflows. Despite no longer being directly involved in alert writing, he retains his interest in genetic toxicology by serving as Secretary of the UK's Industrial Genotoxicology Group (IGG) and has also been part of the International Workgroups on Genotoxicity Testing (IWGT).

Chapter 4
Matrix and Tensor Factorization Methods for Toxicogenomic Modeling and Prediction

Suleiman A. Khan, Tero Aittokallio, Andreas Scherer, Roland Grafström and Pekka Kohonen

Abstract Prediction of unexpected, toxic effects of compounds is a key challenge in computational toxicology. Machine learning-based toxicogenomic modeling opens up a systematic means for genomics-driven prediction of toxicity, which has the potential also to unravel novel mechanistic processes that can help to identify underlying links between the molecular makeup of the cells and their toxicological outcomes. This chapter describes the recent big data and machine learning-driven computational methods and tools that enable one to address these key challenges in computational toxicogenomics, with a particular focus on matrix and tensor factorization approaches. Here we describe these approaches by using exemplary application of a data set comprising over 2.5×10^8 data points and 1300 compounds, with the aim of explaining dose-dependent cytotoxic effects by identifying hidden factors/patterns captured in transcriptomics data with links to structural fingerprints of the compounds. Together transcriptomics and structural data are able to predict pathological states in liver and drug toxicity.

Keywords Machine learning · Group factor analysis · Tensor factorization · Bayesian modeling · Drug sensitivity · Connectivity Map · NCI-60 · Gene expression · Biomarkers

S. A. Khan (✉) · T. Aittokallio · A. Scherer
Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
e-mail: khan.suleiman@gmail.com; suleiman.khan@helsinki.fi

A. Scherer
e-mail: andreasscherer@eatris.eu

T. Aittokallio
Department of Mathematics and Statistics, University of Turku, Turku, Finland

R. Grafström · P. Kohonen
Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Predictomics AB, Stockholm, Sweden

Misvik Biology Oy, Turku, Finland

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_4

Abbreviations

ARD	Automatic Relevance Determination
CCLE	Cancer Cell Line Encyclopedia
CMap	Connectivity Map
CP	Canonical decomposition and Parallel factor analysis
DILI	Drug-induced liver injury
FA	Factor Analysis
FCFP	Functional Connectivity Fingerprints
FDA	Food and Drug Administration
GFA	Group Factor Analysis
GI50	50% Growth Inhibition
IPA	Ingenuity Pathway Analysis
LC50	50% Lethal Concentration
LDA	Latent Dirichlet Allocation
LINCS	Library of Integrated Network-Based Cellular Signatures
MF	Matrix Factorization
MoA	Mode of action
MTF	Multi-tensor Factorization
NCI	National Cancer Institute
PCA	Principal Component Analysis
PTGS	Predictive Toxicogenomics Space
QSAR	Quantitative Structure–Activity Relationship
TF	Tensor Factorization
TGI	Total Growth Inhibition

4.1 Introduction

Cellular responses to drugs and other chemical compounds are increasingly being measured at multiple levels of detail and resolution. For instance, ex vivo toxicity measurements summarize the phenotypic responses in human primary cells [1, 2], while profiling of genome-wide transcriptomic responses opens up a system-level view to the compounds' mode-of-action (MoA) mechanisms. *The study of relationships between genome-wide genomic or molecular responses of the cells to exposure to substances and the corresponding toxicological outcomes is referred to as toxicogenomics*. Understanding these complex relationships can not only identify the molecular mechanisms behind toxicity but also suggest ways to avoid toxic effects in medical or other applications [3–7]. Toxicogenomics may be especially pertinent for analyzing data from cellular assays, and for reducing and eventually replacing the use of animal experiments for toxicity testing during drug development, also referred to as 3R approaches [3, 4, 6]. The reductions in the costs of genomics and transcriptomic assays are enabling factors toward 3R as well [6, 8].

In vitro toxicological outcomes are often based on large-scale compound response profiles, which summarize the responses in a particular cell context. For instance, NCI-60 developmental therapeutics program uses several metrics to quantify dose-responses to a library of thousands of compounds across a panel of 59 human tumor cell lines; such summary metrics include: GI50 (50% Growth Inhibition), TGI (Total Growth Inhibition), and LC50 (50% Lethal Concentration) (https://dtp.cancer.gov/discovery_development/nci-60/). In such a high-throughput setting, the computational task is to search for patterns of toxicity outcomes in correlation with genomic and molecular profiles of the same panel of cell lines. However, cytotoxicity is not a biologically uniform response. Cells use multiple mechanisms that depend on the chemical or drug and the dose at which it is applied to respond to and counter the effects of stressors. Transcriptomic profiling and subsequent analyses using component modeling approaches discussed herein can segment these responses into biologically intelligible and explainable sub-responses, while also providing predictive models.

Therefore, advances in machine learning methodology allow study of toxicogenomic relationships in a more systematic fashion and reveal valuable drug-gene associations. For instance, community efforts have shown great promise to improve in silico predictions of drug sensitivity [9]. In another effort [10] carried out a personalized quantitative structure–activity relationship QSAR analysis by integrating gene expression, drug structures, and drug response profiles using a non-linear machine learning approach. Their study demonstrated the possibility to predict the drug sensitivity outcome for untested drugs even in new cell types. drug-pathway associations can be identified using advanced machine learning methodologies that model the complex molecular interactions [11]. Recently, integrative multitask sparse regression methods have been used to systematically identify biomarker combinations for predicting drug outcomes [12]. Increasing evidence from recent studies thus poses the hypothesis that common patterns in the activity profiles of genes and sensitivity/toxicity profiles of drugs can identify cellular response mechanisms and could be used even in predicting the tissue type or cell context-specific toxicity outcome of drug treatment.

This chapter is organized as follows: Sect. 4.2 introduces representative classes of recent machine learning methods, with a specific emphasis on the matrix and tensor factorization methods. Section 4.3 demonstrates the application of these methods to identification of toxicogenomic relationships in example case studies, followed by a discussion in Sect. 4.4. Section 4.5 concludes the chapter with current limitations and future directions in these developments.

4.2 Machine Learning Methods

Machine learning algorithms search for patterns in data to extract useful information [13, 14]. These algorithms learn a representation a.k.a. the model from existing data samples and then utilize the model in different tasks. When applied to experimental

data, the model formulation and learning processes take into account various forms of inherent noise and corruptions in the measurements to learn a cleaner representation of the data. In toxicogenomics, similar to many other real-life applications, data is expected to be noisy and high dimensional, contains missing values, and may also include correlated variables, which all make direct analysis complicated [15]. In such cases, the key feature of machine learning is to identify a low-dimensional, hidden representation that captures and summarizes the relevant information for the toxicogenomic modeling task. These summaries can then be used to understand the compound's MoA and/or predict the cellular outcomes of the drugs in different cell contexts.

4.2.1 Matrix Factorization

In machine learning, matrix factorization (MF) is a well-established approach to summarize a data set through unobserved features that explain why some parts of the data are similar. MF has applications in broad range of scientific domains, and it is widely used in several applications, including prediction of missing values, dimensionality reduction, as well as data visualization [16, 17]. This wide applicability comes from the assumption that MF can be seen as a means of describing the underlying processes which generated the data. Specifically, MF assumes that measurements have been produced by a combination of a number of latent processes and aims to identify the factors (a.k.a. components) that describe these processes. Figure 4.1 shows a visual illustration of matrix factorization, where a matrix **X** is factorized into distinct low-dimensional components. This component decomposition is valuable for many applications, as the different components can be related to separate mechanisms that may have contributed to the data. Several matrix factorization methods have been proposed for various applications, including factor analysis (FA), principal component analysis (PCA), and Latent Dirichlet Allocation (LDA, see Sect. 2.2) [18–20]. While FA and PCA are designed for continuous data sets, LDA is formulated for discrete data sets.

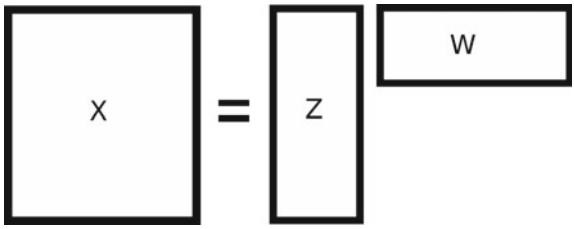

Fig. 4.1 Visual representation of matrix factorization. The data matrix **X** is factorized into low-dimensional matrices **Z** and **W** that capture the key statistical patterns in the data

4.2.2 Latent Dirichlet Allocation

The Latent Dirichlet Allocation (LDA [20, 21]) is a probabilistic formulation of factorization for discrete data sets. Formally, it is a three-level hierarchical Bayesian model that models the probabilities of each input feature to appear in each component. The Dirichlet distribution is a multivariate probability distribution used in LDA to mitigate overfitting and to help LDA to achieve its generalizability beyond the training data. LDA has demonstrated wide applicability in natural language processing, as text data sets can directly be encoded as discrete variables [22–24] as well as in genomic data sets [25–27].

4.2.3 Group Factor Analysis

Group factor analysis (GFA [28, 29]) is a recent machine learning method designed to capture relationships between multiple data sets. GFA models the relationships as statistical dependencies by reducing multiple data sets (also known as views) to learn a joint low-dimensional representation. The joint representation of the data sets is characterized by components that may be active in one or several of the data views as shown in Fig. 4.2. An active component captures underlying relationships between the views in which it is active. For example, the active component of all views captures a common dependency structure between all views, while a component active only in a single view identifies the variance and features unique to that particular view only. GFA learns the components and their activity patterns in a truly data-driven fashion, making it possible to comprehensively capture the interdependencies between all the data views. An easy to use implementation of GFA has been made freely available as an R-package [30].

Formally, for a given collection of M data sets $\mathbf{X}^{(m)} \in \mathcal{R}^{N \times D_m}$ where $m = 1 \ldots M$, having N paired samples and D_m dimensions, GFA learns a joint low-dimensional factorization of the M matrices. The model is formulated as a product of the Gaussian latent variable matrix $\mathbf{Z} \in \mathcal{R}^{N \times K}$ (containing the K components) and view-specific projection weights $\mathbf{W}^{(m)} \in \mathcal{R}^{D_m \times K}$:

$$\mathbf{x}_n^{(m)} \sim \mathcal{N}\left(\mathbf{W}^{(m)} z_n, \sum^{(m)}\right),$$

$$\mathbf{z}_n \sim \mathcal{N}(0, \mathbf{I})$$

$$w_{d,k}^{(m)} \sim h_{m,k}\mathcal{N}\left(0, \left(\alpha_{d,k}^{(m)}\right)^{-1}\right) + (1 - h_{m,k})\delta_0$$

$$h_{m,k} \sim Bernoulli(\pi_k)$$

$$\pi_k \sim Beta(a^{\pi}, b^{\pi})$$

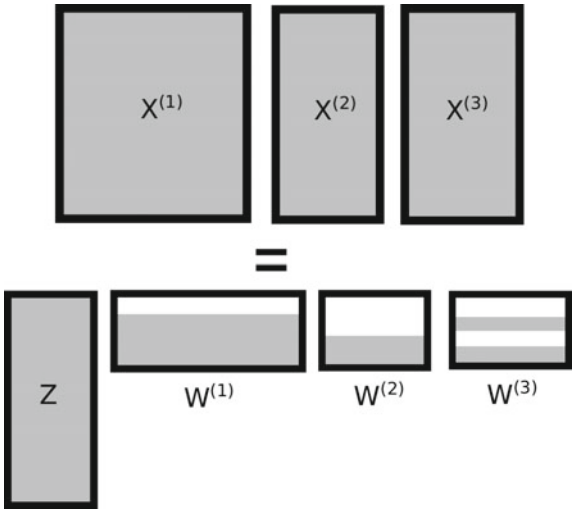

Fig. 4.2 Visual representation of group factor analysis. GFA factorizes a set of data matrices $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ … $\mathbf{X}^{(m)}$, into their joint low-dimensional factors $\mathbf{Z}$. The factors can be active in one or more data matrices through the projection matrices $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$ … $\mathbf{W}^{(m)}$. The $\mathbf{W}$'s are learned to hold a group-wise sparse structure that models the dependency patterns across the data matrices. The sparsity is illustrated by white color which represents zero weights, while shaded color represents nonzero values in the figure

$$\alpha_{d,k}^{(m)} \sim Gamma(a^{\alpha}, b^{\alpha})$$

Here, $\sum^{(m)}$ is a diagonal noise covariance matrix. The latent variable $\mathbf{z}_n$ is common between all the views and captures the response patterns. The projection matrices $\mathbf{w}_{:,k}^{(m)}$ are specific to each view and translate the dependency patterns across views.

GFA achieves the joint factorization by assuming that the projections $\mathbf{w}_{:,k}^{(m)}$ are group-wise sparse. The group sparse projections $\mathbf{w}_{:,k}^{(m)}$ capture both group-specific variations (activity displayed only in one view) as well as dependencies between multiple groups (activity in more than one view). The sparsity is implemented in two layers through a group-wise spike and slab prior formulation using Beta-Bernoulli distribution [29] and an element-wise normal-Gamma Automatic Relevance Determination (ARD) [31]. As a result, the project matrices $\mathbf{W}^{(m)}$ are both group and feature-wise sparse, which is compatible with the biological assumptions of targeted action mechanisms making the results easier to interpret.

4.2.4 Tensor Factorization

A tensor is a multidimensional array $\mathcal{X} \in \mathcal{R}^{I_1 \times I_2 \times \ldots \times I_j}$ and a generalization of matrices and vectors to higher order spaces. Tensors are therefore useful for representing data that has more than two dimensions. Such representation allows investigation of relationships that span multidimensional constructs. Mathematically, a tensor is also commonly defined as an element of space induced by the tensor product of vector spaces.

In order to capture the highly structured patterns of a multidimensional data set, tensor methods employ constrained formulations that help to avoid the overfitting problem [32]. A key characteristic of these tensor formulations is that they have fewer parameters than their matrix counterparts. Analogous to matrix factorizations presented in Sect. 4.2.2, there exist several tensor factorization methods that can be used to discover underlying dependencies in the data [32]. CANDECOMP/PARAFAC (CP; [33, 34]) and Tucker family [35] are the most widely used tensor decomposition methods. The interested reader is referred to [32] for a comprehensive review of various tensor factorization methods.

Tensor factorizations have obtained significant success in a large number of domains, including chemometrics, psychometrics, bioinformatics, and have shown immense promise for advanced applications in toxicology and toxicogenomics. For example, tensor factorization has been used to explore stimuli-variant gene expression patterns [36], as well as in integrating phenotypic responses from multiple studies [37, 38], modeling dependencies between metabolic and gene expression networks [39], as well as in joint QSAR and toxicogenomic analysis [40, 41].

CP factorization, also known as the canonical decomposition or parallel factor analysis [33, 34], is the most widely used tensor factorization method. CP is a natural extension of matrix factorization to arrays of order 3 or more as shown in Fig. 4.3. The method can be seen as carrying out simultaneous factor analysis on multiple slabs (matrices) of a tensor such that the factors of each slab differ just by a scale. CP factorization is defined in a symmetric fashion over all the modes, such that a tensor is decomposed into a sum of rank-one tensors, where each rank-one tensor is the outer product of the latent vector in all modes. For a third order tensor $\mathcal{X} \in \mathcal{R}^{N \times D \times L}$, a rank-K CP is represented as:

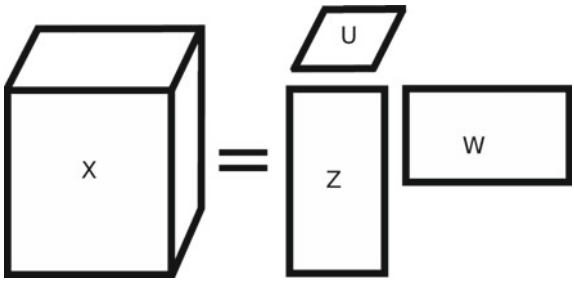

Fig. 4.3 Visual representation of CP factorization of a third order tensor. The data tensor $\mathcal{X}$ is factorized into low-dimensional matrices **Z**, **U,** and **W** that capture the key statistical patterns in the data

$$\mathcal{X} = \sum_{k=1}^{K} \mathbf{z}_k \circ \mathbf{w}_k \circ \mathbf{u}_k + \epsilon$$

where **Z** and **U** and **W** are the latent variables corresponding to the three modes.

Several implementations of CP factorization have existed for quite some time now, for example, the seminal implementation by Andersson and Bro [42]. Recently, CP and other factorizations have gained substantial interest among the machine learning community [43, 44], since recent developments addressed several methodological challenges posed by multi-way data sets. More recently, an easy to use probabilistic implementation of CP was presented by Khan and Ammad-ud-din [45]. The implementation automatically handles missing values in the data, hence making it applicable to a wide selection of real-world data sets. It also features automatic component selection as well as visualization and prediction routines making both exploratory and predictive analytics easier.

4.2.5 Multi-tensor Factorization

Multi-tensor factorization (MTF [40, 41]) is a new machine learning method designed to capture relationships between a collection of tensor data sets. MTF jointly factorizes multiple tensors to learn a joint low-dimensional representation that models the statistical dependencies between the tensors. Interestingly, MTF considers matrices as tensors of order two, thus enabling joint factorization of both matrices and tensors. This characteristic makes it possible to analyze novel data sets composed of matrices as well as tensors in a single joint analysis.

MTF is designed to factorize multiple co-occurring data sets, with the objective of distinguishing the shared and specific components regardless of their matrix or tensor nature. This is achieved by modeling the entire variation of all data sets through a common Factor analysis and CP-type factorization having two keys features. First, the factorization is characterized by latent variables **Z** that are common between all the views (tensor and matrices). This allows the factorization to capture cross-dependencies regardless of the data view. Second, the loadings **W** controls which of the patterns in **Z** are active in each of the views. Learning these **W** loadings makes it possible to identify the dependency patterns in a truly data-driven fashion without any prior information on dependency patterns.

Formally, for multiple paired tensors $\mathcal{X}^{(t)} \in \mathcal{R}^{N \times D_t \times L}$, where $t = 1 \ldots T$, we specify a joint model of matrices and tensor. An indicator variable β_t identifies the tensors ($\beta_t = 1$) and matrices ($\beta_t = 2$), MTF is formulated using normal distributions and conjugate priors as:

$$\mathrm{x}^{(t)}_{n,d_t,l} \sim \mathcal{N}\left(\mathrm{z}_{n,k}.\mathrm{w}_{d_t,k}.\mathrm{u}_{\mathrm{l,k}}, (\tau^{(t)})^{-1}\right)$$

$$\mathbf{Z}, \mathbf{U}^{(t)} \sim \mathcal{N}(0, \mathbf{I})$$

$$w_{d,k}^{(t)} \sim h_{t,k}\mathcal{N}\left(0, \left(\alpha_{d,k}^{(t)}\right)^{-1}\right) + \left(1 - h_{t,k}\right)\delta_0$$

$$h_{t,k} \sim Bernoulli(\pi_k)$$

$$\pi_k \sim Beta(a^{\pi}, b^{\pi})$$

$$\alpha_{d,k}^{(t)} \sim Gamma(a^{\alpha}, b^{\alpha})$$

$$\tau^{(t)} \sim Gamma(a^{\tau}, b^{\tau}).$$

Here, the latent variables **Z** and **U** are common to all the tensors and capture the underlying patterns, while $\mathbf{W}^{(t)}$ translate these patterns for each tensor. The binary variables $h_{t,k}$ control the view activity through a spike and slab prior and are automatically learned from the data. The model also enforces feature-wise sparsity through α to learn sparse features that are easier to interpret. The method is implemented using a Gibbs sampler in *R* programming language and made available freely (http://research.ics.aalto.fi/mi/software/MTF). The implementation learns the model parameters in a Bayesian formulation, while providing default settings for all the hyperparameters.

4.3 Selected Case Studies

4.3.1 Toxicogenomic Data Sets

The toxicogenomic tools described in this chapter are primarily built upon the Connectivity Map (CMap) and NCI60 data sets. CMap, introduced by the US Broad Institute, is a compendium of gene expression response profiles from 1309 small molecules comprising mostly FDA approved drugs ([46]; https://www.broadinstitute.org/connectivity-map-cmap). The post-treatment measurements originated from three main cancer cell lines spanning different tissues or cell types, namely, breast (MCF7), prostate (PC3), and blood (HL60). CMap has been widely used to study interactions between small molecules, genes, and diseases for various purposes including understanding the drug MoA, identifying biologically similar compounds as well as molecular mechanisms of toxicity. The treatment versus control differential gene expression (log2 readout) was obtained from the CMap data set, such that positive expression values represent up-regulation and negative represent down-regulation as a result of treatment [4].

The NCI60 is a unique data repository from the US National Cancer Institute (NCI) that screened thousands of compounds over 59 cancer cell lines to provide measurements of drug responses (Shoemaker 2006; https://dtp.cancer.gov/discovery_development/nci-60). Drug response metrics include GI50 (50% Growth Inhibition), total growth inhibition (TGI), and LC50 (50% Lethal Concentration). A large number

of common compounds tested in CMap cell lines were also profiled by the NCI60 program. This presents the unique opportunity to study toxic effects by integrating these two large-scale data sets (see Sect. 4.3.3). For each CMap drug-cell pair which was also screened by NCI60, a dose-dependent toxicity score was computed such that positive values indicated that the CMap instance is profiled at a drug concentration higher than GI50, TGI, or LC50, and therefore suggest a dose-dependent cytotoxic response.

4.3.2 Multi-view Toxicogenomic Using Group Factor Analysis

To study the gene–toxicity relationships, we performed an integrated modeling of the two data sets, CMap and NCI60. The CMap data comprised detailed gene level differential expression profiles that represent the molecular response space across 11,350 genes, measured after 222 drug treatments across 3 cell lines. These data were preprocessed as previously described [47]. To focus the analysis, the 1106 highest variance genes were selected to form an expression matrix consisting of 222 drug-cell samples $\times 1106$ genes. The toxicity values described in Sect. 4.3.1 were used to represent profiles of 222 drug-cell samples $\times 3$ toxicity measures.

Group factor analysis (GFA) is designed to model the relationships between multiple data sets. Here, GFA was used to identify the toxicogenomic dependencies between drug-induced gene expression changes and toxicity scores. These dependencies, once identified, can elicit insights into molecular mechanisms of toxicity. GFA was run with large enough components as specified by Virtanen et al. [28], identifying 8 shared components that capture cross-expression and toxicity relationships, as shown in Fig. 4.4, whereas a number of components found were specific to one of the data sets only. The shared components model the dependencies between the data sets while those specific to gene expression capture patterns that are not correlated with toxicity and vice versa. The components 1 through 8 had varying numbers of genes attached to them: 518, 748, 39, 90, 27, 45, 16, and 20. The first two components included an excess of up-regulated genes (component 1:316) and down-regulated (component 2:706) genes, respectively.

Functional analysis of the eight components was performed with Ingenuity Pathway Analysis (IPA) which indicated that the first two components captured the largest number of biological mechanisms. The first component is highlighted here, as up-regulated genes are most informative for biomarker analysis applications (Fig. 4.5). Component 1 enriched for many organ toxicity-related gene lists, including hepatic cholestasis and liver necrosis as well as functional pathways related to oxidative stress, the p. 53 pathway activation and Nf-kappa *B* signaling and Toll-like receptor (TLR) activation. RELA, the NF-kappa *B* regulator, was predicted to be most clearly effected ($p < 10–16$ and Z-score 3.5). Others included the TP53 ($p < 10–10$, $Z > 1$), TLR-related ECSIT ($p < 10–15$, $Z > 3.5$), and NR3C1 ($p < 10–14$, $Z < -05$),

Fig. 4.4 Toxicogenomic component activity plot. The plot shows the components that are found by the GFA model as active in the joint gene expression and toxicity data set. The *y*-axis shows the component number in ascending order while the *x*-axis shows the two data sets. The components colored black are active. The model was run for $K = 40$ components and a total of 8 components (bottom black in both gene expression and toxicity) are found as shared between the two data sets. These components capture statistical patterns that are correlated across the two data sets and hence can be hypothesized for representing molecular mechanisms of toxicity

supporting the role of NF-kappa *B* signaling and the Toll-like receptor activation for component 1. The GFA model identified related drugs across all three cancer types, hence suggesting a generic response of the drugs.

The second component included many cell cycle-related genes, as could be expected for a component which mainly contained down-regulated genes. Similar pathways were found activated among the 14 predictive toxicogenomic space PTGS components derived using the LDA analysis, and there is an average of almost 40% overlap between the PTGS genes and the GFA genes [4]. It is interesting to note that the first two GFA components were much larger than the other six, whereas the PTGS components had more equal numbers of genes that were significantly associated with them. Further studies would be needed to verify the utility of the GFA components for toxicity-mode-of-action studies, including biomarker discovery and drug-induced liver injury (DILI) prediction.

4.3.3 Structural Toxicogenomic Using Multi-tensor Factorization

Toxicogenomic applications can be extended to simultaneously include a quantitative structure activity response (QSAR) analysis, by modeling the dependencies between cellular responses of drugs and their structural descriptors. The formulation can, therefore, explore, identify, and predict genomic responses linked to drugs toxicity, while simultaneously discovering their cancer specificity and correspondence to structural properties of the drugs.

Data collection for such analysis can be represented as a set of multiple tensors and matrices. In this example, we specified two tensors and one matrix. The post-treatment gene expression data from CMap was represented as the first tensor of drugs *times* cancers *times* genes dimensions. Multiple toxicity measures such as GI50, TGI, and LC50 from the NCI60 formed the second tensor of drugs *times* cancers *times*

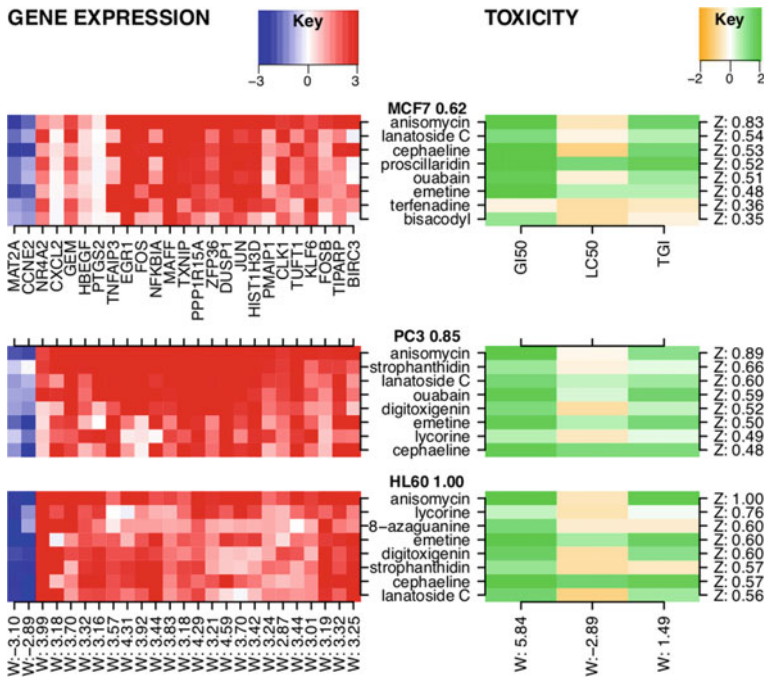

Fig. 4.5 First shared component found by GFA. The plot shows the gene expression profiles of the top genes for the top drugs of the component across the three cancer cell lines (MCF7, PC3, and HL60). Red represents up-regulated expression while blue is down-regulated expression. The correspondingly active toxicity profiles of the same drugs are shown on the right. Here green represents high dose-dependent toxicity values

toxicity measures. Finally, the structural properties of the drugs were represented as a matrix of drugs *times* descriptors.

The expression and toxicity data sets from CMap and NCI60 were processed as described in [4]. For drug structures, the modeling could make use of one or more different types of structures based on the hypothesis being tested; for example [48] used both 3D descriptors and 2D fingerprints of the drugs for structure response analysis. In this example, functional connectivity fingerprints FCFP4 were used for representing the structural properties of the drugs. FCFP4 are advanced 2D circular topological fingerprints that have been designed for modeling of structure–activity relationships.

The multi-tensor factorization (MTF) method of Sect. 4.2.3 [41] was used to explore the structural toxicogenomic relationships. The model identified three key response components that are shared between gene expression, toxicity, and structural data sets, revealing findings that are both recently established as biological insights, as well as new biological discoveries that may have potential impact. The first component identified a response primarily driven by three heat shock protein (HSP) inhibitor drugs, geldanamycin, tanespimycin, and alvespimycin, all of which

are structurally analogous drugs. The drugs demonstrated an HSP response of the cells as through up-regulation of key HSP genes. This pan-cancer response across all three cancers is linked to the toxicity outcomes of the drugs. HSP90 is a molecular chaperone protein that is essential for stabilization of a variety of other proteins [49], and HSP90 inhibitors bind to the protein, resulting in its loss of function. HSP90 inhibitors have been evaluated for their therapeutic efficacy in multiple cancers [50, 51]. This component, therefore, presents a well-known HSP90 response of cancer cells. For details on this and other components, see [41].

4.3.4 Predictive Toxicogenomic Space (PTGS)

Predictive toxicogenomics space (PTGS [4]) is a recent "big data compacting and data fusion" methodology to model various adverse and toxic outcomes on cellular and organism levels. A machine learning-based data summarization approach was applied on a large transcriptomics data set. This methodology formed a predictive tool termed PTGS that used as features over 1000 genes distributed over 14 overlapping cytotoxicity-related gene space components, as described in [4]. Specifically, a LDA matrix factorization-based method was applied to the gene profiles from the Connectivity Map data set, and the resulting summarized components were fused with cytotoxicity data from the NCI-60 cancer cell line screens to generate the PTGS. The PTGS tool was validated for predicting drug-induced liver injury (DILI) and liver cytopathological changes by calculating PTGS component scores within three liver-related subsets of the independent TG-GATEs database [52], being the largest public toxicogenomics database. It was shown to successfully capture all the studied liver pathological changes in rats, and in conjunction with human therapeutic drug exposure levels (C_{max}), was able to facilitate the use of cell culture-derived toxicogenomics experiments with human and rat hepatocytes to predict DILI with greater accuracy than other in vitro methods [4].

4.4 Discussion

Recent advances in machine learning methodologies have made it possible to perform integrated analysis of the gene expression response data and toxicity profiles directly. Such detailed analysis offers deeper insights by linking the activity patterns of the genes directly with the toxicity responses, and hence enriching the factor components with detailed interactions. As molecular responses of cancer cells are known to depend on a multitude of factors, including drug MoA, cell type, and cellular states, simultaneous modeling of these various factors is beginning to attract attention. Specifically, in cancer, cells are known to be heterogeneous and respond selectively to targeted drugs, making it valuable to systematically model the various factors and segregate responses specific to a particular cancer-type from those which are generic.

A limitation of current methods is the ability to handle missing values particularly when considering overlap between different data sets. Compared to many other "Big Data" study areas, biomedical data is less extensive and contains more missing values. The LDA method used with the PTGS had the advantage that the entire CMap data set could be used to derive the initial components, whereas the GFA method required at least some overlap between all variables, reducing the amount of gene expression data used. Tensor factorization methods are even less tolerant of missing values. Therefore unique methodological considerations and trade-offs apply to each study.

A key outcome of this joint analysis is the ability to predict the toxicity outcomes of compound treatment. The prediction of unexpected toxic effects is a challenging and important goal in toxicology. The presented first steps in computational toxicogenomic open up a systematic way for genomics-driven prediction of toxic effects. In addition, these provide novel mechanistic insights into the links between genomic measurements of cells and toxicological profiles of drugs. Gene expression response profiles of drugs present a popular systems-level view, while toxicity profiles summarize the drugs' phenotypic behavior. Large repositories of gene expression and drug sensitivity profiles such as those emerging from NCI60, CMap, CCLE, Sanger, and LINCS profile cellular responses at several levels of detail in a cell context-specific manner. With the emergence of heterogeneous and partially paired data sets, joint factorizations are gaining popularity to identify novel dependency patterns, as well as to design powerful predictive applications [48, 53]. These recent advances in machine learning, and especially the methods described in Sect. 4.2, enable systematic analysis of such large data repositories to provide novel toxicogenomic insights and predictions.

4.5 Conclusion and Future Directions

State-of-the-art machine learning methods have been presented here for modeling various toxicogenomic relationships. These advanced computational methodologies enable integration of disparate, high-dimensional data sources, including but not limited to omics, drug screening, chemical structures, and drug-targets to achieve novel toxicogenomic analysis in terms of:

(i) providing means for predicting personalized toxicity outcomes,
(ii) identifying toxic modes of action, and
(iii) enabling quantitative structure activity modeling.

The here presented works suggest novel directions for future analysis. From the application perspective, matrix and tensor factorization methods can serve to stimulate integrative analysis of various toxicological and toxicogenomic data sets to suggest novel hypotheses. For example, a joint analysis of omics, toxicity, and drug-target data sets can help to identify disparate target-driven and toxic molecular mechanisms. Integrative analysis with drug-side effect repositories can help draw novel interactions between disease, side effect, and toxicity mechanisms. From a holistic

angle, a large-scale analysis may even help us understand the different toxic states of a cell and the molecular drivers of each cellular state.

While there are limitations in the current analysis, future extensions of the analyses can advance our knowledge in various directions. First, using detailed drug-target interactions in the models could help classify the on-target and off-target effects more reliably; however, a key limitation here is to obtain large-scale standardized drug-target profiles. Very recently works in standardizing the drug-target interactions have come up on a large-scale [54] and exploring these for an integrated drug-target-toxicogenomic analysis would be an interesting future direction. Secondly, a large majority of toxicity analysis is performed on data originating from cell line panels. It would be valuable to explore if tissue-specific toxicity profiles are available for a more robust and practically applicable analysis. Third, organism-level toxicity data is limited to only a few organisms only; it is important to evaluate how comprehensive such modeling is in general and how widely the results can be applied across organisms.

In terms of future developments in the toxicology practices, studies, and risk assessment strategies, we hope the presented works could stimulate the integration of advanced machine learning models. For example, the methods presented here can be used to identify the markers of toxic response toward a data and knowledge-driven approach for risk assessment.

References

1. Grabinger T et al (2014) Ex vivo culture of intestinal crypt organoids as a model system for assessing cell death induction in intestinal epithelial cells and enteropathy. Cell Death Dis 5(5):e1228
2. Aberdam E et al (2017) Induced pluripotent stem cell-derived limbal epithelial cells (LiPSC) as a cellular alternative for in vitro ocular toxicity testing. PLoS ONE 12(6):e0179913
3. Hartung T et al (2012) Food for thought … systems toxicology. ALTEX 29(2):119–128
4. Kohonen P et al (2017) A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. Nat Commun 8:15932
5. Kohonen P et al (2014) Cancer biology, toxicology and alternative methods development go hand-in-hand. Basic Clin Pharmacol Toxicol 115:50–58
6. Grafström RC et al (2015) Toward the replacement of animal experiments through th bioinformatics-driven analysis of 'omics' data from human cell cultures. Altern Lab Anim 43:325–332
7. Nymark P et al (2018) A data fusion pipeline for generating and enriching adverse outcome pathway descriptions. Toxicol Sci 162(1):264–275
8. Yeakley JM et al (2017) A trichostatin a expression signature identified by TempO-Seq targeted whole transcriptome profiling. PLoS One 12(5)
9. Costello JC et al (2014) A community effort to assess and improve drug sensitivity prediction algorithms. Nat Biotechnol 32(12):1202–1212
10. Ammad-Ud-Din M et al (2014) Integrative and personalized QSAR analysis in cancer by Kernelized Bayesian matrix factorization. J Chem Inf Model 54(8):2347–2359
11. Ammad-ud-din M et al (2016) Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization. Bioinformatics 32(17):i455–i463

12. Ammad-ud-din M et al (2017) Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. Bioinformatics 33(14):i359–i368
13. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York, USA
14. Gelman A et al (2013) Bayesian data analysis. Chapman and Hall/CRC
15. Bartholomew DJ et al (2011) Latent variable models and factor analysis: a unified approach. Wiley, 904
16. Salakhutdinov R, Mnih A (2008) Bayesian probabilistic matrix factorization using markov chain monte carlo. In: Proceedings of the 25th international conference on Machine learning, pp 880–887
17. Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. J R Stat Soc: Ser B (Stat Methodol) 61(3):611–622
18. Witten DM et al (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, Biostatistics, pp 1–8
19. Kossenkov AV, Ochs MF (2009) Matrix factorization for recovery of biological processes from microarray data. Methods Enzymol 467:59–77
20. Blei DM et al (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
21. Ghahramani Z (2015) Probabilistic machine learning and artificial intelligence. Nature 28:452–459
22. Guo Y et al (2017) Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tour Manag 59:467–483
23. Moro S et al (2015) Business intelligence in banking: A literature analysis from 2002–2013 using text mining and latent Dirichlet allocation. Expert Syst Appl 42(3):1314–1324
24. Krestel R et al (2009) Latent dirichlet allocation for tag recommendation. In: Proceedings of the third ACM conference on recommender systems, pp 61–68
25. Caldas J et al (2009) Probabilistic retrieval and visualization of biologically relevant microarray experiments. Bioinformatics 25(12):145–153
26. Pinoli P et al (2014) Latent Dirichlet allocation based on Gibbs sampling for gene function prediction. In: Computational intelligence in bioinformatics and computational biology, pp 1–8
27. Backenroth D et al (2018) FUN-LDA: a latent dirichlet allocation model for predicting tissue-specific functional effects of noncoding variation: methods and applications. Am J Hum Genet 102(5):920–942
28. Virtanen S et al (2012) Bayesian group factor analysis. In: Artificial Intelligence and Statistics, pp 1269–1277
29. Klami A et al (2015) Group factor analysis. IEEE Trans Neural Netw Learn Syst 26(9):2136–2147
30. Leppäaho E et al (2017) GFA: exploratory analysis of multiple data sources with group factor analysis. J Mach Learn Res 18(1):1294–1298
31. Neal RM (1996) Bayesian learning for neural networks. Springer-Verlag
32. Kolda T, Bader B (2009) Tensor decompositions and applications. SIAM Review 51(3):455–500
33. Carroll JD, Chang JJ (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. Psychometrika 35(3):283–319
34. Harshman RA (1970) Foundations of the parafac procedure: models and conditions for an explanatory multimodal factor analysis, UCLA Working Papers in Phonetics, 16:1–84
35. Tucker LR (1966) Some mathematical notes on three-mode factor analysis. Psychometrika 31(3):279–311
36. Yener B et al (2008) Multiway modeling and analysis in stem cell systems biology. BMC Syst Biol 2(1):63
37. Omberg L et al (2007) A tensor higher-order singular value decomposition for integrative analysis of dna microarray data from different studies. Proc Natl Acad Sci 104(47):18371–18376
38. Li W et al (2011) Integrative analysis of many weighted co-expression networks using tensor computation. PLoS Comput Biol 7(6):e1001106

39. Brink-Jensen K et al (2013) Integrative analysis of metabolomics and transcriptomics data: a unified model framework to identify underlying system pathways. PLoS ONE 8(9):e72116
40. Khan SA, Kaski S (2014) Bayesian multi-view tensor factorization. In: Joint european conference on machine learning and knowledge discovery in databases. Springer, Berlin, Heidelberg, pp 656–671
41. Khan SA et al (2016) Bayesian multi-tensor factorization. Mach Learn 105(2):233–253
42. Andersson CA, Bro R (2000) The N-way toolbox for MATLAB. Chemometr Intell Lab Syst 52(1):1–4
43. Mørup M, Hansen LK (2009) Automatic relevance determination for multiway models. J Chemom 23(7):352–363
44. Xiong L (2010) Temporal collaborative filtering with bayesian probabilistic tensor factorization, vol. 10. In: Proceedings of SIAM data mining, pp 211–222
45. Khan SA, Ammad-ud-din M (2017) TensorBF: an R package for Bayesian tensor factorization, bioRxiv, 6097048 1–6
46. Lamb J et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes and disease. Science 313(5795):1929–1935
47. Khan SA et al (2012) Comprehensive data-driven analysis of the impact of chemoinformatic structure on the genome-wide biological response profiles of cancer cells to 1159 drugs. BMC Bioinform 13(1):112–127
48. Khan SA (2014) Identification of structural features in chemicals associated with cancer drug response: a systematic data-driven analysis. Bioinformatics 30(17):i497–i504
49. Shoemaker RH (2006) The nci60 human tumour cell line anticancer drug screen. Nat Rev Cancer 6(10):813–823
50. Isaacs JS et al (2003) Heat shock protein 90 as a molecular target for cancer therapeutics. Cancer Cell 3(3):213–217
51. Neckers L, Workman P (2012) Hsp90 molecular chaperone inhibitors: are we there yet? Clin Cancer Res 18(1):64–76
52. Igarashi Y et al (2015) Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res 43:D921–D927
53. Hore V et al (2016) Tensor decomposition for multiple-tissue gene expression experiments. Nat Genet 48(9):1094
54. Tang J et al (2018) Drug target commons: a community effort to build a consensus knowledge base for drug-target interactions. Cell Chem Biol 25(2):224–229

Suleiman Ali Khan received his Ph.D. in Machine Learning from Aalto University, Finland in 2015. He has specialized in Bayesian machine learning and its applications in bioinformatics. He subsequently joined FIMM as a postdoctoral researcher and has developed novel statistical machine learning methods for various drug discovery and drug repurposing approaches. He is currently an Academy of Finland postdoctoral researcher working on identification of predictive genomic markers through advanced and integrative machine learning methods. His current research interests include statistical machine learning, explainable artificial intelligence, computational biology, and medicine.

Tero Aittokallio received his Ph.D. in Applied Mathematics from the University of Turku in 2001. He did his postdoctoral training in the Systems Biology Lab at the Institut Pasteur (2006–2007), where he focused on network biology applications using high-throughput experimental assays. Dr. Aittokallio then launched his independent career as a principal investigator in the Turku Biomathematics Research Group in 2007 and received a five-year appointment as an Academy of Finland Research Fellow (2007–2012). In 2011, he started as EMBL Group Leader at FIMM, where his research group is developing computational biology approaches to personalized medicine.

Andreas Scherer received his Ph.D. in molecular biology at the Max-Planck-Institute for Immunobiology in Germany. After a postdoctoral position at the UT Southwestern Medical Center in Dallas, USA, until 1999, he has been working on biomarker research in various positions at Novartis in Basel, CH, from 2000 until 2007. He then worked as consultant and bioinformatician in his own company in Finland, and as Manager of the Bioinformatics Team at the Australian Genome Research Facility (AGRF) in Melbourne, AU. He is currently working at the Institute for Molecular Medicine Finland (FIMM) at the University of Helsinki. He is the National Coordinator for EATRIS ERIC, the European Research Infrastructure for Translational Medicine. He has been elected co-chair of the EATRIS Biomarker Platform and is chair of the EATRIS Quality Initiative. He has focused on technological, biological, clinical, and legal, and ethical aspects of personalized health.

Roland Grafström is a Professor in Risk Assessment at the Karolinska Institutet since 2000. His speciality is biocomputing and informatics for toxicogenomics, cancer biology, personalized medicine, and safety-by-design work. It includes mechanistic and mode-of-action-driven modeling and data interpretation related to toxicity effects and adverse outcome pathways. He is also the CEO of Predictomics AB (SME, Stockholm, Sweden), being a subsidiary to Cercon AB, a biomedical consulting business in operation since 1983. Customers are trained for using in vitro and in silico-based predictive workflows based on twenty-first century toxicology and 1R (replacement category) concept, including frequent organization of courses and research guidance for industry and academia, and developing IPR related to the above fields. He serves on this basis as science advisor for Misvik Biology Oy, Turku, Finland and BioTeSys GmbH, Esslingen, Germany.

Pekka Kohonen is an assistant Professor in Risk Assessment at the Karolinska Institutet. He is an expert on toxicogenomics, bioinformatics, and data sciences for toxicology. Interest areas include integrative omics analyses, high-throughput/high content screening analysis, and Big Data analytics to predict toxicity effects of chemicals and nanomaterials, using primarily cell culture data (1R). He co-developed with Roland Grafström and others the predictive toxicogenomics space (PTGS) scoring concept for predicting drug-induced liver injury (DILI). PK is also affiliated with Predictomics AB (SME, Stockholm, Sweden). Won the 2014 Lush Science prize together with Prof. Roland Grafström for contributions to twenty-first century Toxicology to replace animal use in toxicity testing.

Chapter 5
Cardio-oncology: Network-Based Prediction of Cancer Therapy-Induced Cardiotoxicity

Feixiong Cheng

Abstract The growing awareness of cardiotoxicities associated with cancer treatment has led to the emerging field of cardio-oncology (also known onco-cardiology), which centers on screening, monitoring, and treating cancer patients with cardiac dysfunction before, during, or after cancer treatment. The classical approach centered on the hypothesis of 'one gene, one drug, one disease' in the traditional drug discovery paradigm may have contributed to unanticipated off-target cardiotoxicity. However, there are no guidelines in terms of how to prevent and efficiently treat new cardiotoxicities in drug discovery and development. Novel approaches, such as network-based drug-disease proximity, shed light on the relationship between drugs and diseases, offering novel tools for risk assessment of drug-induced cardiotoxicity. In this chapter, we will introduce an integrated, network-based, systems pharmacology approach that incorporates disease-associated proteins/genes, drug-target networks, and the human protein-protein interactome, for risk assessment of drug-induced cardiotoxicity. Specifically, we will introduce available bioinformatics resources and quantitative network analysis tools. In addition, we will showcase how to use network proximity for risk assessment of drug-induced cardiotoxicity and for understanding of their underlying cardiotoxicity-related mechanism-of-action (e.g., multi-targeted kinase inhibitors). Finally, we will discuss existing challenges and highlight future directions of network proximity approaches for comprehensive assessment of oncological drug-induced cardiotoxicity in the early stage of drug discovery, clinical trials, and post-marketing surveillance.

F. Cheng (✉)
Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Avenue Cleveland, Cleveland, OH 44195, USA
e-mail: chengf@ccf.org

Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44106, USA

Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA

Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_5

Keywords Cardio-oncology · Cardiotoxicity · Drug-target network · Disease module · Human protein-protein interactome · Network proximity · Systems pharmacology

Abbreviations

ABL1	Abelson murine leukemia viral oncogene homolog 1
AP-MS	Affinity purification followed by mass spectrometry
ATC	Anatomical therapeutic chemical
BCR	Breakpoint cluster region protein
CM	Cardiomyopathy
CTD	The comparative toxicogenomics database
CV	Cardiovascular
CVD	Cardiovascular disease
ERBB2	Erb-b2 receptor tyrosine kinase 2
FDA	Food and drug administration
FGFR	Fibroblast growth factor receptor
GWAS	Genome-wide association studies
HER2	Human epidermal growth factor receptor 2
KIT	KIT proto-oncogene receptor tyrosine kinase
MeSH	Medical Subject Headings
MoA	Mechanism-of-action
NCBI	National Center for Biotechnology Information
OMIM	Online Mendelian Inheritance in Man
PDGFR	Platelet-derived growth factor receptor
PDGFRA	Platelet-derived growth factor receptor A
PDGFRB	Beta-type platelet-derived growth factor receptor
PheWAS	Phenome-wide association study
PPI	Protein-protein interaction
QSP	Quantitative and systems pharmacology
SNP	Single-nucleotide polymorphisms
SRC	Proto-oncogene tyrosine-protein kinase Src
TTD	Therapeutic Target Database
UMLS	Unified Medical Language System
VEGFR	Vascular endothelial growth factor receptor
Y2H	Yeast two-hybrid

5.1 Introduction

Cardiovascular disease (CVD) is a leading cause of death and the second leading cause of mortality and morbidity in cancer survivors after recurrent malignancy in the USA [1, 2]. Comorbidity between cardiovascular disease and cancer suggests an underlying shared disease etiology, which can be both genetic and environmental [3–5]. One critical issue regarding environmental factors is that comorbidity between cardiovascular disease and cancer is typically associated with various anticancer treatments [6], including cytotoxic chemotherapies [7], radiotherapy [8], molecularly targeted therapies [9, 10], and immunotherapies [11, 12]. For example, a growing number of cancer survivors (>5 million) are exposed with an increased lifetime risk of anthracycline-induced cardiovascular complications [2, 13].

There are several different mechanisms-of-action (Fig. 5.1) for drug-induced cardiotoxicities, including both on-target [14] and off-target effects [9, 10]. For example, previous studies in a genetically engineered mutant mouse model have suggested that the loss of ERBB2 (erb-b2 receptor tyrosine kinase 2) in the heart can lead to heart failure and increased susceptibility to cardiotoxicity of HER2 (human epidermal growth factor receptor 2) inhibitors (e.g., trastuzumab) [14]. Most kinase inhibitors often reveal 'promiscuous' profiles [15] via inhibiting many other kinases rather than cancer-related targets, resulting in a high risk of off-target cardiotoxicities [16–18]. Furthermore, these reports may only represent the tip of the iceberg for cancer therapies [19]. We believe a key factor in the high risk of cancer treatment-related cardiotoxicities is the continued adherence to the classical 'one gene (product), one drug, one disease' paradigm in the traditional oncological drug development and regulatory reviews [20–22]. For instance, imatinib, the first approved molecularly targeted agent for the treatment of chronic myeloid leukemia, was reported to bind over 40 different human proteins, which associates with multiple cardiac complications [23, 24]. Ponatinib was approved for chronic myeloid leukemia with a fairly broad label in the USA; however, later studies reported its cardiotoxicity due to its promiscuous profiles on multiple kinases, including SRC (proto-oncogene tyrosine-protein kinase Src), PDGFR (platelet-derived growth factor receptor), FGFR (fibroblast growth factor receptor), and VEGFR (vascular endothelial growth factor receptor), which now has led to its restricted use in patients with the BCR-ABL1^{T315I} or in whom other kinase inhibitors are not effective [16–18].

The growing awareness of cancer treatment-related cardiotoxicities has led to the emerging field of cardio-oncology (also known onco-cardiology), which centers on screening, monitoring, and treating cancer patients with cardiac dysfunction before, during, or after cancer treatment [2, 6]. Furthermore, it is also an exciting field because there are no guidelines and no available US Food and Drug Administration (FDA)-approved therapeutics for preventing and treating new cardiotoxicity in cancer survivors. There is an increasing recognition that our current disease categorization approaches are inadequate to describe the scope and patient heterogeneity of complex diseases and understand the mechanism-of-action of therapeutics. Quantitative and systems pharmacology (QSP) refers to a multidisciplinary approach for the emerging

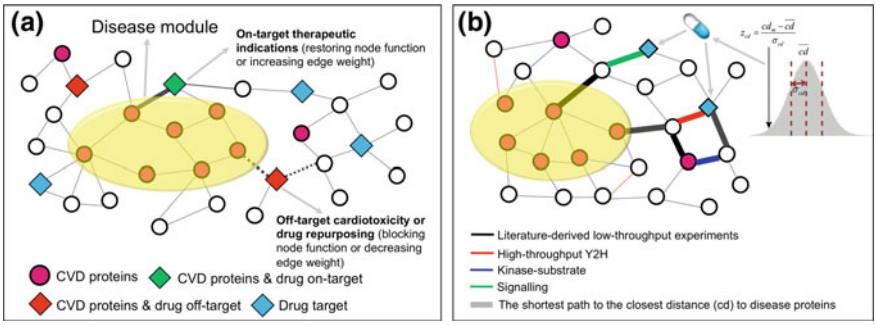

Fig. 5.1 A biological hypothesis for the network proximity approach. **a** A proposed network-based hypothesis of drug cardiotoxicity under the human protein-protein interactome model. Drug targets representing nodes within cellular networks are often intrinsically coupled in both therapeutic and adverse effects (e.g., cardiotoxicity). We, therefore, asserted that for a drug with multiple targets to be on-target effective for a disease or to cause off-target cardiotoxicity, its target proteins should be within or in the immediate vicinity of the corresponding cardiovascular disease module; **b** A diagram illustrating network proximity that quantifies the interplay between disease modules and drug targets on the carefully curated human protein-protein interactome

development of efficacious therapies from an integrated context using informatics tools and experimental pharmacology approaches, offering an innovative way to identify actionable biomarkers to predict and prevent cancer treatment-related cardiotoxicities. In the past few years, we have demonstrated that systems pharmacology and network-based approaches offered possibilities for identifying novel therapeutic targets, disease pathways, and network modules in cancer [25–47], cardiovascular disease [48], pulmonary fibrosis [49], and infectious disease [50, 51]. However, traditional gene-overlap approaches and machine learning-based approaches [52] often have potential limitations in understanding drug mechanism-of-action (MoA) owing to data incompleteness, literature data bias, and the complexities of human cellular systems.

Novel network approaches, such as a network-based drug-disease proximity that sheds light on the relationship between drugs (e.g., drug targets) and diseases (e.g., molecular disease determinants in disease modules within the human interactome), offer powerful tools for efficient screening of potentially new indications for approved drugs, or for previously unidentified adverse events [48]. In this chapter, we will introduce an integrated, network-based, systems pharmacology approach that we recently developed [48]. Specifically, this network approach incorporates disease-associated proteins/genes, drug-target networks, and the human protein-protein interactome, for efficient risk assessment of drug-induced cardiotoxicities. We will showcase how to use network proximity to identify the underlying mechanisms-of-action of cardiotoxicities induced by various oncological drugs (e.g., multi-targeted kinase inhibitors). Finally, we will discuss several existing challenges and highlight future directions of network proximity approaches for comprehensive risk assessment of drug-induced

cardiotoxicities in the early stage of drug discovery, clinical trials, and post-marketing surveillance.

5.2 Method and Materials

5.2.1 Computers

1. Computer requirements, laptop/desktop computers, or high-performance computing clusters with UNIX/LINUX operating systems.
2. Network analysis and visualization tools, Cytoscape (https://cytoscape.org/) or Gephi (https://gephi.org/).
3. Python, Java, or other environments.

5.2.2 Reconstruction of Drug-Target Interaction Network

The drug-target network can be described as a bipartite graph $G(D, T, P)$, where the drug set denotes as $D = \{d_1, d_2, \ldots, d_n\}$, target set as $T = \{t_1, t_2, \ldots, t_m\}$, and interaction set as $P = \left\{p_{ij}, d_i \in D, t_j \in T\right\}$. An interaction is drawn between d_i and t_j when drug d_i binds with target t_j with binding affinity (such as IC_{50}, K_i, or K_d) less than a given threshold value. Mathematically, a drug-target bipartite network can be presented by an $n \times m$ adjacent matrix $\left\{p_{ij}\right\}$, where $p_{ij} = 1$ if the binding affinity between d_i and t_j is less than 10 μM, otherwise $p_{ij} = 0$, as described in Eq. (5.1).

$$p_{ij} = \begin{cases} 1 \ IC_{50}(K_i) \leq 10\,\mu M \\ 0 \ IC_{50}(K_i) > 10\,\mu M \end{cases} \tag{5.1}$$

In general, we can collect drug-target interaction information from the DrugBank database (v4.3) [53], the Therapeutic Target Database (TTD) [54], and the PharmGKB database [55]. Specifically, bioactivity data for drug-target pairs can be collected from ChEMBL (v20) [56], BindingDB [57], and IUPHAR/BPS Guide to PHARMACOLOGY [58]. Detailed drug-target databases are provided in Table 5.1. To improve the quality of data, we usually focus on physical drug-target interactions based on the following three criteria: (i) The human target is represented by a unique UniProt accession number; (ii) the target is marked as 'reviewed' in the UniProt database [59]; and (iii) binding affinities, including K_i, K_d, $IC_{50,}$ or EC_{50} each $\leq$10 μM. In addition, we can also build functional drug-gene association networks from drug-induced transcriptomics data or proteomics data derived from human cells (Table 5.1).

Table 5.1 Summary of chemoinformatics and bioinformatics resources for re-constructing drug-target network

Name of databases	Description	The number of interactions, drugs, and targets	Web site	References
Section 1. Databases for collecting physical drug-target interactions				
ChEMBL	Chemical properties and biological activities of drug-like molecules	2,036,512 compounds against 11,224 targets and 14,371,197 bioactivity records	https://www.ebi.ac.uk/chembldb	[98]
BindingDB	Binding affinities of proteins with small drug-like ligands	565,136 compounds against 6612 proteins and 1,279,670 binding affinity data	http://www.bindingdb.org	[99]
PubChem	Repository of small molecule biological activities	More than 230 million bioactivities connecting 9.3 million compounds and 9851 targets	http://puchem.ncbi.nlm.nih.gov	[100]
DrugBank	Detailed drug data with comprehensive target information	8250 drug entries including 2016 FDA-approved small molecule drugs and over 6000 experimental drugs	http://www.drugbank.ca/	[53]
TTD	Information on the therapeutic targets	31,614 drugs and 2589 targets	http://bidd.nus.edu.sg/group/ttd	[54]
DGIdb	The drug-gene interaction database	26,298 unique drug-gene interactions connecting 7569 drugs and 7524 unique genes	http://dgidb.genome.wustl.edu/	[101]

(continued)

Table 5.1 (continued)

Name of databases	Description	The number of interactions, drugs, and targets	Web site	References
STITCH	Experimental and predicted compound-protein interactions	1.6 billion interactions connecting 0.5 million compounds and 9.6 million proteins from 2031 organisms	http://stitch.embl.de/	[102]
SuperPred	Experimental and predicted compound-protein interactions	341,000 compounds, 1800 targets and 665,000 compound-target interactions	http://prediction.charite.de/index.php	[103]
Section 2. Databases for collecting functional drug-gene interactions				
CMap (v 2.0)		Gene-expression signatures to connect 1309 small molecules and ~7000 genes in 4 cancer cell lines	https://www.broadinstitute.org/cmap	[104]
LINCS		Library of integrated network-based cellular signatures for ~1 million of gene expression profiles	http://www.lincscloud.org/	[105]
open TG-GATEs		A large-scale toxicogenomics database	http://toxico.nibio.go.jp/english/index.html	[106]
DrugMatrix		Molecular toxicology reference database and informatics system	http://ntp.niehs.nih.gov/drugmatrix/index.html	[107]

5.2.3 Reconstruction of the Human Protein-Protein Interactome

There are several experimental strategies for mapping protein-protein interactions (PPIs), such as yeast two-hybrid assay (Y2H) that measures direct physical interactions in cells and affinity purification mass spectrometry that measure the composition of protein complexes. Specifically, we can reconstruct the human protein-protein interactome network by assembling various publicly available PPI data: (i) binary, physical PPIs tested by high-throughput Y2H systems from public available high-quality Y2H datasets [60, 61]; (ii) High-quality PPIs from the published protein structure databases, such as Interactome3D [62], Instruct [63], and Interactome INSIDER [64]; (iii) kinase-substrate interactions by literature-derived low-throughput and high-throughput experiments from KinomeNetworkX [65], Human Protein Resource Database (HPRD) [66], PhosphoNetworks [67, 68], PhosphositePlus [69], DbPTM 3.0 [70], and Phospho. ELM [71]; (iv) signaling network by literature-derived low-throughput experiments as annotated in SignaLink2.0 [72]; (v) protein complexes data identified by a robust affinity purification mass spectrometry methodology collected from BioPlex V2.0 [73]; and (vi) carefully literature-curated PPIs identified by affinity purification followed by mass spectrometry (AP-MS) and by literature-derived low-throughput experiments from BioGRID [74], PINA [75], HPRD [66], MINT [76], IntAct [77], and InnateDB [78]. The detailed bioinformatics resources for human protein-protein interactions are provided in Table 5.2.

5.2.4 Collection of Disease-Associated Genes/Proteins

In general, we can integrate disease-gene annotation data from multiple commonly used bioinformatics resources currently available (Table 5.2).

OMIM, The OMIM database (Online Mendelian Inheritance in Man, http://www.omim.org/) [79] is a comprehensive collection covering literature-curated human disease genes with high-quality experimental validation evidence.

CTD, The Comparative Toxicogenomics Database (http://ctdbase.org/) [80] provides information about interactions between chemicals and gene products, and their association with various diseases. Here, only manually curated gene-disease interactions from the literature were used.

ClinVar, ClinVar is a public archive of relationships among sequence variation and various human phenotypes (https://www.ncbi.nlm.nih.gov/clinvar/) [81]. To improve the data quality, only clinically significant relationships among variants and disease traits annotated in ClinVar can be used.

GWAS Catalog, The NHGRI-EBI Catalog of published genome-wide association studies (GWAS, https://www.ebi.ac.uk/gwas/) [82] provides unbiased (single-nucleotide polymorphism) SNP-trait associations with genome-wide significance. Usually, a SNP-trait with genome-wide significance ($p < 5 \times 10^{-8}$) will be used.

Table 5.2 Summary of representative bioinformatics resources and network tools for building disease modules from the human protein-protein interactome

Name of databases	Description	Webs	References
Section 1. Systems biology resources			
BioGRID	Integrated protein-protein interaction data	http://thebiogrid.org	[74]
HPRD	Human protein-protein interaction data	http://www.hprd.org	[108]
Interactome3D	Manually curated PPIs with known three-dimensional structure information	http://interactome3d.irbbarcelona.org	[62]
STRING	Functional protein association networks database	http://string-db.org	[109]
MINT	Protein-protein interactions in refereed journals	http://mint.bio.uniroma2.it/mint	[76]
KinomeNetworkX	An integrative kinase-substrate database		[65]
PhosphoNetworks	A high-resolution phosphorylation network connects the specific phosphorylation sites present in substrates with their upstream kinases	http://www.phosphonetworks.org/	[67, 68]
PhosphositePlus	A database and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, acetylation, and more	https://www.phosphosite.org/homeAction.action	[69]
Section 2. Systems biology resources			
OMIM	A comprehensive collection covering literature-curated human disease genes with experimental evidence	http://www.omim.org/	[79]
CTD	A database containing literature-curated interactions connecting chemical, genes, and diseases	http://ctdbase.org/	[80]
ClinVar	A public archive of relationships among sequence variation and various human phenotypes	https://www.ncbi.nlm.nih.gov/clinvar/	[81]

(continued)

Table 5.2 (continued)

Name of databases	Description	Webs	References
GWAS Catalog	A database contains unbiased SNP-trait associations with genome-wide significance	https://www.ebi.ac.uk/gwas/	[82]
GWASdb	A data curation and knowledge database for SNP-trait associations from GWAS for PubMed	http://jjwanglab.org/gwasdb	[83]
PheWAS Catalog	A catalog contains SNP-trait associations identified by the phenome-wide association study (PheWAS)	http://phewas.mc.vanderbilt.edu	[84]
HuGE navigator	An integrated disease candidate gene database based on the core data from PubMed abstracts using text mining algorithms	https://phgkb.cdc.gov/PHGKB/	[85]
DisGeNET	A disease-gene database by assembling expert-curated databases and text-mined data	http://www.disgenet.org/	[86]
Section 3. Network analysis and visualization tools			
Cytoscape	An open-source software platform for visualizing complex networks	https://cytoscape.org/	[110]
Gephi	An open graph visualization platform	https://gephi.org/	[111]

GWASdb, GWASdb includes a more comprehensive data curation and knowledge integration for SNP-trait associations from GWAS for PubMed and other resources [83]. For example, the curated moderate SNP-trait associations ($p < 1.0 \times 10^{-3}$) are annotated in GWASdb. However, this low threshold of SNP-trait associations ($p < 1.0 \times 10^{-3}$) often causes potential false positive rate.

PheWAS Catalog, The PheWAS Catalog contains SNP-trait associations identified by the phenome-wide association study (PheWAS) paradigm within electronic medical records, an unbiased approach to replication and discovery that interrogates relationships between targeted genotypes and multiple phenotypes [84]. Thus, the PheWAS Catalog may complement data incompleteness from traditional GWAS.

HuGE Navigator, The HuGE Navigator is an integrated disease candidate gene database based on the core data from PubMed abstracts using text mining algorithms [85]. To keep the data quality, the literature-reported disease-gene annotation data with known PubMed IDs from HuGE Navigator are often used.

DisGeNET, DisGeNET is a comprehensive database for collecting disease-associated genes [86]. In October 26, 2018, DisGeNET contains over 561,119 associations connecting 17,074 genes and over 20,000 diseases, disorders, and traits by integrating expert-curated databases and text-mined data.

To improve the data quality during data integration, medical terms of diseases, disorders, and traits are often annotated by Medical Subject Headings (MeSH) and Unified Medical Language System (UMLS) vocabularies (https://www.nlm.nih.gov/mesh/MBrowser.html) [87]. In addition, protein-coding genes are further annotated by gene Entrez ID, chromosome location, and the official gene symbols from the National Center for Biotechnology Information (NCBI) database [88]. A detailed description of disease-gene annotation data integration is provided in Table 5.2.

5.2.5 *Network Proximity*

Given the set of drug targets, X, and the set of disease proteins, Y, we can calculate the network topological distance $d(x, y)$ between nodes x and y in the human protein-protein interactome. In general, there are four different distance measures that take into account the path lengths between drug targets (X) and the set of disease proteins (Y): (a) the closest measure, representing the average shortest path length between targets of X and the nearest proteins of Y; (b) the shortest measure, representing the average shortest path length among all targets of drugs; (c) the kernel measure, down-weighting longer paths via an exponential penalty; and (d) the center measure, representing the shortest path length among all targets of drugs with the greatest closeness centrality among proteins in X and Y. We define those four distance measures in Eqs. (5.2–5.6).

$$\text{Closest, } \langle \text{cd}_{XY} \rangle = \frac{1}{\|X\| + \|Y\|} \left(\sum_{x \in X} \min_{y \in Y} d(x, y) + \sum_{y \in Y} \min_{x \in X} d(x, y) \right) \tag{5.2}$$

$$\text{Shortest, } \langle \text{sd}_{XY} \rangle = \frac{1}{\|X\| + \|Y\|} \sum_{x \in X, y \in Y} d(x, y) \tag{5.3}$$

$$\text{Kernel, } \langle \text{kd}_{XY} \rangle = \frac{-1}{\|X\| + \|Y\|} \left(\sum_{x \in X} \ln \sum_{y \in Y} \frac{\text{e}^{-(d(x,y)+1)}}{\|Y\|} + \sum_{y \in Y} \ln \sum_{x \in X} \frac{\text{e}^{-(d(x,y)+1)}}{\|X\|} \right) \tag{5.4}$$

$$\text{Centre, } \left\langle \text{cd}'_{XY} \right\rangle = d(\text{centre}_X, \text{centre}_Y) \tag{5.5}$$

where centre_Y, the topological center of X, is defined as

$$\text{centre}_Y = \text{argmin}_{u \in Y} \sum_{y \in Y} d(y, u) \tag{5.6}$$

If the centre$_X$ or centre$_Y$ is not unique, all the nodes in centre$_X$ or centre$_Y$ are used to define the center, and shortest path lengths between these nodes are averaged. If the centre$_Y$ is not unique, all nodes are used to define the center and the shortest path lengths to these nodes are averaged.

Finally, the significance of the measure is evaluated by comparing to the reference distance distribution corresponding to the expected network topological distance between two randomly selected groups of proteins matched to size and degree (connectivity) distribution as the original disease proteins and drug targets in the human interactome. This procedure was repeated 1000 times. As illustrated in Fig. 5.1a, for closest distance measure (cd in Eq. 5.2), the mean distance ($\overline{\text{cd}}$) and standard deviation (σ_{cd}) of the reference distribution are used to calculate a z-score (z) by converting an observed closest distance to a normalized distance using Eq. (5.7).

$$z_{\text{cd}} = \frac{\text{cd} - \overline{\text{cd}}}{\sigma_{\text{cd}}} \tag{5.7}$$

A detailed description for z-score calculation can be found in our recent study [48].

5.2.6 *Network Visualization and Statistical Analysis*

Networks can be analyzed and visualized by Cytoscape (v3.2.0, http://www.cytoscape.org/) and Gephi (v0.9.2, https://gephi.org/). Statistical analysis can be performed by the Python (v3.2, http://www.python.org/) or R platforms (v3.01, http://www.r-project.org/).

5.3 Results/Case Studies

The basis for the proposed network-based methodology in this chapter rests on the notion that the proteins that associate with and functionally govern a disease phenotype are localized in the corresponding disease module or subnetwork (graph) within the comprehensive human protein-protein interactome network [48]. As shown in Fig. 5.2, our preliminary network analysis reveals that cardiomyopathy-associated proteins form a statistically, significantly clustered, distinct module in the human protein-protein interactome, as we have previously shown for 23 other types of cardiovascular outcomes as well [48].

To examine drug effects on the cardiovascular system, our previous study [48] quantifies the interplay between diseases and drug targets on the human protein-protein interactome using a network proximity measure (Fig. 5.1b). Figure 5.3 reveals the globally predicted drug-disease network using z-score (z) <-4.0, which connects 23 CV events and approximately 600 FDA-approved non-CVD drugs grouped by

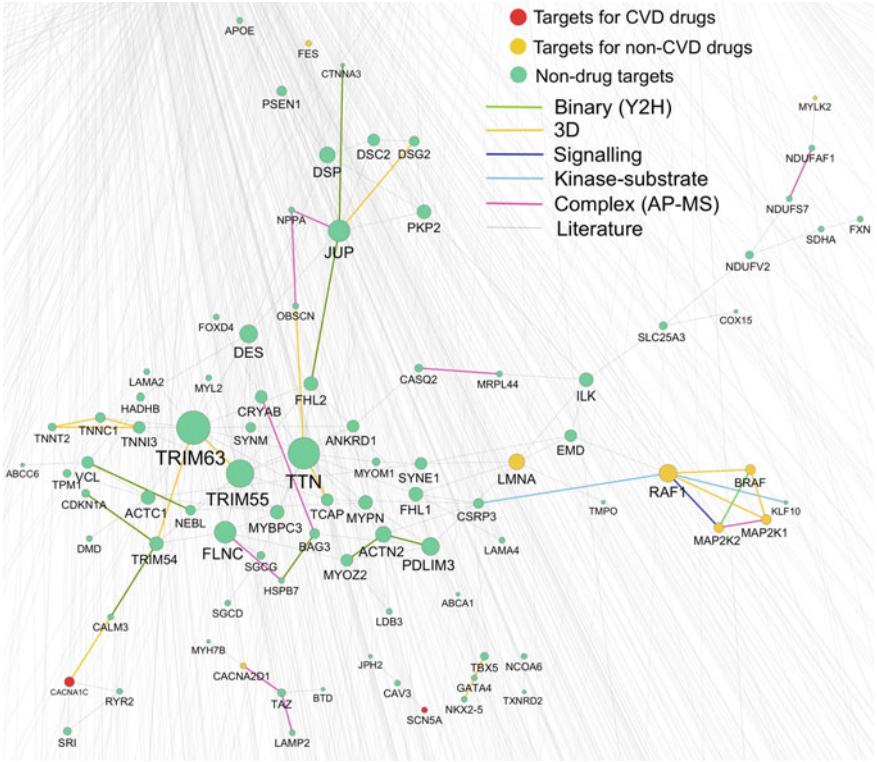

Fig. 5.2 Subnetwork of the full protein-protein interaction (PPI) network highlighting the disease module for cardiomyopathy (CM). CM gene-coding proteins are grouped by targets for known cardiovascular disease (CVD) drugs (red) or non-CVD drugs (yellow), and non-drug targets (green), collected from OMIM data [79] as shown in Table 5.2. The PPIs are labeled by six different types of experimental evidence, which served as the basis for constructing the PPI. Background light gray lines represent other edges in the dense PPI unrelated to the CM disease module. Nodes at the bottom are unconnected to the module, likely owing to the incompleteness of the PPI. Networks were visualized by the spring-embedded layout algorithm in Cytoscape (Table 5.2)

the first-level anatomical therapeutic chemical (ATC) classification system codes as described previously [48]. This network of drug effects on the cardiovascular system offers unexpected opportunities in identifying previously unrecognized associations between drugs and cardiovascular outcomes. To be specific, we examined predicted drug-disease pairs for non-cardiovascular drugs across different drug categories defined by the first-class ATC codes (Fig. 5.3). We found that FDA-approved drugs often generated effects on the cardiovascular system, such as drugs that affect the alimentary tract and metabolism [A], and antineoplastic and immunomodulating agents [L]. For example, previous studies have suggested that comorbidity between CVD and cancer is typically associated with various cytotoxic chemotherapies, such as anthracyclines (e.g., doxorubicin) [89]. Figure 5.4 shows that doxorubicin is pre-

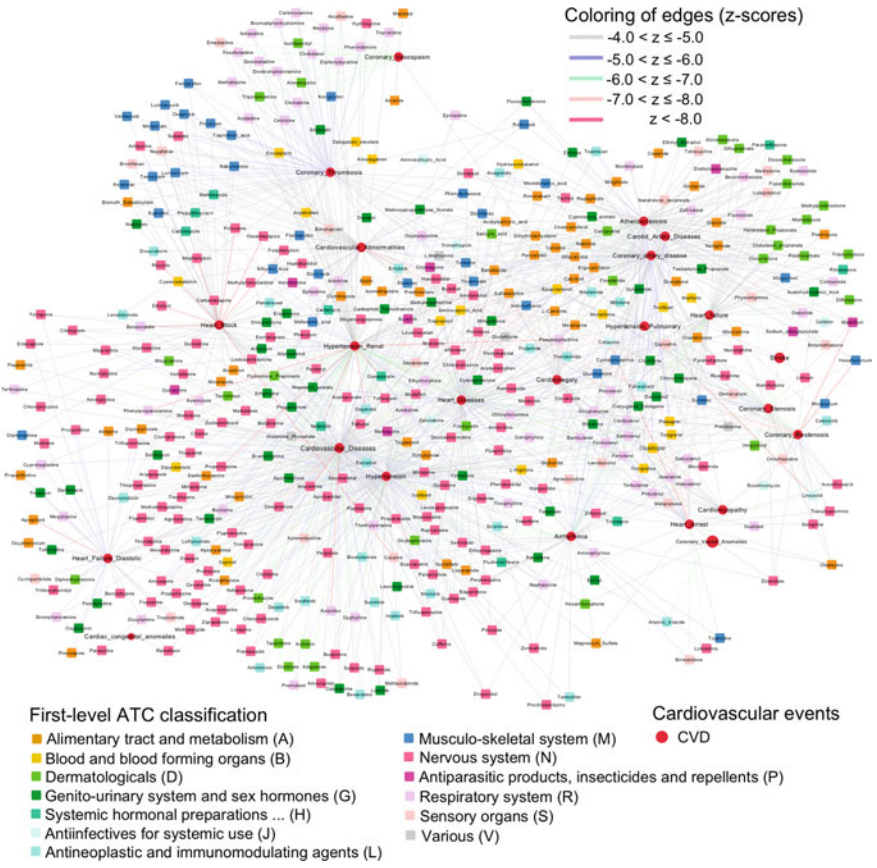

Fig. 5.3 Network-predicted drug-disease associations. The globally predicted drug-disease association network, connecting 23 types of cardiovascular (CV) events (red circles) and 707 FDA-approved non-CV drugs (squares) [48]. The edges between drugs and diseases represent the predicted z-score <−4.0. Drugs are colored by the first-level anatomical therapeutic chemical (ATC) classification system codes from the DrugBank data (Table 5.1)

dicted to have a significant association with multiple types of cardiovascular events, such as heart block ($z = -5.06$), cardiovascular abnormalities ($z = -4.04$), arrhythmia ($z = -3.46$), CAD ($z = -2.71$), and heart failure ($z = -2.61$). Similar trends were observed for daunorubicin (Fig. 5.4).

Recent studies have suggested that molecularly targeted cancer therapies (e.g., multiple-target kinase inhibitors) often cause cardiotoxicities as well [16–18, 23, 24]. Figure 5.4 shows the significant associations of multiple cardiovascular events with several multiple-target kinase inhibitors, such as sorafenib ($z = -7.51$), dasatinib ($z = -6.37$), sunitinib ($z = -6.27$), and nilotinib ($z = -5.54$), and predictions are consistent with several recent clinical reports [18, 90]. Interestingly, imatinib, the first approved targeted agent for the treatment of chronic myeloid leukemia,

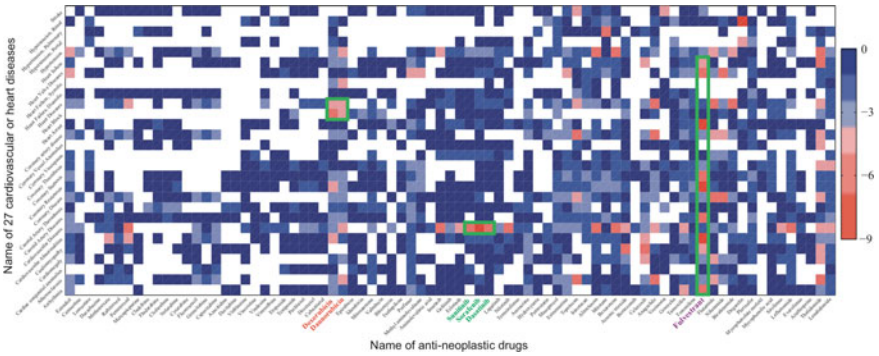

Fig. 5.4 Network-predicted cardiotoxicities induced by approved oncological drugs. In total, network-predicted associations (*Z*-scores showing the color key) between 79 approved anti-cancer drugs and 27 cardiovascular diseases are shown by a network proximity approach [48]. The lower negative z-scores show more significant associations

was reported to target over 40 different human proteins (Fig. 5.5), which associates multiple cardiac side effects [23, 24]. However, imatinib was also reported to have potentially therapeutic effect on pulmonary hypertension in clinical studies (ClinicalTrials, NCT00902174 and NCT00477269) [91, 92]. Drug-target network analysis reveals that inhibition on platelet-derived growth factor (PDGFRA [platelet-derived growth factor receptor A] and PDGFRB [beta-type platelet-derived growth factor receptor]) and KIT (KIT proto-oncogene receptor tyrosine kinase) by imatinib may contribute to its potentially therapeutic effects on pulmonary hypertension [93, 94]. Preliminary drug-target network analysis from the human protein-protein interactome thereby offers potential underlying mechanism-of-action of imatinib on cardiovascular systems (Fig. 5.5). However, mechanistic pre-clinical and clinical studies are warranted. Fulvestrant, a recently FDA-approved drug for the treatment of hormone receptor-positive metastatic breast cancer, was predicted to associate significantly with multiple cardiovascular events, such as coronary restenosis ($z = -7.86$), cardiovascular abnormalities ($z = -7.60$), cardiac arrest ($z = -7.38$), arrhythmia ($z = -5.78$), and heart failure ($z = -4.14$), indicating the importance of evaluating potential cardiotoxicities during fulvestrant treatment. Those new significant associations among CVD and antineoplastic drugs identified by network proximity analysis offer a useful resource for characterizing the pharmacologic underpinnings of cardio-oncology [89].

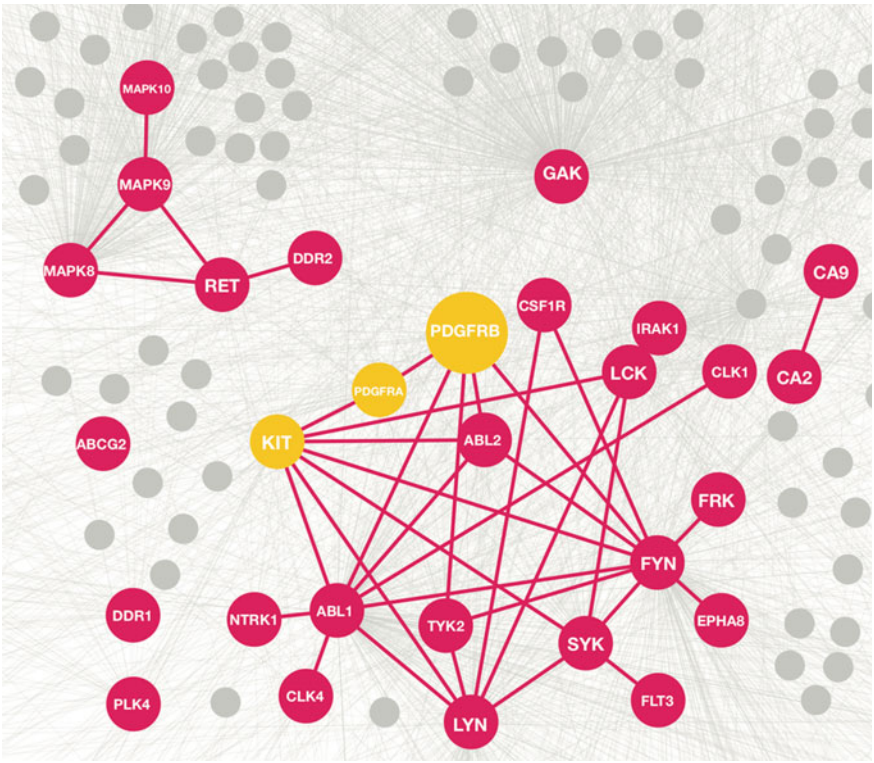

Fig. 5.5 Proposed mechanistic model of imatinib-induced cardiotoxicity. A highlighted subnetwork shows the predicted mechanism-of-action for imatinib-induced cardiotoxicity by drug-target network analysis under the human protein-protein interactome model. Imatinib's targets were collected from the DrugBank database [53]. The human protein-protein interactions and the known cardiovascular disease genes were collected from bioinformatics resources as described in Table 5.2

5.4 Conclusion and Future Directions

In this chapter, we introduce a network-based, systems pharmacology approach for quantifying drug-disease associations under the human protein-protein interactome network model as developed recently [48]. This network-based approach can be used for drug repurposing or for risk assessment of drug-induced cardiotoxicity. We showcased that this network-based, systems pharmacology approach can be used to identify well-known cardio-toxic chemotherapeutic agents and novel oncological drug-induced cardiotoxicities.

We acknowledge several potential limitations in the current systems pharmacology framework. The total size of the human protein-protein interactome is estimated to be ~650,000 interactions [95]. Data incompleteness of the human interactome and known drug-target networks should be of concern. In addition, potential literature

bias of the human interactome warrants inspection in the future. Second, our current network-based in silico models cannot separate therapeutic effects from side effects owing to the lack of detailed functional effects of drug targets and disease proteins. Drug targets representing nodes within cellular networks are often intrinsically coupled in both therapeutic and adverse profiles. For example, drugs can inhibit or activate protein functions (including antagonists vs. agonists), while disease alleles from genetic or genomic studies contain loss-of-function or gain-of-function. An inhibitor that targets loss-of-function disease proteins often causes adverse effects. In addition, dose-dependent cardiotoxicity cannot be evaluated by the current network-based systems pharmacology framework. Finally, translating network-based prediction to regulatory science during drug discovery and development remains challenging.

There are several new directions to improve network proximity approach further. Adding genome/proteome-wide drug-induced transcriptome or proteome data such as the Connectivity Map may help overcome data incompleteness of known targets on approved drugs (Table 5.1). In addition, integration of functional genomic assays or large-scale disease gene expression profiles (upregulation or downregulation), along with patient data validation and in vitro or in vivo mechanistic studies will improve network proximity approaches further. Utilizing network proximity approaches to investigate the metabolic intervention and dietary regulation may offer novel chemical intervention strategies for cancer treatment-related cardiotoxicities. In addition, implementing dynamics data (e.g., time series drug-protein binding affinity [k_{on} and k_{off}]) via network control approaches [96, 97] and pharmacokinetics-based mathematical modeling into the network-based systems pharmacology framework can be used to assess dose-dependent cardiotoxicities. Finally, assembling multi-omics data, including genomics, transcriptomics, and proteomics from individual patients, under a network proximity framework, may offer novel actionable biomarkers for characterization of heterogeneities of cancer treatment-induced cardiotoxicities, in the personalized cardio-oncology era.

Acknowledgements This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number K99HL138272 and R00HL138272.
Competing Interests The author has declared that no conflict of interest exists.

References

1. Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016. CA Cancer J Clin 66(1):7–30
2. Brown SA, Sandhu N, Herrmann J (2015) Systems biology approaches to adverse drug effects, the example of cardio-oncology. Nat Rev Clin Oncol 12(12):718–731
3. Cheng F, Loscalzo J (2018) Pulmonary comorbidity in lung cancer. Trends Mol Med 24(3):239–241
4. Cheng F, Nussinov R (2018) KRAS activating signaling triggers arteriovenous malformations. Trends Biochem Sci 43(7):481–483
5. Pullamsetti SS, Kojonazarov B, Storn S, Gall H, Salazar Y, Wolf J et al (2017) Lung cancer-associated pulmonary hypertension, Role of microenvironmental inflammation based on tumor cell-immune cell cross-talk. Sci Transl Med 9(416):eaai9048

6. Lenneman CG, Sawyer DB (2016) Cardio-oncology, an update on cardiotoxicity of cancer-related treatment. Circ Res 118(6):1008–1020
7. Han X, Zhou Y, Liu W (2017) Precision cardio-oncology, understanding the cardiotoxicity of cancer therapy. NPJ Precis Oncol 1(1):31
8. Saiki H, Petersen IA, Scott CG, Bailey KR, Dunlay SM, Finley RR et al (2017) Risk of heart failure with preserved ejection fraction in older women after contemporary radiotherapy for breast cancer. Circulation 135(15):1388–1396
9. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M et al (2011) Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med 365(14):1273–1283
10. Sharma A, Burridge PW, McKeithan WL, Serrano R, Shukla P, Sayed N et al (2017) High-throughput screening of tyrosine kinase inhibitor cardiotoxicity with human induced pluripotent stem cells. Sci Transl Med 9(377):eaaf2584
11. Cheng F, Loscalzo J (2017) Autoimmune cardiotoxicity of cancer immunotherapy. Trends Immunol 38(2):77–78
12. Johnson DB, Balko JM, Compton ML, Chalkias S, Gorham J, Xu Y et al (2016) Fulminant myocarditis with combination immune checkpoint blockade. N Engl J Med 375(18):1749–1755
13. Ranchoux B, Gunther S, Quarck R, Chaumais MC, Dorfmuller P, Antigny F et al (2015) Chemotherapy-induced pulmonary hypertension, role of alkylating agents. Am J Pathol 185(2):356–371
14. Chien KR (2006) Herceptin and the heart–a molecular modifier of cardiac failure. N Engl J Med 354(8):789–790
15. Cheng F (2018) In silico oncology drug repositioning and polypharmacology. Methods Mol Biol 1878:243–261
16. Elkins JM, Fedele V, Szklarz M, Abdul Azeez KR, Salah E, Mikolajczyk J et al (2016) Comprehensive characterization of the published kinase inhibitor set. Nat Biotechnol 34(1):95–103
17. Knight ZA, Lin H, Shokat KM (2010) Targeting the cancer kinome through polypharmacology. Nat Rev Cancer 10(2):130–137
18. Li W, Croce K, Steensma DP, McDermott DF, Ben-Yehuda O, Moslehi J (2015) Vascular and metabolic implications of novel targeted cancer therapies, focus on kinase inhibitors. J Am Coll Cardiol 66(10):1160–1178
19. Mullard A (2017) 2016 FDA drug approvals. Nat Rev Drug Discov 16(2):73–76
20. Loscalzo J, Barabasi AL, Silverman EK (2017) Network medicine, complex systems in human disease and therapeutics. Publisher, Harvard University Press Publication Date 2/1/2017. ISBN 9780674436534
21. Antman EM, Loscalzo J (2016) Precision medicine in cardiology. Nat Rev Cardiol 13(10):591–602
22. Greene JA, Loscalzo J (2017) Putting the patient back together—social medicine, network medicine, and the limits of reductionism. N Engl J Med 377(25):2493–2499
23. Kerkela R, Grazette L, Yacobi R, Iliescu C, Patten R, Beahm C et al (2006) Cardiotoxicity of the cancer therapeutic agent imatinib mesylate. Nat Med 12(8):908–916
24. Distler JH, Distler O (2007) Cardiotoxicity of imatinib mesylate, an extremely rare phenomenon or a major side effect? Ann Rheum Dis 66(6):836
25. Cheng F, Hong H, Yang S, Wei Y (2017) Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. Brief Bioinform 18(4):682–697
26. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G et al (2012) Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol 8(5):e1002503
27. Cheng F, Liu C, Lin CC, Zhao J, Jia P, Li WH et al (2015) A gene gravity model for the evolution of cancer genomes, a study of 3000 cancer genomes across 9 cancer types. PLoS Comput Biol 11(9):e1004497
28. Cheng F, Jia P, Wang Q, Lin CC, Li WH, Zhao Z (2014) Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. Mol Biol Evol 31(8):2156–2169

29. Cheng F, Liu C, Shen B, Zhao Z (2016) Investigating cellular network heterogeneity and modularity in cancer, a network entropy and unbalanced motif approach. BMC Syst Biol 10(Suppl 3):65
30. Cheng F, Zhao J, Fooksa M, Zhao Z (2016) A network-based drug repositioning infrastructure for precision cancer medicine through targeting significantly mutated genes in the human cancer genomes. J Am Med Inform Assoc 23(4):681–691
31. Cheng F, Zhao J, Hanker AB, Brewer MR, Arteaga CL, Zhao Z (2016) Transcriptome- and proteome-oriented identification of dysregulated eIF4G, STAT3, and Hippo pathways altered by PIK3CA (H1047R) in HER2/ER-positive breast cancer. Breast Cancer Res Treat 160(3):457–474
32. Fang J, Cai C, Wang Q, Lin P, Zhao Z, Cheng F (2017) Systems pharmacology-based discovery of natural products for precision oncology through targeting cancer mutated genes. CPT Pharmacometrics Syst Pharmacol 6(3):177–187
33. Fang J, Liu C, Wang Q, Lin P, Cheng F (2018) In silico polypharmacology of natural products. Brief Bioinform 19(6):1153–1171
34. Fang J, Wu Z, Cai C, Wang Q, Tang Y, Cheng F (2017) Quantitative and systems pharmacology. 1. In silico prediction of drug-target interactions of natural products enables new targeted cancer therapy. J Chem Inf Model 57(11):2657–2671
35. Jiang X, Lu W, Shen X, Wang Q, Lv J, Liu M et al (2018) Repurposing sertraline sensitizes non-small cell lung cancer cells to erlotinib by inducing autophagy. JCI Insight 3(11):e98921
36. Li J, Lei K, Wu Z, Li W, Liu G, Liu J et al (2016) Network-based identification of microRNAs as potential pharmacogenomic biomarkers for anticancer drugs. Oncotarget 7(29):45584–45596
37. Lu W, Cheng F, Yan W, Li X, Yao X, Song W et al (2017) Selective targeting p53(WT) lung cancer cells harboring homozygous p53 Arg72 by an inhibitor of CypA. Oncogene 36(33):4719–4731
38. Shen Q, Cheng F, Song H, Lu W, Zhao J, An X et al (2017) Proteome-scale investigation of protein allosteric regulation perturbed by somatic mutations in 7000 cancer genomes. Am J Hum Genet 100(1):5–20
39. Wang J, Hu K, Guo J, Cheng F, Lv J, Jiang W et al (2016) Suppression of KRas-mutant cancer through the combined inhibition of KRAS with PLK1 and ROCK. Nat Commun 7:11363
40. Wu D, Wang W, Chen W, Lian F, Lang L, Huang Y et al (2018) Pharmacologic inhibition of dihydroorotate dehydrogenase induces apoptosis and differentiation in acute myeloid leukemia cells. Haematologica 103(9):1472–1483
41. Wu Z, Cheng F, Li J, Li W, Liu G, Tang Y (2017) SDTNBI, an integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. Brief Bioinform 18(2):333–347
42. Wu Z, Lu W, Yu W, Wang T, Li W, Liu G et al (2018) Quantitative and systems pharmacology 2. In silico polypharmacology of G protein-coupled receptor ligands via network-based approaches. Pharmacol Res 129:400–413
43. Wu Z, Lu W, Wu D, Luo A, Bian H, Li J et al (2016) In silico prediction of chemical mechanism of action via an improved network-based inference method. Br J Pharmacol 173(23):3372–3385
44. Yu W, Lu W, Chen G, Cheng F, Su H, Chen Y et al (2017) Inhibition of histone deacetylases sensitizes EGF receptor-TK inhibitor-resistant non-small-cell lung cancer cells to erlotinib in vitro and in vivo. Br J Pharmacol 174(20):3608–3622
45. Zhao J, Cheng F, Wang Y, Arteaga CL, Zhao Z (2016) Systematic prioritization of druggable mutations in approximately 5000 genomes across 16 cancer types using a structural genomics-based approach. Mol Cell Proteomics 15(2):642–656
46. Zhao J, Cheng F, Zhao Z (2017) Tissue-specific signaling networks rewired by major somatic mutations in human cancer revealed by proteome-wide discovery. Cancer Res 77(11):2810–2821
47. Zhao Y, Hu Q, Cheng F, Su N, Wang A, Zou Y et al (2015) SoNar, a highly responsive NAD+/NADH sensor, allows high-throughput metabolic screening of anti-tumor agents. Cell Metab 21(5):777–789

48. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabasi AL et al (2018) Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun 9(1):2691
49. Lu W, Yao X, Ouyang P, Dong N, Wu D, Jiang X et al (2017) Drug repurposing of histone deacetylase inhibitors that alleviate neutrophilic inflammation in acute lung injury and idiopathic pulmonary fibrosis via inhibiting leukotriene A4 hydrolase and blocking LTB4 biosynthesis. J Med Chem 60(5):1817–1828
50. Cheng F, Murray JL, Zhao J, Sheng J, Zhao Z, Rubin DH (2016) Systems biology-based investigation of cellular antiviral drug targets identified by gene-trap insertional mutagenesis. PLoS Comput Biol 12(9):e1005074
51. Cheng F, Murray JL, Rubin DH (2016) Drug repurposing, new treatments for Zika virus infection? Trends Mol Med 22(11):919–921
52. Cai C, Fang J, Guo P, Wang Q, Hong H, Moslehi J et al (2018) In silico pharmacoepidemiologic evaluation of drug-induced cardiovascular complications using combined classifiers. J Chem Inf Model 58(5):943–956
53. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y et al (2014) DrugBank 4.0, shedding new light on drug metabolism. Nucleic Acids Res 42(Database issue):D1091–1097
54. Yang H, Qin C, Li YH, Tao L, Zhou J, Yu CY et al (2016) Therapeutic target database update 2016, enriched resource for bench to clinical drug target and targeted pathway information. Nucleic Acids Res 44(D1):D1069–D1074
55. Hernandez-Boussard T, Whirl-Carrillo M, Hebert JM, Gong L, Owen R, Gong M et al (2008) The pharmacogenetics and pharmacogenomics knowledge base, accentuating the knowledge. Nucleic Acids Res 36(Database issue):D913–918
56. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A et al (2012) ChEMBL, a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40(D1):D1100–D1107
57. Liu TQ, Lin YM, Wen X, Jorissen RN, Gilson MK (2007) BindingDB, a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 35:D198–D201
58. Pawson AJ, Sharman JL, Benson HE, Faccenda E, Alexander SPH, Buneman OP et al (2014) The IUPHAR/BPS Guide to PHARMACOLOGY, an expert-driven knowledgebase of drug targets and their ligands. Nucleic Acids Res 42(D1):D1098–D1106
59. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S et al (2004) UniProt, the universal protein knowledgebase. Nucleic Acids Res 32:D115–D119
60. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N et al (2014) A proteome-scale map of the human interactome network. Cell 159(5):1212–1226
61. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N et al (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437(7062):1173–1178
62. Mosca R, Ceol A, Aloy P (2013) Interactome3D, adding structural details to protein networks. Nat Methods 10(1):47–53
63. Meyer MJ, Das J, Wang X, Yu H (2013) INstruct, a database of high-quality 3D structurally resolved protein interactome networks. Bioinformatics 29(12):1577–1579
64. Meyer MJ, Beltran JF, Liang S, Fragoza R, Rumack A, Liang J et al (2018) Interactome INSIDER, a structural interactome browser for genomic studies. Nat Methods 15(2):107–114
65. Cheng F, Jia P, Wang Q, Zhao Z (2014) Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. Oncotarget 5(11):3697–3710
66. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B et al (2004) Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res 32(Database issue):D497–501
67. Newman RH, Hu J, Rho HS, Xie Z, Woodard C, Neiswinger J et al (2013) Construction of human activity-based phosphorylation networks. Mol Syst Biol 9:655
68. Hu J, Rho HS, Newman RH, Zhang J, Zhu H, Qian J (2014) PhosphoNetworks, a database for human phosphorylation networks. Bioinformatics 30(1):141–142

69. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E (2015) PhosphoSitePlus, 2014, mutations, PTMs and recalibrations. Nucleic Acids Res 43(Database issue):D512–520
70. Lu CT, Huang KY, Su MG, Lee TY, Bretana NA, Chang WC et al (2013) DbPTM 3.0, an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. Nucleic Acids Res 41(Database issue):D295–305
71. Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ et al (2011) Phospho.ELM, a database of phosphorylation sites–update 2011. Nucleic Acids Res 39(Database issue):D261–267
72. Fazekas D, Koltai M, Turei D, Modos D, Palfy M, Dul Z et al (2013) SignaLink 2—a signaling pathway resource with multi-layered regulatory networks. BMC Syst Biol 7:7
73. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K et al (2017) Architecture of the human interactome defines protein communities and disease networks. Nature 545(7655):505–509
74. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D et al (2015) The BioGRID interaction database, 2015 update. Nucleic Acids Res 43(Database issue):D470–478
75. Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM et al (2012) PINA v2.0, mining interactome modules. Nucleic Acids Res 40(Database issue):D862–865
76. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E et al (2012) MINT, the molecular interaction database, 2012 update. Nucleic Acids Res 40(Database issue):D857–861
77. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F et al (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 42(Database issue):D358–363
78. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R et al (2013) InnateDB, systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res 41(Database issue):D1228–1233
79. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A (2015) OMIM.org, Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res 43(Database issue):D789–798
80. Davis AP, Grondin CJ, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL et al (2015) The comparative toxicogenomics database's 10th year anniversary, update 2015. Nucleic Acids Res 43(Database issue):D914–920
81. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM et al (2014) ClinVar, public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42(Database issue):D980–985
82. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H et al (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42(Database issue):D1001–1006
83. Li MJ, Liu Z, Wang P, Wong MP, Nelson MR, Kocher JP et al (2016) GWASdb v2, an update database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res 44(D1):D869–D876
84. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD et al (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 31(12):1102–1110
85. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ (2008) A navigator for human genome epidemiology. Nat Genet 40(2):124–125
86. Pinero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M et al (2015) DisGeNET, a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford), bav028
87. Bodenreider O (2004) The unified medical language system (UMLS), integrating biomedical terminology. Nucleic Acids Res 32(Database issue):D267–270

88. Coordinators NR (2016) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44(D1):D7–D19
89. Moslehi JJ (2016) Cardiovascular toxic effects of targeted cancer therapies. N Engl J Med 375(15):1457–1467
90. Schmidinger M, Zielinski CC, Vogl UM, Bojic A, Bojic M, Schukro C et al (2008) Cardiac toxicity of sunitinib and sorafenib in patients with metastatic renal cell carcinoma. J Clin Oncol 26(32):5204–5212
91. Ghofrani HA, Seeger W, Grimminger F (2005) Imatinib for the treatment of pulmonary arterial hypertension. N Engl J Med 353(13):1412–1413
92. Ghofrani HA, Morrell NW, Hoeper MM, Olschewski H, Peacock AJ, Barst RJ et al (2010) Imatinib in pulmonary arterial hypertension patients with inadequate response to established therapy. Am J Respir Crit Care Med 182(9):1171–1177
93. Farha S, Dweik R, Rahaghi F, Benza R, Hassoun P, Frantz R et al (2014) Imatinib in pulmonary arterial hypertension, c-Kit inhibition. Pulm Circ 4(3):452–455
94. Schermuly RT, Dony E, Ghofrani HA, Pullamsetti S, Savai R, Roth M et al (2005) Reversal of experimental pulmonary hypertension by PDGF inhibition. J Clin Invest 115(10):2811–2821
95. Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M et al (2008) Estimating the size of the human interactome. Proc Natl Acad Sci USA 105(19):6959–6964
96. Gao J, Barzel B, Barabasi AL (2016) Universal resilience patterns in complex networks. Nature 530(7590):307–312
97. Vinayagam A, Gibson TE, Lee HJ, Yilmazel B, Roesel C, Hu Y et al (2016) Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. Proc Natl Acad Sci USA 113(18):4976–4981
98. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M et al (2014) The ChEMBL bioactivity database, an update. Nucleic Acids Res 42(Database issue):D1083–1090
99. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J (2016) BindingDB in 2015, a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res 44(D1):D1045–D1053
100. Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO et al (2010) An overview of the PubChem BioAssay resource. Nucleic Acids Res 38(Database issue):D255–266
101. Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM et al (2016) DGIdb 2.0, mining clinically relevant drug-gene interactions. Nucleic Acids Res 44(D1):D1036–1044
102. Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, Jensen LJ et al (2014) STITCH 4, integration of protein-chemical interactions with user data. Nucleic Acids Res 42(Database issue):D401–407
103. Nickel J, Gohlke BO, Erehman J, Banerjee P, Rong WW, Goede A et al (2014) SuperPred, update on drug classification and target prediction. Nucleic Acids Res 42(Web Server issue):W26–31
104. Lamb J (2007) The connectivity map, a new tool for biomedical research. Nat Rev Cancer 7(1):54–60
105. Duan Q, Flynn C, Niepel M, Hafner M, Muhlich JL, Fernandez NF et al (2014) LINCS Canvas Browser, interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. Nucleic Acids Res 42(Web Server issue):W449–460
106. Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T et al (2015) Open TG-GATEs, a large-scale toxicogenomics database. Nucleic Acids Res 43(Database issue):D921–927
107. Ganter B, Snyder RD, Halbert DN, Lee MD (2006) Toxicogenomics in drug discovery and development, mechanistic analysis of compound/class-dependent effects using the DrugMatrix database. Pharmacogenomics 7(7):1025–1044
108. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S et al (2009) Human protein reference database–2009 update. Nucleic Acids Res 37(Database issue):D767–772

109. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M et al (2017) The STRING database in 2017, quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45(D1):D362–D368
110. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al (2003) Cytoscape, a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504
111. Bastian M, Heymann S, Jacomy M (2009) Gephi, an open source software for exploring and manipulating networks. In: International AAAI conference on web and social media, North America. Available at, https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

Feixiong Cheng, Ph.D. is a principal investigator with Cleveland Clinic's Genomic Medicine Institute. Dr. Cheng is a computational biologist by training, with expertise in analyzing, visualizing, and mining data from real world (e.g., electronic health records, and health care claims) and experiments that profile the molecular state of human cells and tissues by interactomics, transcriptomics, genomics, proteomics, and metabolomics for drug discovery and precise patient care. Dr. Cheng is working to develop computational and experimental network medicine technologies for advancing the characterization of disease heterogeneity, thereby approaching the goal of coordinated, patient-centered strategies to innovative diagnostics and therapeutics development. The primary goal of Dr. Cheng's lab is to combine tools from genomics, network medicine, bioinformatics, computational biology, chemical biology, and experimental pharmacology and systems biology assays (e.g., single-cell sequencing), to address the challenging questions toward understanding of various human complex diseases (e.g., cardio-oncology, pulmonary vascular diseases, and cancer), which could have a major impact in identifying novel real-world data-driven diagnostic biomarkers and therapeutic targets for precision medicine.

Chapter 6
Mode-of-Action-Guided, Molecular Modeling-Based Toxicity Prediction: A Novel Approach for *In Silico* Predictive Toxicology

Ping Gong, Sundar Thangapandian, Yan Li, Gabriel Idakwo, Joseph Luttrell IV, Minjun Chen, Huixiao Hong and Chaoyang Zhang

Abstract Computational toxicology is a sub-discipline of toxicology concerned with the development and use of computer-based models and methodology to understand and predict chemical toxicity in a biological system (e.g., cells and organisms). Quantitative structure–activity relationship (QSAR) has been the predominant approach in computational toxicology. However, classical QSAR methodology has often suffered from low prediction accuracy, largely owing to the lack or non-integration of toxicological mechanisms. To address this lingering problem, we have

P. Gong (✉) · S. Thangapandian
Environmental Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS 39180, USA
e-mail: Ping.Gong@usace.army.mil

S. Thangapandian
e-mail: Sundarapandian.Thangapandian@usace.army.mil

Y. Li
Bennett Aerospace, Inc., Cary, NC 27518, USA
e-mail: yli@bennettaerospace.com

G. Idakwo · J. Luttrell IV · C. Zhang
School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS 39406, USA
e-mail: Gabriel.Idakwo@usm.edu

J. Luttrell IV
e-mail: Joseph.Luttrell@usm.edu

C. Zhang
e-mail: Chaoyang.Zhang@usm.edu

M. Chen · H. Hong
Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA
e-mail: Minjun.Chen@fda.hhs.gov

H. Hong
e-mail: Huixiao.Hong@fda.hhs.gov

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_6

developed a novel *in silico* toxicology approach that is based on molecular modeling and guided by mode of action (MoA). Our approach is implemented through a target-specific toxicity knowledgebase (TsTKb), consisting of a pre-categorized database of chemical MoA (ChemMoA) and a series of pre-built, category-specific classification and quantification models. ChemMoA serves as the depository of chemicals with known MoAs or molecular initiating events (i.e., known target biomacromolecules) and quantitative information for measured toxicity endpoints (if available). The models allow a user to qualitatively classify an uncharacterized chemical by MoA and quantitatively predict its toxicity potency. This approach is currently under development and will evolve to incorporate physiologically based pharmacokinetic (PBPK) modeling to address absorption, distribution, metabolism and excretion (ADME) processes in a biological system. The fully developed approach is believed to significantly advance *in silico*-based predictive toxicology and provide a new powerful toolbox for regulators, the chemical industry and the relevant academic communities.

Keywords Mode of action (MoA) · Molecular dynamics (MD) simulation · Molecular docking · Deep learning · Predictive toxicology · Target-specific toxicity knowledgebase (TsTKb) · Chemical mode of action database (ChemMoA) · Qualitative classification · Quantitative prediction · Quantitative structure–activity relationship (QSAR)

Abbreviations

3D	Three-dimensional
3Rs	Refine, reduce, and replace
ACToR	Aggregated Computational Toxicology Online Resource
ADME	Absorption, distribution, metabolism, and excretion
AOP	Adverse outcome pathway
BLAST	Basic local alignment search tool
BPA	Bisphenol A
ChemMoA	Chemical MoA
DSSTox	Distributed Structure-Searchable Toxicity
dyPLID	Dynamic protein–ligand interaction descriptors
EADB	Estrogenic Activity Database
EDKB	Endocrine Disruptor Knowledge Base
EDSP	Endocrine Disruptor Screening Program
EPA	Environmental Protection Agency
EU	European Union
FDA	Food and Drug Administration
iPSC	Induced Pluripotent Stem Cell
LTKB	Liver Toxicity Knowledge Base
MD	Molecular Dynamics
MIE	Molecular Initiating Event

MMDB	Molecular Modeling DataBase
MoA	Mode of Action
NCBI	National Center for Biotechnology Information
NRC	National Research Council
OECD	Organization for Economic Cooperation and Development
PBPK	Physiologically based pharmacokinetic
PDB	Protein Data Bank
QSAR	Quantitative structure–activity relationship
RC	Reference chemical
REACH	Registration, Evaluation, Authorization, and Restriction of Chemicals
Risk21	Risk Assessment in the twenty-first century
SEURAT	Safety Evaluation Ultimately Replacing Animal Testing
SPLIF	Structural protein–ligand interaction fingerprints
T3DB	Toxin and Toxin Target Database
Tox21	Toxicology in the twenty-first century
ToxCast	Toxicity Forecaster
TsTKb	Target-specific toxicity knowledgebase
USDA	US Department of Agriculture

6.1 Introduction

All chemical substances are required to be tested for their toxicological and environmental properties before being approved for use by regulatory authorities, such as the US Environmental Protection Agency (EPA), the US Food and Drug Administration (FDA) and the US Department of Agriculture (USDA). Three strategies are commonly adopted for toxicity testing: in vivo, in vitro, and in silico [1–4]. Driven by public concerns about animal welfare in research and testing, the European Union (EU) and most individual countries now have laws, policies, and regulations that aim to refine, reduce, and replace animal use (commonly referred to as the "3Rs") [5]. For instance, the US Animal Welfare Act requires consideration of alternatives whenever procedures involve more than slight or momentary pain or distress for warm-blooded animals in research and testing [6, 7]. The US FDA endorses the effort "to reduce animal testing [and] to work toward replacement of animal testing" as a basis for regulatory action [8]. In Europe, Directive 2010/63/EU [9] legislates an end to "the use of animals in toxicology and biomedical research as soon as scientifically feasible to do so." The EC 1907/2006 Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) regulation [10] restricts animal testing only "as a last resort to satisfy registration information requirements", and Regulation 1223/2009/EU [11] introduces a comprehensive ban on marketing within the EU of any cosmetic product (or ingredient thereof) that has been tested on animals since March 2013 [8, 12].

Another incentive for modernizing chemical toxicity testing is the combination of the large number of chemicals found in the environment with no or little toxicity data and the prohibitively high costs associated with traditional toxicity testing methods.

In 2007, the US National Research Council (NRC) published a landmark report entitled "Toxicity Testing in the twenty-first Century: A Vision and a Strategy", which envisioned a paradigm shift from in vivo animal-based studies to target-specific, mechanism-based toxicity pathway perturbations using in vitro and computational modeling approaches [13]. Implementation of this new vision is believed to be able to transform toxicology from a largely observational science to a more predictive one [14]. Since then, the toxicology community has made significant progress developing in vitro assays and computational tools that help achieve the predictive toxicology goal outlined in the seminal NRC report [8]. Regulatory bodies are also increasingly turning to alternative toxicity testing methods, among which at least 63 have been approved or endorsed by US federal regulatory agencies and international test guideline organizations, such as the Organization for Economic Cooperation and Development (OECD) [15–17].

6.1.1 Highlights of Recent Progress in the Development of Alternative Testing Methods

Over the past decade, a number of efforts have been initiated toward developing innovative in vitro and *in silico* tools and methodology for toxicity testing. The most prominent ones include use of induced pluripotent stem cell (iPSC)-derived human cells, development of defined heterotypic cell and three-dimensional (3D) cell/tissue models, engineered human "organ-on-a-chip" microscale physiological systems, mathematical modeling of cellular processes and morphogenesis, adverse outcome pathways (AOPs), a molecular initiating event (MIE) atlas for toxicities, and next-generation quantitative structure–activity relationship (QSAR) models [8, 12, 18]. These efforts have been mostly carried out within large-scale research programs, such as toxicology in the twenty-first Century (Tox21) [19], Toxicity Forecaster (ToxCast) [20], Safety Evaluation Ultimately Replacing Animal Testing (SEURAT)-1 [21], Risk Assessment in the twenty-first Century (Risk21) [22], human-on-a-chip [23], the Human Toxome Project [24], carcinoGENOMICS [22], the US Endocrine Disruptor Screening Program (EDSP) [8], and virtual tissues [25] (e.g., virtual brain [26], embryo, liver and thyroid models; see also www.epa.gov/chemical-research/virtual-tissue-models-predicting-how-chemicals-impact-development). Looking at Tox21 as an example, high-throughput in vitro screening assays have been developed using a highly automated robotics platform to quickly and efficiently assess whether certain chemical compounds have the potential to disrupt processes in the human body and possibly lead to negative health effects [19]. All these efforts also share a common strategic goal—turn the knowledge of toxicological modes of action (MoAs) and perturbed toxicity pathways or AOPs into in vitro and *in silico* models that quantitatively predict points of departure or other end points (e.g., ED_{50} and AC_{50}) for chemical toxicity/risk/safety assessment [8, 19–24].

6.1.2 *Mechanism-Based Toxicity Prediction: MIE and "Critical Target" Concept*

Allen et al. recently defined an AOP as a conceptual framework, presented as a logical sequence of events or processes within biological systems, which can be used to understand adverse effects and refine current risk assessment practices in ecotoxicology, and an MIE as the initial interaction between a chemical and a biomolecule that can be causally linked to an outcome via a pathway [27]. According to the AOP concept, any phenotypic endpoint of toxicity, i.e., an adverse outcome resulting from a series of biological processes can be eventually linked to a unique MIE [18, 28]. It is now widely accepted that QSAR-based predictive toxicology has more success when the mechanism of toxicity end point is well understood or can be linked to a well-defined molecular target [29]. In a recent perspective review on the future research of predictive toxicology, Daston et al. [30] recommended to focus on the identification of "critical biological targets" relevant for toxicity and to test their suitability for being used as anchors for predicting toxicity. Chemicals can interfere with normal biological processes or pathways at the molecular level through a multitude of different mechanisms that vary from non-selective binding (to intracellular proteins) to selective agonism/antagonism of a particular nuclear or another receptor [30]. Decades of conventional animal toxicity testing have accumulated a myriad of information on toxicological mechanisms that may be further explored for use in in vitro and computational modeling-based predictive toxicology. For instance, Tox21 researchers have developed high throughput, cell-based or cell-free in vitro assays that evaluate critical cellular targets or processes involved in toxicity response [19]. They have also published a prioritized set of 2750 targeted sentinel genes (referred to as Human S1500+ Gene Set Ver2) whose transcriptional changes are responsive to exposures to a wide variety of toxic agents [31].

6.1.3 *Limitations of Current In Vitro and* In Silico *Approaches*

While significant progress has been made in developing mechanism-based in vitro assays and testing platforms (including high-content toxicogenomics platforms), it is recognized that quantitative in vitro to in vivo extrapolation still faces many technical barriers, such as the choice of appropriate cell lines and the lack of metabolism [8, 19]. Although *in silico* approaches are rapid and inexpensive compared to experimental approaches, the growth of *in silico*-based predictive toxicology tools is unsatisfactory. Conventional QSAR models often suffer from low prediction accuracy [32] because they do not consider the structure and flexibility of target biomacromolecules, especially when applied toward the more elusive goal of predicting potential toxicity outcomes for in vitro cell cultures or in vivo animal test systems [29]. In these systems, the toxicity end point (e.g., cytotoxicity, mutagenicity, developmental toxicity

and cancer) tends to be less well understood and often encompasses multiple mechanisms and pathways to adverse outcome. Consequently, the success of system-level modeling or simulation of dynamic biological processes leading to toxicity is mostly limited to "fit-for-purpose" [8, 33] due to incomplete knowledge of complex biological systems. Hence, for the purpose of accurately predicting chemical toxicity, it is currently unrealistic and unnecessary to capture and reconstruct all the cascading processes leading from an MIE to an adverse outcome.

6.1.4 Molecular Docking for Virtual Chemical Screening: A Green Toxicology Approach

The pharmaceutical and pesticide/herbicide industries have a long history of using molecular docking as a key tool in computer-assisted virtual chemical design and efficacy screening of candidate compounds. A wide variety of ligand–protein docking methods have been developed to predict the predominant conformation and orientation [i.e., pose(s) or binding mode(s)] of a ligand within a targeted binding site of a biomacromolecule (e.g., a protein) with a known 3D structure [34–36]. These methods can model the interaction between a small molecule (chemical) and a biomacromolecule at the atomic level. This allows us to characterize the behavior of small molecules in the binding site of target biomacromolecules and elucidate fundamental biochemical processes [37].

In essence, this virtual screening approach falls within the scope of an emerging discipline called green toxicology. Similar to the green chemistry movement, it moves the toxicity and risk or safety assessment schemes to the beginning of the production cycle of a chemical or a product, i.e., to the molecular design [38]. Green toxicology uses predictive toxicology tools for the design of less harmful substances, tests early in the development process to prioritize less dangerous chemicals, and reduces exposures—thereby "designing out" undesirable human health and environmental risks, reducing animal testing demands, and increasing the likelihood of launching a successful, sustainable product [21].

However, molecular docking-based virtual screening has not been applied to quantitative assessment of long-term chemical toxicity. Its application to qualitative mechanistic studies is also limited, largely due to the historical unavailability of 3D macromolecular structures of many toxicity targets and the extremely high computational expenses associated with allowing conformational flexibility of both the ligand and the protein. Recently, advances in structural biology (e.g., high-throughput protein purification, crystallography and nuclear magnetic resonance spectroscopy techniques [35]), rapid progress in algorithm development, and great advances in supercomputing resources (e.g., high-performance computing technology) [39] have paved the way for pursuing this approach and its potential to scale-up for quickly screening thousands of critical toxicity targets. For instance, the molecular modeling database (MMDB) [40], which is based on the Protein Data Bank (PDB) [41, 42]

and maintained by the National Center for Biotechnology Information (NCBI), now contains 144,042 records of resolved structures (as of September 7, 2018) for chemically bound or unbound proteins, DNAs and RNAs (as well as their complexes). Consequently, recent years have seen increased applications of molecular docking in qualitative toxicological MoA studies; for example: (1) screening of endocrine disrupting environmental compounds through docking to the ligand-binding domain of estrogen receptor α [43], (2) predicting idiosyncratic drug reactions via examining the binding modes of drugs in the human leukocyte antigens [44], and (3) evaluating the endocrine disrupting activity of 45 bisphenol A (BPA) replacement compounds using molecular dynamics (MD) simulations [45].

6.2 Methodology

6.2.1 Approach Overview

In view of the existing limitations of current approaches briefly reviewed above, we have developed a novel, state-of-the-art predictive toxicology approach that is guided by MoA, MIE or other toxicological mechanism information at the molecular level, as illustrated in Fig. 6.1. This approach is also based on molecular modeling that integrates structural biology principles, computational chemistry tools, and machine learning techniques. Historically, molecular modeling (molecular docking and MD simulation in particular) has been applied to qualitative studies for elucidating the mechanism of molecular interactions between a ligand and a target biomacromolecule [45, 46] or to high-throughput preliminary screening of drug candidates for their potency on disease targets [34, 35, 47]. Here we expand its application to quantitative toxicity assessment. Meanwhile, machine learning (especially deep learning) methods that have been widely applied in predictive toxicology are employed to train and validate toxicity prediction models for qualitative categorization and quantitative estimation of uncharacterized chemicals [48, 49].

6.2.2 Approach Implementation

Our approach is implemented through a target-specific toxicity knowledgebase (TsTKb) that consists of a pre-categorized database of chemical MoAs (ChemMoA) and a library of pre-built, category-specific classification and quantification models (see [50] for more information). ChemMoA serves as the depository of chemicals with known MoAs or MIEs (i.e., known target biomacromolecules) and quantitative information for measured toxicity endpoints (Fig. 6.1). The following information is curated in ChemMoA: chemical data (e.g., IUPAC name, identifier, SMILES structure, and 1D to 3D molecular descriptors), target data (i.e., the 3D structures of

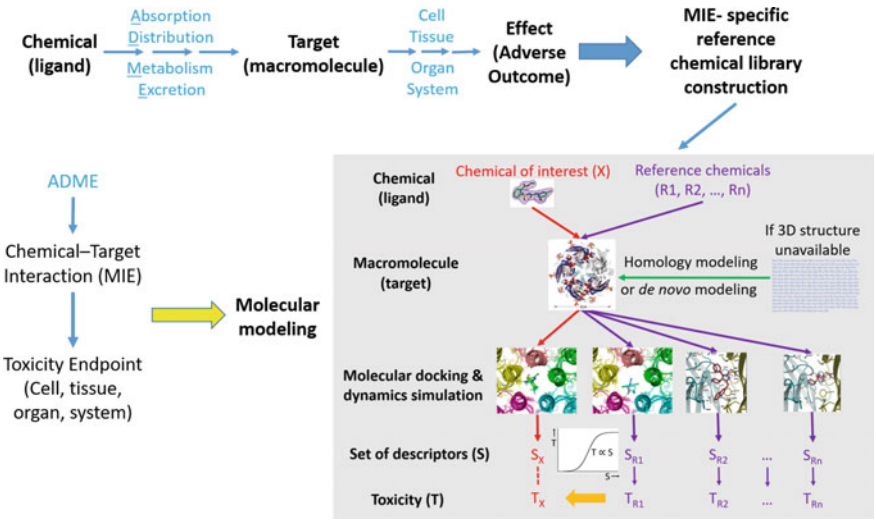

Fig. 6.1 Schematic workflow of the mode of action (MoA)/molecular initiating event (MIE)-guided, molecular modeling-based *in silico* predictive toxicology approach

biomacromolecular targets retrieved from the PDB [40–42] or built via homology or de novo modeling), toxicity data, experimental data, and citation data. The model library is built using data curated in ChemMoA and allows a user to qualitatively classify an uncharacterized chemical by MoA and quantitatively predict its toxicity potency. For instance, the TsTKb enables one to *in silico* screen an uncharacterized chemical (X) for each potential macromolecular target in ChemMoA, estimate the interaction activity in terms of a set of molecular descriptor (S) including dynamic protein–ligand interaction descriptors (dyPLIDs) [51] and scoring function-based binding scores [34, 52], and identify potential toxicity target(s) (Fig. 6.1). Then, the toxicity (T) of the chemical of interest (X) can be derived quantitatively as a function of the binding scores and other variables (e.g., molecular descriptors) that are relative to the measured toxicity of well-characterized reference chemicals (RCs) that elicit toxicity through interfering with the same macromolecular target. For quantitative prediction, a significant correlation must exist between T and S, i.e., $T \propto S$. When scaling up, this workflow can be replicated simultaneously for other uncharacterized chemicals and many different toxicity targets.

At the current stage of development, we have not yet considered the effect of modulators in the molecular modeling although they may influence the binding affinity between a chemical and its target biomacromolecule. Furthermore, the influence of absorption, distribution, metabolism, and excretion (ADME) processes on chemical toxicity in vivo [53] remains to be accounted for (Fig. 6.1). These impacts can be built into a quantitative prediction model: $T = f(S_{\text{target}}, S_{\text{modulator}}, S_{\text{adme}})$, where S_{target} stands for the binding activity of a chemical to a target receptor, $S_{\text{modulator}}$ represents

the modification of S_{target} by modulators, and S_{adme} accounts for the influence of ADME on chemical bioavailability biomacromolecules.

6.3 Results

As an ongoing project, we divide the development of our novel MoA/MIE-guided, molecular modeling-based approach for *in silico* predictive toxicology into two major phases. In Phase 1, we intend to focus on prediction of in vitro toxicity endpoints. In Phase 2, we will expand to in vivo toxicity end points by incorporating ADME processes into prediction modeling (Fig. 6.1). In the following, we provide an update on the Phase 1 status of the TsTKb development. For more details about the TsTKb, one may refer to our recent publication [50].

6.3.1 Libraries of Reference Chemicals in ChemMoA

We have curated a library of reference chemicals for each MoA/MIE or toxicity target according to the following criteria: (1) the availability of toxicity data, (2) the uniqueness of toxicity target (to avoid chemicals interacting with multiple targets within a toxicity pathway and eliciting the same toxicity at the organ/system level), and (3) their reported toxicity spanning a wide potency spectrum. We have queried more than a dozen publicly accessible databases: Aggregated Computational Toxicology Online Resource (ACToR) [54] and Distributed Structure-Searchable Toxicity (DSSTox) [55] databases, both developed by the US EPA; PubChem [56]; ChEMBL [57]; ZINC15 [58, 59]; Estrogenic Activity Database (EADB), Endocrine Disruptor Knowledge Base (EDKB), and Liver Toxicity Knowledge Base (LTKB), all of which were developed by researchers at the US FDA [60]; SuperTarget [61]; SuperToxic [62]; Toxin and Toxin Target Database (T3DB) [63, 64]; TG-GATE [65]; and TOXNET [66, 67]. We retain chemicals that cause a wide variety of toxic effects (e.g., acetolactate synthase inhibition, $GABA_A$ receptor antagonism, hepatic steatosis, acetylcholinesterase inhibition, androgen receptor antagonism/agonism, and estrogen receptor antagonism/agonism) through interacting with their respective toxicity targets. MoA/MIE data for these chemicals are retrieved and categorized, whereas toxicity data are normalized and harmonized. The AOP knowledgebase, developed as part of the OECD AOP Development Effort [18, 68] and hosted at the AOP Wiki Web portal (https://aopwiki.org/) is also consulted with regard to MoA-based chemical categorization. An example of the curated hepatotoxin library is provided in Fig. 6.2.

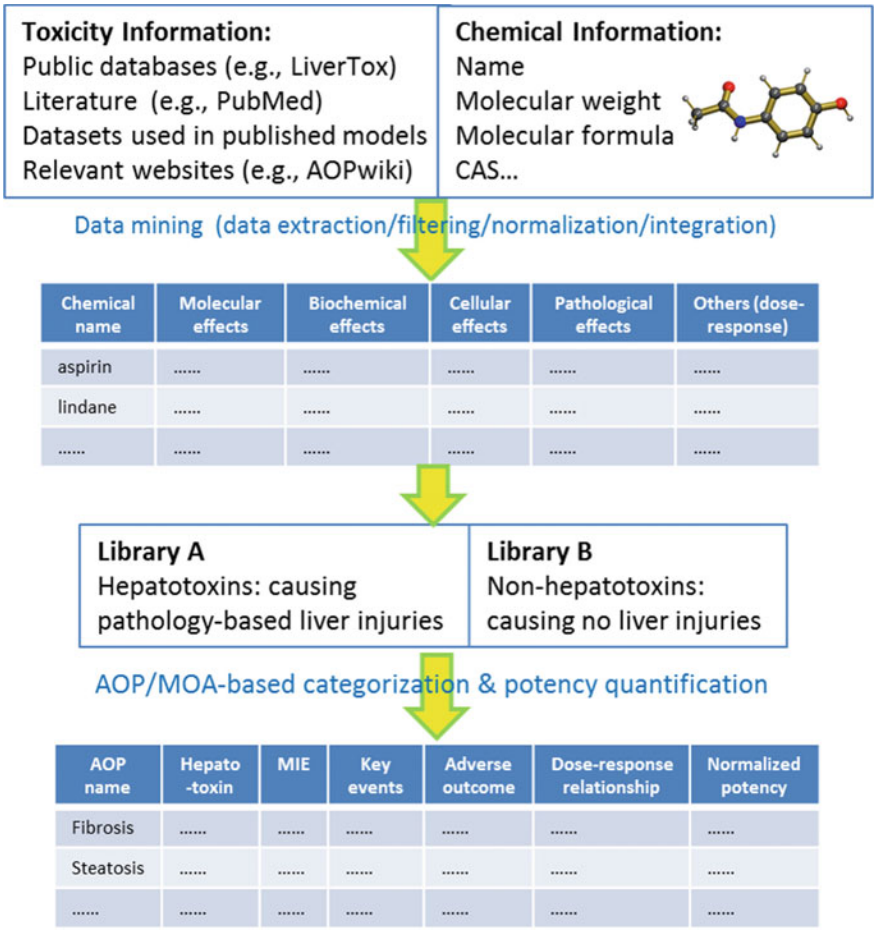

Fig. 6.2 An example of reference chemical library preparation based on toxicological mode of action (MoA) or molecular initiating event (MIE) in the framework of adverse outcome pathway (AOP)

6.3.2 Structural Models of Biomacromolecular Targets

The 3D crystal structures of toxicity targets are first searched in RCSB's PDB [69] (https://www.rcsb.org/) and NCBI's MMDB [40] (https://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml). If the 3D structure of a biomarcomolecular target is not yet resolved, we first retrieve its protein sequence from relevant databases, such as UniProtKB [70] (https://www.uniprot.org). Then, using BLAST (Basic Local Alignment Search Tool) [71] to search for homologous proteins with available 3D structures, we choose the one with the highest similarity to the homologous template. Homology-based modeling tools, such as Modeller [72] and Swiss-Model

[73], are employed to build homology models for the target biomacromolecule. The stereochemical quality of the homology models is assessed using PROCHECK [74], whereas the local and global model quality is estimated using the QMEAN scoring function [75]. If a homologous protein of known 3D structure cannot be found, the de novo modeling approach QUARK [76] is employed to produce ab initio a template-free predicted model for the target of interest.

6.3.3 Molecular Docking and MD Simulation for dyPLID Generation

With the 3D structure of a target biomacromolecule in hand, FINDSITE [77], a threading algorithm, or POCKET [78], a cavity detection program, is used to identify putative active binding sites in the target biomacromolecule if the sites are unknown. Then, AutoDock Vina [79] is utilized to dock a chemical into the binding sites in the target biomacromolecule. The top-scoring binding pose with a favorable calculated binding energy is selected and further refined using Amber18 (http://ambermd.org/), an MD simulation program package [80, 81]. The obtained MD trajectories and the VMD program [82] are used to calculate the dyPLIDs as the quantitative measurements for possible target-chemical interactions. The binding energy is recalculated after MD simulations, and the refined binding energy estimate is included in the set of target-chemical interaction descriptors.

6.3.4 In Vitro *Toxicity Prediction Mode Libraries*

We are currently developing toxicity classification and quantification models using machine learning (including classical SVM, random forest, and deep learning algorithms such as deep neural networks [48, 83]) methods. For instance, using the androgen receptor bioassay dataset for more than ten thousand chemicals made available through the Tox21 data challenge [84] (https://tripod.nih.gov/tox21/challenge/), we have built a set of models to classify these chemicals into agonists, antagonists, inactive ligands, and inconclusive compounds (i.e., neither active nor inactive) (G. Idakwo et al. Manuscript under review) and to quantitatively predict the degree of agonism or antagonism for 273 active compounds [51].

6.3.5 Web Portal for ChemMoA/TsTKb

It has been our intention to make the toolkits we developed publicly accessible so that our research findings and products can be disseminated to the relevant communities

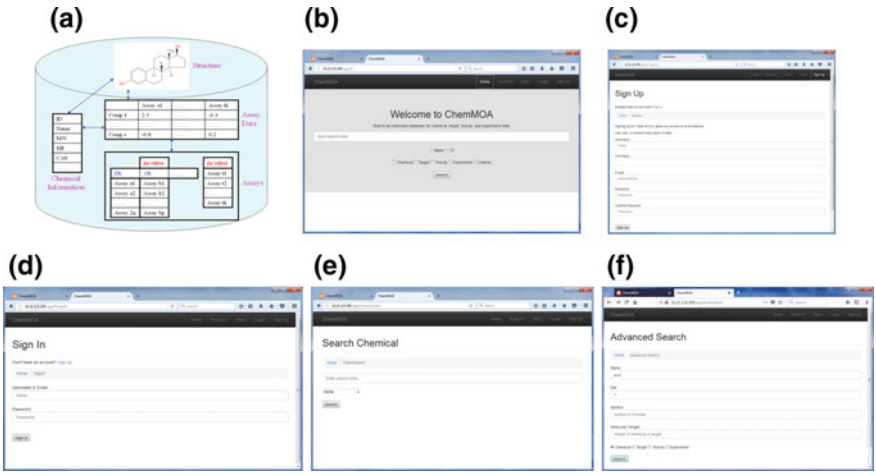

Fig. 6.3 Architecture of the relational ChemMoA database (**a**) and captured screenshots of the prototype ChemMoA web portal: **b** Homepage, **c** Sign-up page, **d** Sign-in page, **e** Simple search, and **f** Advanced search

in a timely fashion. Hence, we have started to develop a Web portal for ChemMoA. At this time, a prototype of the relational ChemMoA database has been completed (see [50] for more details). The architecture and a few screenshots of the ChemMoA database are provided in Fig. 6.3. We plan to add toxicity target structure models and toxicity prediction model libraries to ChemMoA and eventually turn it into the TsTKb with more features and functionalities (e.g., target-specific in vitro toxicity classification and prediction of uncharacterized chemicals).

6.4 Discussion

The field of *in silico* predictive toxicology has been dominated by conventional QSAR-based approaches. For many decades, QSAR modeling techniques have undergone continuous development and refinement dedicated primarily to enhance prediction accuracy based on the relationships between physicochemical properties of chemical substances and their biological activities. For instance, previous efforts have resulted in the advent of 0D to 3D molecular descriptors for chemical ligands. A molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number, or the result of some standardized experiment [85]. 0D descriptors are atom counts and sums, and 1D descriptors are constitutional parameters, such as molecular weight and the list of substructural fragments and bonds. 2D descriptors are based on molecular topology and include graph invariants (topological indices) and topographic descriptors, and 3D descriptors expand to

geometrical parameters covering molecular surfaces and fields as well as parameters calculated in quantum chemistry.

QSAR approaches have also evolved from 1D to 6D [86, 87] and have the following properties: 1D-QSAR correlates biological activity with physiochemical properties, such as pK_a and logP [88]; 2D-QSAR correlates activity with structural patterns, such as connectivity indices, and 2D-pharmacophores. [89]; 3D-QSAR correlates activity with non-covalent interaction fields surrounding the ligand (e.g., CoMFA [90]) or receptor (COMBINE [91]) in an alignment-dependent (CoMFA [90] and COMBINE [91]) or independent (WHIM [92] and COMPASS [93]) fashion; 4D-QSAR adds ensemble sampling of conformation, orientation and protonation state as the fourth dimension [94, 95]; 5D-QSAR allows for a multiple representation of the topology of the quasi-atomistic receptor surrogate (i.e. the fifth dimension), leading to less biased induced-fit models [96]; and 6D-QSAR further allows for the simultaneous consideration of different solvation models (i.e., the sixth dimension), reflecting varying solvent accessibility [97]. However, all these QSAR strategies either do not consider ligand-receptor interaction (1D- and 2D-QSAR), or restrain the flexibility of the ligand and/or receptor by modeling the binding interaction in a predefined grid box (3D- to 6D-QSAR). It was not until recently that molecular docking and MD simulation were performed to infer optimal conformations with minimal binding free energy in order to compute structural protein–ligand interaction fingerprints (SPLIF) [98] and 3D-D Moments/WHIM descriptors [99], respectively.

To the best of our knowledge, QSAR modeling approaches have given little or no consideration to the dynamic nature in chemical-target biomacromolecule interactions, leading to limited success of QSAR models in toxicity prediction. Our novel MoA-guided and molecular modeling-based approach is geared to overcome the aforesaid drawbacks. Despite being an ongoing effort, this approach is believed to improve the accuracy and efficiency of predictive toxicology, which is supported by our preliminary results. There are three aspects of novelty in our approach: (1) pre-categorized reference chemical libraries organized by their documented MoA/MIE [50], (2) the generation of >5000 dyPLIDs that give full consideration to the flexibility of both ligands (small chemical molecules) and receptors (toxicity biomacromolecular targets) [51], and (3) the application of machine learning, especially deep learning, in the development of prediction models [48, 49].

6.5 Conclusion and Future Directions

The motivation for developing the novel approach presented in this chapter was to improve *in silico* toxicity characterization and risk assessment of existing chemicals as well as prediction of adverse biological effects for emerging or novel chemicals undergoing development. Ultimately, our work may lead to the following outcomes: (1) a reduction in animal use for toxicity testing, (2) early detection of toxicological properties, and (3) an increase in the likelihood of launching a sustainable "green" product without incurring undesirable human health and environmental risks. Specif-

ically, our approach addresses several challenges in model development for toxicity prediction and quantification: (1) what properties and features of the ligand-target biomacromolecule (e.g., receptor and enzyme protein) interactions should be taken into consideration; (2) how to capture such dynamic interactions and incorporate them into QSAR modeling; (3) how to characterize chemical toxicities beyond binary classes (toxic/non-toxic) for the purpose of lead optimization, mechanism elucidation and analogue prioritization; (4) how to handle the commonly encountered class imbalance problem in chemical classification; and (5) how to develop novel machine learning (e.g., deep learning) approaches that include solid theoretical foundation and advanced optimization techniques for rapid and accurate quantitative toxicity prediction. We anticipate finding interdisciplinary solutions for these challenges in the course of developing the TsTKb. We believe the fully developed TsTKb will significantly advance *in silico*-based predictive toxicology and provide a new powerful toolbox for regulators, the chemical industry and relevant academic communities.

References

1. Sachana M, Hargreaves AJ (2018) Toxicological testing. *In vivo* and *in vitro* models. In: Gupta RC (ed) Veterinary toxicology: basic and clinical principles, 3rd edn. Elsevier, London, UK, pp 145–161
2. Eisenbrand G, Pool-Zobel B, Baker V et al (2002) Methods of *in vitro* toxicology. Food Chem Toxicol 40:193–236. https://doi.org/10.1016/S0278-6915(01)00118-1
3. Jain AK, Singh D, Dubey K et al (2018) Models and methods for *in vitro* toxicity. In: Dhawan, Alok; Kwon S (ed) *In vitro* toxicology, 1st ed. Elsevier, London, UK, pp 45–65
4. Parthasarathi R, Dhawan A (2018) *In silico* approaches for predictive toxicology. In: Dhawan, Alok; Kwon S (ed) *In vitro* toxicology, 1st ed. Elsevier, London, UK, pp 91–109
5. Stokes WS (2015) Animals and the 3Rs in toxicology research and testing: The way forward. Hum Exp Toxicol 34:1297–1303. https://doi.org/10.1177/0960327115598410
6. United States Code (2014) Animal Welfare Act: Title 7, Chapter 54, Sections 2131–2159
7. Code of Federal Regulations (2014) Title 9: animals and animal products; Chapter I, subchapter A—animal welfare (Parts 1–4)
8. Knudsen TB, Keller DA, Sander M et al (2015) FutureTox II: *In vitro* data and *in silico* models for predictive toxicology. Toxicol Sci 143:256–267. https://doi.org/10.1093/toxsci/kfu234
9. European Union (2010) Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes
10. European Union (2006) Regulation (EC) No 1907/2006—Registration, evaluation, authorisation and restriction of chemicals (REACH)
11. European Union (2009) Regulation (EC) no 1223/2009 of the european parliament and of the council of 30 November 2009 on cosmetic products
12. Elmore SA, Ryan AM, Wood CE et al (2014) FutureTox II: contemporary concepts in toxicology: pathways to prediction. *In vitro* and *in silico* models for predictive toxicology. Toxicol Pathol 42:940–942. https://doi.org/10.1177/0192623314537135
13. National Research Council (2007) Toxicity testing in the 21st century: a vision and a strategy. National Academies Press, Washington, D.C.
14. Collins FS, Gray GM, Bucher JR (2008) Toxicology. Transforming environmental health protection. Science 319:906–907. https://doi.org/10.1126/science.1154619
15. Stokes WS (2014) Validation and regulatory acceptance of toxicological testing methods and strategies. In: Hayes AW, Kruger CL (eds) Hayes' principles and methods of toxicology, 6th edn. CRC Press, Boca Raton, pp 1103–1128

16. ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2012) Biennial progress report 2010–2011, NIH Publication No. 12-7873. Research Triangle Park, NC
17. ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2014) Biennial Progress Report 2012-2013. Research Triangle Park, NC
18. OECD (Organization for Economic Co-operation and Development) (2013) Guidance document on developing and assessing adverse outcome pathways. OECD environment, health and safety publications—series on testing and assessment, No. 184. Paris, France
19. Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121:756–765. https://doi.org/10.1289/ehp.1205784
20. Kavlock R, Chandler K, Houck K et al (2012) Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 25:1287–1302. https://doi.org/10.1021/tx3000939
21. COACH Consortium (2015) Mechanism-based methods for improved toxicity testing, SEURAT-1 Annual Report, Vol 5
22. Seidle T, Spielmann H (2012) Alternative testing strategies progress report 2012 and AXLR8-3 workshop report on a 'Roadmap to next generation safety testing under horizon 2020'. Berlin, Germany
23. Zhang C, Zhao Z, Abdul Rahim NA et al (2009) Towards a human-on-chip: culturing multiple cell types on a chip with compartmentalized microenvironments. Lab Chip 9:3185–3192. https://doi.org/10.1039/b915147h
24. Bouhifd M, Andersen ME, Baghdikian C et al (2015) The human toxome project. Altex 32:112–124. https://doi.org/10.14573/altex.1502091
25. Shah I, Wambaugh J (2010) Virtual tissues in toxicology. J Toxicol Environ Heal Part B 13:314–328. https://doi.org/10.1080/10937404.2010.483948
26. Sanz Leon P, Knock SA, Woodman MM et al (2013) The virtual brain: a simulator of primate brain network dynamics. Front Neuroinform 7:10. https://doi.org/10.3389/fninf.2013.00010
27. Allen TEH, Goodman JM, Gutsell S, Russell PJ (2014) Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem Res Toxicol 27:2100–2112. https://doi.org/10.1021/tx500345j
28. Ankley GT, Bennett RS, Erickson RJ et al (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730–741. https://doi.org/10.1002/etc.34
29. Cherkasov A, Muratov EN, Fourches D et al (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57:4977–5010. https://doi.org/10.1021/jm4004285
30. Daston G, Knight DJ, Schwarz M et al (2015) SEURAT: safety evaluation ultimately replacing animal testing–recommendations for future research in the field of predictive toxicology. Arch Toxicol 89:15–23. https://doi.org/10.1007/s00204-014-1421-5
31. Mav D, Shah RR, Howard BE et al (2018) A hybrid gene selection approach to create the S1500 + targeted gene sets for use in high-throughput transcriptomics. PLoS ONE 13:e0191105. https://doi.org/10.1371/journal.pone.0191105
32. Benigni R, Netzena TI, Benfenati E et al (2007) The expanding role of predictive toxicology: an update on the (Q)SAR models for mutagens and carcinogens. J Environ Sci Heal Part C 25:53–97. https://doi.org/10.1080/10590500701201828
33. Sturla SJ, Boobis AR, FitzGerald RE et al (2014) Systems toxicology: from basic research to risk assessment. Chem Res Toxicol 27:314–329. https://doi.org/10.1021/tx400410s
34. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3:935–949. https://doi.org/10.1038/nrd1549
35. Meng X-Y, Zhang H-X, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7:146–157
36. Morris GM, Lim-Wilby M (2008) Molecular docking. Methods Mol Biol 443:365–382. https://doi.org/10.1007/978-1-59745-177-2_19

37. Mcconkey BJ, Sobolev V, Edelman M (2002) The performance of current methods in ligand–protein docking. Curr Sci 83:
38. Maertens A (2014) Green toxicology. Altex 31:243–249. https://doi.org/10.14573/altex.1406181
39. Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 106:1589–1615. https://doi.org/10.1021/cr040426m
40. Madej T, Lanczycki CJ, Zhang D et al (2014) MMDB and VAST+ : tracking structural similarities between macromolecular complexes. Nucleic Acids Res 42:D297–D303. https://doi.org/10.1093/nar/gkt1208
41. Berman HM (2008) The protein data bank: a historical perspective. Acta Crystallogr A 64:88–95. https://doi.org/10.1107/S0108767307035623
42. Dutta S, Burkhardt K, Young J et al (2009) Data deposition and annotation at the worldwide protein data bank. Mol Biotechnol 42:1–13. https://doi.org/10.1007/s12033-008-9127-7
43. McRobb FM, Kufareva I, Abagyan R (2014) *In silico* identification and pharmacological evaluation of novel endocrine disrupting chemicals that act via the ligand-binding domain of the estrogen receptor α. Toxicol Sci 141:188–197. https://doi.org/10.1093/toxsci/kfu114
44. Luo H, Du T, Zhou P et al (2015) Molecular docking to identify associations between drugs and class I human leukocyte antigens for predicting idiosyncratic drug reactions. Comb Chem High Throughput Screen 18:296–304
45. Ng HW, Shu M, Luo H et al (2015) Estrogenic activity data extraction and in silico prediction show the endocrine disruption potential of bisphenol a replacement compounds. Chem Res Toxicol 28:1784–1795. https://doi.org/10.1021/acs.chemrestox.5b00243
46. Sakkiah S, Kusko R, Pan B et al (2018) Structural changes due to antagonist binding in ligand binding pocket of androgen receptor elucidated through molecular dynamics simulations. Front Pharmacol 9:492. https://doi.org/10.3389/fphar.2018.00492
47. Thangapandian S, John S, Sakkiah S, Lee KW (2010) Docking-enabled pharmacophore model for histone deacetylase 8 inhibitors and its application in anti-cancer drug discovery. J Mol Graph Model 29:382–395. https://doi.org/10.1016/j.jmgm.2010.07.007
48. Idakwo G, Luttrell J, Chen M et al (2018) A review on machine learning methods for *in silico* toxicity prediction. J Environ Sci Heal Part C - Environ Carcinog Ecotoxicol Rev 36:169–191
49. Tang W, Chen J, Wang Z, Xie H, Hong H (2018) Deep learning for predicting toxicity of chemicals: a mini review. J Environ Sci Heal Part C - Environ Carcinog Ecotoxicol Rev 36:252–271
50. Li Y, Idakwo G, Thangapandian S, Chen M, Hong H, Zhang C, Gong P (2018) Target-specific toxicity knowledgebase (TsTKb): a novel toolkit for *in silico* predictive toxicology. J Environ Sci Heal Part C—Environ Carcinog Ecotoxicol Rev 36:219–236
51. Thangapandian S, Idakwo G, Luttrell J, Hong H, Zhang C, Gong P (2019) Quantitative target-specific toxicity prediction modeling (QTTPM): a proof-of-concept case study on androgen receptor. Chem Sci (Submitted)
52. Trott O, Olson AJ (2010) AutoDock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461. https://doi.org/10.1002/jcc
53. Roncaglioni A, Toropov AA, Toropova AP, Benfenati E (2013) *In silico* methods to predict drug toxicity. Curr Opin Pharmacol 13:802–806. https://doi.org/10.1016/j.coph.2013.06.001
54. Judson RS, Martin MT, Egeghy P et al (2012) Aggregating data for computational toxicology applications: the U.S. environmental protection agency (EPA) aggregated computational toxicology resource (ACToR) System. Int J Mol Sci 13:1805–1831. https://doi.org/10.3390/ijms13021805
55. Williams AJ, Grulke CM, Edwards J et al (2017) The comptox chemistry dashboard: a community data resource for environmental chemistry. J Cheminform 9:61. https://doi.org/10.1186/s13321-017-0247-6
56. Brown N, Cambruzzi J, Cox PJ et al (2018) Big data in drug discovery. Prog Med Chem 57:277–356. https://doi.org/10.1016/bs.pmch.2017.12.003
57. Gaulton A, Hersey A, Nowotka M et al (2017) The ChEMBL database in 2017. Nucleic Acids Res 45:D945–D954. https://doi.org/10.1093/nar/gkw1074

58. Irwin JJ, Sterling T, Mysinger MM et al (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757–1768. https://doi.org/10.1021/ci3001277
59. Sterling T, Irwin JJ (2015) ZINC 15–ligand discovery for everyone. J Chem Inf Model 55:2324–2337. https://doi.org/10.1021/acs.jcim.5b00559
60. Hong H, Chen M, Ng HW, Tong W (2016) QSAR models at the US FDA/NCTR. Methods Mol Biol 1425:431–459. https://doi.org/10.1007/978-1-4939-3609-0_18
61. Hecker N, Ahmed J, von Eichborn J et al (2012) SuperTarget goes quantitative: update on drug-target interactions. Nucleic Acids Res 40:D1113–D1117. https://doi.org/10.1093/nar/gkr912
62. Schmidt U, Struck S, Gruening B et al (2009) SuperToxic: a comprehensive database of toxic compounds. Nucleic Acids Res 37:D295–D299. https://doi.org/10.1093/nar/gkn850
63. Lim E, Pon A, Djoumbou Y et al (2010) T3DB: a comprehensively annotated database of common toxins and their targets. Nucleic Acids Res 38:D781–D786. https://doi.org/10.1093/nar/gkp934
64. Wishart D, Arndt D, Pon A et al (2015) T3DB: the toxic exposome database. Nucleic Acids Res 43:D928–D934. https://doi.org/10.1093/nar/gku1004
65. Sandhu KS, Veeramachaneni V, Yao X et al (2015) Release of (and lessons learned from mining) a pioneering large toxicogenomics database. Pharmacogenomics 16:779–801. https://doi.org/10.2217/pgs.15.38
66. Fowler S, Schnall JG (2014) TOXNET: information on toxicology and environmental health. Am J Nurs 114:61–63. https://doi.org/10.1097/01.NAJ.0000443783.75162.79
67. Wexler P (2001) TOXNET: an evolving web resource for toxicology and environmental health information. Toxicology 157:3–10
68. OECD (2018) Users' handbook supplement to the guidance document for developing and assessing AOPs. France, Paris
69. Burley SK, Berman HM, Kleywegt GJ et al (2017) Protein data bank (PDB): the single global macromolecular structure archive. Methods Mol Biol 1607:627–641. https://doi.org/10.1007/978-1-4939-7000-1_26
70. Boutet E, Lieberherr D, Tognolli M et al (2007) UniProtKB/Swiss-Prot. Methods Mol Biol 406:89–112
71. Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
72. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 374:461–491. https://doi.org/10.1016/S0076-6879(03)74020-8
73. Biasini M, Bienert S, Waterhouse A et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258. https://doi.org/10.1093/nar/gku340
74. Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res 29:221–222
75. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27:343–350. https://doi.org/10.1093/bioinformatics/btq662
76. Xu D, Zhang Y (2012) *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins Struct Funct Bioinforma 80:1715–1735. https://doi.org/10.1002/prot.24065
77. Brylinski M, Skolnick J (2008) A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc Natl Acad Sci U S A 105:129–134. https://doi.org/10.1073/pnas.0707684105
78. Levitt DG, Banaszak LJ (1992) POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. J Mol Graph 10:229–234
79. Trott O, Olson AJ (2010) AutoDock Vina. J Comput Chem 31:445–461. https://doi.org/10.1002/jcc.21334
80. Case DA, Cheatham TE, Darden T et al (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688. https://doi.org/10.1002/jcc.20290

81. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci 3:198–210. https://doi.org/10.1002/wcms.1121
82. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(33–8):27–28
83. Wu K, Wei G-W (2018) Quantitative toxicity prediction using topology based multitask deep neural networks. J Chem Inf Model 58:520–531. https://doi.org/10.1021/acs.jcim.7b00558
84. Huang R, Xia M, Nguyen D-T et al (2016) Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. Front Environ Sci 3:85. https://doi.org/10.3389/fenvs.2015.00085
85. Todeschini R, Consonni V (2009) Molecular descriptors for chemoinformatics, 2nd edn. Wiley-VCH, Weinheim, Germany
86. Damale MG, Harke SN, Kalam Khan FA et al (2014) Recent advances in multidimensional QSAR (4D-6D): a critical review. Mini Rev Med Chem 14:35–55
87. Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design–a review. Curr Top Med Chem 10:95–115
88. Hansch C, Maloney PP, Fujita T, Muir RM (1962) Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. Nature 194:178–180. https://doi.org/10.1038/194178b0
89. Hong H, Xie Q, Ge W et al (2008) Mold ^{2}, Molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. J Chem Inf Model 48:1337–1344. https://doi.org/10.1021/ci800038f
90. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967. https://doi.org/10.1021/ja00226a005
91. Lushington GH, Guo J-X, Wang JL (2007) Whither combine? New opportunities for receptor-based QSAR. Curr Med Chem 14:1863–1877
92. Todeschini R, Gramatica P (1997) The whim theory: new 3d molecular descriptors for QSAR in environmental modelling. SAR QSAR Environ Res 7:89–115. https://doi.org/10.1080/10629369708039126
93. Jain AN, Koile K, Chapman D (1994) Compass: predicting biological activities from molecular surface properties. performance comparisons on a steroid benchmark. J Med Chem 37:2316–2327
94. Polanski J, Bak A (2003) Modeling steric and electronic effects in 3D-and 4D-QSAR schemes: predicting benzoic pK_a values and steroid CBG binding affinities. J Chem Inf Comput Sci 43:2081–2092. https://doi.org/10.1021/ci034118l
95. Hopfinger AJ, Wang S, Tokarski JS et al (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. J Am Chem Soc 119:10509–10524. https://doi.org/10.1021/JA9718937
96. Vedani A, Dobler M (2002) 5D-QSAR: the key for simulating induced fit? J Med Chem 45:2139–2149. https://doi.org/10.1021/jm011005p
97. Vedani A, Dobler M, Lill MA (2005) Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor †. J Med Chem 48:3700–3703. https://doi.org/10.1021/jm050185q
98. Da C, Kireev D (2014) Structural protein-ligand interaction fingerprints (SPLIF) for structure-based virtual screening: method and benchmark study. J Chem Inf Model 54:2555–2561. https://doi.org/10.1021/ci500319f
99. Ash J, Fourches D (2017) Characterizing the chemical space of ERK2 kinase inhibitors using descriptors computed from molecular dynamics trajectories. J Chem Inf Model 57:1286–1299. https://doi.org/10.1021/acs.jcim.7b00048

Ping Gong is Principal Investigator with multidisciplinary expertise in environmental genomics, bioinformatics, bioengineering, mechanistic and predictive toxicology, and molecular modeling. He earned a Bachelor's degree in Environmental Biology and Ecology from Peking University and a PhD in Environmental Toxicology from the Institute of Applied Ecology, Chinese Academy of Sciences. He completed his postdoctoral studies in the Technology University of Berlin (TU Berlin), Swedish University of Agricultural Sciences (SLU), and Biotechnology Research Institute of National Research Council of Canada. As a research biologist, he has led a number of multi-year projects that have greatly improved the fundamental and mechanistic understanding underlying the observed toxicities of military unique and other environmental contaminants. His current research interests focus on mode of action-guided predictive toxicology, genetic variations-conferred herbicide resistance, epigenetics-driven transgenerational inheritance of phenotypic traits, and synthetic biology-based bioengineering and biocontrol of invasive species. He has advised more than 18 graduate students and published over 75 peer-reviewed research articles and book chapters. He also served on the Board of Directors of MidSouth Computational Biology and Bioinformatics Society (MCBIOS) as well as the editorial boards of Frontiers in Genetics and Environmental Toxicology and Chemistry.

Sundar Thangapandian is Computational Chemist/Biophysicist at Environmental Laboratory of US Army Engineer Research and Development Center (ERDC) in Vicksburg, MS with an appointment at Oak Ridge Associated Universities (ORAU). His current research focuses on development of computational toxicity prediction efforts. Before joining ERDC, he was a postdoctoral research associate at the Theoretical and Computational Biophysics Group (TCBG), University of Illinois at Urbana-Champaign (UIUC), IL and spent two years in designing novel therapeutics for the pharmaceutical industry. He earned his Ph.D. in computational biophysics from Gyeongsang National University, South Korea.

Yan Li is Senior Scientist with Bennett Aerospace, Inc. He earned his Ph.D. in Computational Chemistry from the Institute of Process Engineering, Chinese Academy of Sciences, Beijing, China in 2002. As a computational chemist, Dr. Li has extensive experience in solving biological and chemical questions using computational methods. He has been closely collaborating with multidisciplinary scientists in the discovery of anti-cancer agents and development of toxicity databases. His research interests include molecular dynamics simulation, computer-aided drug discovery and design, receptor–ligand interaction, computational toxicity prediction, and protein structure prediction.

Gabriel Idakwo is Ph.D. student and research assistant in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi. His research has focused on machine learning applied to problems in the field of computational chemistry, including chemical structure-activity relationship modeling, toxicity prediction, and feature selection. Gabriel received his BS from Ahmadu Bello University in 2008 and his MS from the University of Southern Mississippi in 2014.

Joseph Luttrell IV is Ph.D. student and research assistant in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi. His research has focused on machine learning and its application to problems in the areas of bioinformatics and image processing, including protein residue–residue contact prediction, face recognition, and object detection. Joseph received his BS in computer science from the University of Southern Mississippi in 2016.

Minjun Chen is Principal Investigator working at the Division of Bioinformatics and Biostatics of the US FDA's National Center for Toxicological Research and serves as the adjunct faculty and mentor for the bioinformatics program offered jointly by the University of Arkansas at Little Rock (UALR) and the University of Arkansas for Medical Sciences (UAMS). He received the FDA award for outstanding junior investigator (2012), the NCTR scientific achievement award (2014)

and the FDA scientific achievement award (2017). He has been co-chairing the FDA Liver Toxicity Working Group since 2014 and edited the book "Drug-induced Liver Toxicity" that was published by the Springer in 2018. His primary research interests encompass drug-induced liver injury, drug safety, bioinformatics, and personalized medicine. He has authored or co-authored more than 90 original publications and book chapters.

Huixiao Hong is Chief of Bioinformatics Branch, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research (NCTR), US Food and Drug Administration (FDA), working on the scientific bases for regulatory applications of bioinformatics. Before joining the FDA, he was Manager of Bioinformatics Division of Z-Tech, an ICFI company. He was a research scientist at Sumitomo Chemical Company in Japan and a visiting scientist at National Cancer Institute at National Institutes of Health. He was also Associate Professor and Director of Laboratory of Computational Chemistry at Nanjing University in China. Dr. Hong is a member of the steering committee of OpenTox, a member of board directors of US MidSouth Computational Biology and Bioinformatics Society, and is in the leadership circle of US FDA modeling and simulation working group. He published more than 180 scientific papers and served as Editor-in-Chief and Editorial Board member for multiple peer-reviewed journals. He received his Ph.D. from Nanjing University in China in 1990 and conducted research in Leeds University in England in 1990–1992.

Chaoyang Zhang is Professor in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi (USM). His research interests include data mining, machines learning, big data analytics, and their applications to interdisciplinary areas of bioinformatics, health care, and toxicity analysis. His research has been supported by the National Science Foundation (NSF), Department of Defense (DOD), National Academy of Sciences (NAS), and National Institute of Health (NIH). Dr. Zhang received his MS in computer science and Ph.D. in computational analysis and modeling from Louisiana Tech University. Prior to joining USM in 2003, he was a research assistant professor in the Department of Computer Science at the University of Vermont.

Chapter 7
A Review of Feature Reduction Methods for QSAR-Based Toxicity Prediction

Gabriel Idakwo, Joseph Luttrell IV, Minjun Chen, Huixiao Hong, Ping Gong and Chaoyang Zhang

Abstract Thousands of molecular descriptors (1D to 4D) can be generated and used as features to model quantitative structure–activity or toxicity relationship (QSAR or QSTR) for chemical toxicity prediction. This often results in models that suffer from the "curse of dimensionality", a problem that can occur in machine learning practice when too many features are employed to train a model. Here we discuss different methods of eliminating redundant and irrelevant features to enhance prediction performance, increase interpretability, and reduce computational complexity. Several feature selection and extraction methods are summarized along with their strengths and shortcomings. We also highlight some commonly overlooked challenges such as algorithm instability and selection bias while offering possible solutions.

Keywords Molecular descriptors · Feature selection · Feature extraction · Toxicity prediction · Machine learning · Quantitative Structure–Activity or toxicity relationship (QSAR or QSTR) · Prediction accuracy

G. Idakwo · J. Luttrell IV · C. Zhang (✉)
School of Computing Sciences and Computer Engineering,
University of Southern Mississippi, Hattiesburg, MS, USA
e-mail: Chaoyang.Zhang@usm.edu

G. Idakwo
e-mail: Gabriel.Idakwo@usm.edu

J. Luttrell IV
e-mail: Joseph.Luttrell@usm.edu

M. Chen · H. Hong
Division of Bioinformatics and Biostatistics, National Center for Toxicological Research,
US Food and Drug Administration, Jefferson, AR, USA
e-mail: Minjun.Chen@fda.hhs.gov

H. Hong
e-mail: Huixiao.Hong@fda.hhs.gov

P. Gong
Environmental Laboratory, US Army Engineer Research and Development Center,
Vicksburg, MS, USA
e-mail: Ping.Gong@usace.army.mil

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_7

Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
4D	Four-dimensional
ACO	Ant colony optimization
ECFP	Extended connectivity fingerprints
GA	Genetic algorithm
KPCA	Kernel principal component analysis
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
LOOCV	Leave-one-out cross-validation
MACCS	Molecular access system
MDS	Multi-dimensional scaling
PCA	Principal component analysis
PSO	Particle swarm optimization
QSAR	Quantitative structure–activity relationship
QSTR	Quantitative structure–toxicity relationship
RFE	Recursive feature elimination
SA	Simulated annealing
SAR	Structure–activity relationship
SFFS	Sequential floating forward selection
SFS	Sequential forward selection
STR	Structure–toxicity relationship
SVM	Support vector machine
Tox21	Toxicology in the twenty-first century
t-SNE	t-Distributed stochastic neighbor embedding

7.1 Introduction

The limitations of in vivo and in vitro approaches for determination of the biological activity of chemicals have fostered the development of in silico approaches [1]. In silico predictive toxicology is designed to complement experimental efforts with a view toward improving the quality of toxicity predictions for safety assessment while decreasing the associated time, cost, and ethical conflicts (animal testing) [2–4]. Methodology for in silico predictive toxicology has been dominated by (quantitative) structure–activity or toxicity relationship [(Q)SAR or (Q)STR] (hereafter called SAR). Traditional SAR models describe a relationship between the chemical structure of molecules (numerically encoded as molecular descriptors) and their activity against a specific biological target [1]. This is achieved by establishing a trend in the molecular descriptor space that links to a biological activity. Thus, all SAR models

are developed on the assumption of a similarity principle. That is, molecules with similar structures (and descriptors, consequently) will have similar biological activity [4, 5]. A SAR model to predict toxicity (T) is given in Eq. (1)

$$T = g\left(D_f\right) \tag{1}$$

where $\left(D_f\right)$ represents the feature space of molecular descriptors as chemical properties and g is a function that relates T to $\left(D_f\right)$ [2]. The accuracy of the model or function g has been shown to depend on the most representative set of molecular descriptors that will encode the useful properties of the molecules for prediction.

Molecular descriptors, being numerical features extracted from molecular structures, are the most common variables used for SAR-based toxicity prediction modeling [6]. The information encoded by descriptors depends on the molecular representation or "dimensionality" of the compound as well as the algorithm used to calculate the descriptors [7]. One-dimensional (1D) descriptors are scalars encoding physiochemical properties (molecular weight, *logP*) and constitutional parameters, such as number of atoms, bond count, atom type, ring count, and fragment counts. 1D descriptors are insensitive to the topology of the molecule and tend to be similar for distinct compounds. As a result, they are often used in combination with other descriptors. Two-dimensional (2D) descriptors are more frequently used for chemical space description. 2D descriptors, including topological indices and structural fragments, are calculated from the connection table (chemical graph) representation of a molecule. They are not only independent of the conformation of the molecule but also graph invariant (not sensitive to altering the number of graph nodes). Three-dimensional (3D) descriptors provide a more complete characterization of molecular structures. 3D descriptors require conformational searching and can discriminate between isomers; this comes at the price of being computationally expensive. The ability to discriminate between isomers can translate to less redundant features. Examples of 3D descriptors include geometric, electrostatic, quantum chemical, and WHIM & GETAWAY. Four-dimensional (4D) descriptors are much like 3D descriptors that evaluate multiple structural conformations simultaneously. Fingerprints are another form of molecular descriptors [7–9]. Commonly used fingerprints include the Molecular ACCess System (MACCS) [10] substructure fingerprints, PubChem [11], and extended-connectivity fingerprints (ECFP) [12]. These fingerprints and 2D descriptors were widely used in the Tox21 data challenge [13] where the winning submissions used over 2500 predefined features covering a wide range of data from topological and physical properties to fingerprints [14].

As shown above, the chemical structures used in SAR modeling are characterized by many molecular descriptors. It is common to generate thousands of descriptors for a single molecule [14]. It is well known that the accuracy of predictive models is not positively correlated to the dimensionality of the data, as overfitting tends to become an issue [15–17]. High-dimensional spaces are prone to include irrelevant and noisy features [18]. SARs developed using such features tend to focus on the peculiarities of molecules and fail to be generalizable [19]. In the chemical space

for a given library, each descriptor adds a dimension to the n-dimensional chemical space. Every molecule in the library is assigned a coordinate depending on its values for all the descriptors. A reduction in the dimensionality of the chemical space correlates with an increasing similarity between molecules. This is important because the underlying assumption in SAR modeling posits that molecules with similar structures should have similar activity [20, 21]. Thus, one of the most important tasks prior to modeling is dimension reduction focused on keeping the most important and relevant descriptors with the maximum amount of biologically meaningful information required for predicting the desired toxicity end point. Shen et al. [13] demonstrated the usefulness of feature selection for toxicity prediction, particularly for interpreting the role of the features. By reducing the feature space, they were able to pinpoint *MolRef* and *AlogP* as the most important descriptors for predicting the toxicity of aromatic compounds.

In simple terms, dimensionality reduction is considered desirable for activity prediction modeling for the following reasons [22]:

(i) Employing fewer descriptors means that the model can focus on important information for establishing a relationship, thus improving prediction accuracy and reducing overfitting (Models with many features enjoy more discriminating power during training but are often not generalizable).
(ii) As the number of features decreases, interpretability of certain models increases.
(iii) Computational costs reduce significantly as the complexity of many learning algorithms is greater than linear [19, 23].
(iv) Elimination of irrelevant descriptors can help remove activity cliffs [7].
(v) Machine learning algorithms are statistical in nature; hence, they suffer from the "curse of dimensionality", which is common with optimization problems as described by Bellman [24].

As the dimensionality increases, the amount of data needed to develop generalizable models increases exponentially [25, 26]. SAR data rarely have an abundance of labeled molecules and, as such, the final model and resulting toxicity prediction will benefit from a reduction in dimension as a smaller dimension means fewer samples will be required during training. The optimal subset of a feature space is one which has the least number of dimensions yet offers the best learning accuracy [26]. Two techniques used to alleviate the challenges of high dimension in SAR datasets include feature selection and feature extraction.

In this review, we discuss different methods for both feature selection and feature extraction techniques, as well as their applications in SAR modeling. In the next two sections, we discuss feature selection and feature extraction methods consecutively. In the last section, we highlight important aspects that must be considered while attempting feature space reduction, such as the stability and validation of the methods.

7.2 Feature Selection

Feature selection works by selecting a subset of features from the original feature set and removing irrelevant features without altering the original representation of the data, on the basis of certain relevance criteria [18, 26–28]. The physical meanings of the features are retained.

Mathematically, considering a descriptor space $X = \{x_i, i = 1 \ldots n\}$, find a subset Y_k (with $k < n$) that maximizes an objective function $J(X)$ for the probability P that a compound is correctly predicted as active or inactive using Eq. (2).

$$Y_k = \left\{x_{(1),} x_{(2),} \ldots, x_{(k)}\right\} = \text{argmax}_{Y_k \subseteq X} J(Y_k) \tag{2}$$

Thus, the ultimate goal of feature selection is to define a subset of Y_k relevant descriptors (obtained from an initial set of X descriptors) which holds the most useful molecular structure information for learning the underlying pattern present in the data.

One pronounced benefit of feature selection is that it can be used to avoid overfitting. Models with high dimension offer many degrees of freedom and tend to learn random patterns and noise instead of important underlying patterns between descriptors and the target end point [29, 30]. Many feature selection algorithms have been documented. Broadly, these algorithms can be grouped into the following three categories depending on the availability of class labels for the training set: supervised [22, 25, 28, 31], semi-supervised [18, 32], and unsupervised [18, 33]. The choice of an appropriate method is dependent on the learning algorithm to be employed and the data to be used [34]. The focus of this review is on supervised feature selection methods. Supervised feature selection requires that the entire training dataset be labeled. Feature selection is achieved by eliminating descriptors that have a low correlation with the toxicity end point to be predicted [28]. Feature selection methods applied to supervised tasks can be classified into filter, wrapper, and embedded methods [28]. We discuss each of these methods and further describe Hybrid [35, 36] and Ensemble [37–39] methods, which are a blend of the earlier listed methods. These methods are illustrated in Fig. 7.1.

7.2.1 Filter

Filter methods evaluate the relevance of a feature based on its intrinsic properties and are completely independent of the learning algorithm [18, 27, 28, 40]. The majority of filter methods are univariate, where each feature is considered independently of the feature space. Multivariate methods, such as correlation-based scores and paired -scores, have also been used to assess the relevance of feature pairs and how well they synergize to enhance prediction of the desired end point [41]. Filter methods are computationally efficient and fast in comparison with wrapper methods. Their lack

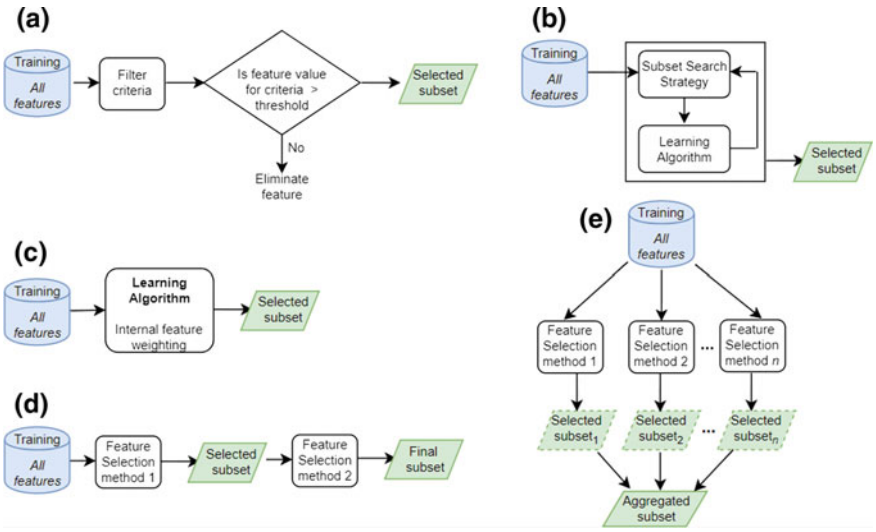

Fig. 7.1 An illustration of different feature selection methods: **a** Filter **b** Wrapper **c** Embedded **d** Hybrid **e** Ensemble

of dependence on any learning algorithm means that the features they select can be used with almost any learning algorithm. However, this independence often results in varied performance from these different learning algorithms [28]. Statistical methods make the assumption that the data they are applied on are normally distributed [40]. By not taking the learning algorithm into consideration, filter methods also turn a blind eye to the heuristics and biases of these algorithms, which may impair their predictive abilities [25].

Filter methods use feature ranking and filtering techniques as the basis for selection. Features are first evaluated and ranked based on a criterion. Then, a threshold is used to select all features above the mark that are considered to be relevant for predicting the end point [18, 28, 41], as shown in Fig. 7.1a. The elimination of low-variance and highly correlated descriptors is a common filtering technique applied to SAR datasets [14, 23, 42]. Several criteria have been employed for filtering descriptors, including variance score [32], correlation coefficient [25, 34], fisher [28, 43], and information gain [44].

7.2.2 Wrapper

Wrapper methods use learning algorithms to evaluate the relevance of a feature, where the learning algorithm's error rate or accuracy is treated as the objective function/criterion for evaluating a feature. A wrapper method begins by selecting a subset of the features heuristically or sequentially, and then a learning algorithm

of choice is used to evaluate this subset. This process of subset generation and testing is repeated until the desired objective function is achieved [27, 28] (Fig. 7.1b). Wrappers tend to perform better than filters in selecting features since they consider feature dependencies and directly incorporate the specific biases and heuristics of the learning algorithm into the selection process. However, this implies that the selected features are unlikely to be optimal for any other classifiers [18].

The size of search space for m features is $O(2^m)$ [28]. Since evaluating the subsets of such a search space is considered an NP-hard problem, the computational inefficiency of wrappers becomes evident when using larger datasets. However, search algorithms have been proposed for selecting optimal subsets of the feature space. Broadly, we consider two groups of search strategies for wrappers: sequential and heuristic selection algorithms [25].

7.2.2.1 Sequential Selection Algorithms

Sequential selection can be achieved in two ways: forward selection and backward elimination. Sequential forward selection (SFS) begins with an empty set of features, and features are progressively incorporated into larger and larger subsets (one at a time) until no further improvement is recorded in the evaluation criterion. A backward elimination algorithm begins with the full set of features and iteratively eliminates the least relevant features [28].

The sequential floating forward selection (SFFS) [45, 46] algorithm has been suggested as an improvement over SFS because it includes flexible backtracking capabilities. Similar to SFS, SFFS adds one feature at a time as determined by the objective function. Meanwhile, it backtracks by eliminating one feature at a time from the initial subset, followed by an evaluation. If an improvement is noticed in the objective function, it leaves that feature out and moves on to add a new feature. This process goes on iteratively until the desired goal is met with the fewest number of features.

7.2.2.2 Heuristic Selection Algorithms

Heuristic search algorithms evaluate different subsets to optimize the objective function. Subsets can be generated by evaluating a search space or by generating solutions to the optimization problem, with the learning algorithm's performance being the objective function [25]. Simulated annealing (SA) [47] and genetic algorithms (GA) [48], two widely used heuristic algorithms, find a subset of features for wrappers. A hybrid of these methods has also been suggested [49]. In GA, the chromosome bits indicate if a feature should be included or not. SA, a stochastic algorithm, solves for the global minimum of a function by improving the initial solution repeatedly using small local perturbations until no such perturbations yield an improvement in the objective function. This process is randomized such that there are occasional and intentional deviations from the solution to lessen the probability of becoming stuck

in local optima. The use of GA to preselect descriptor subsets for SAR modeling of artificial and real data was shown to be successful in [13] where 2D descriptors were employed to discriminate between active and inactive compounds. Particle swarm optimization (PSO) [47] and ant colony optimization (ACO) [50] algorithms may also be employed for heuristic subset search. For instance, it has been shown that the ACO algorithm is a useful method for selecting descriptors for predicting cyclooxygenase inhibitors [50].

7.2.3 Embedded

Embedded feature selection methods incorporate feature selection into the model training process. Embedded feature learning, much like wrapper methods, takes the potential dependencies among features into consideration while being more computationally efficient and less prone to overfitting as compared to wrappers [18, 27, 28, 41]. A common embedded feature selection algorithm is random forest. A random forest is an ensemble of learners with a built-in mechanism for feature selection, such as ID3 and C4.5 [28, 51]. Base learners, i.e., decision trees, look at each feature in the feature space individually and assign importance to them based on how well they contribute to the model attaining an optimal fit. Features with the lowest importance are discarded, and the forest with the least number of features and highest predictive performance is selected [28] (Fig. 7.1c). Using the top 20 molecular descriptors from the random forest predictor importance method, Newby et al. [44] obtained more accurate decision tree classification models in most cases, compared to the use of filter methods such as information gain, chi-square, and greedy search.

Pruning is another embedded feature selection approach that has been applied to neural networks as well as classical learning algorithms, specifically support vector machines (SVMs) [25]. For instance, SVM-recursive feature elimination (SVM-RFE) begins with all the features and recursively removes features that do not contribute positively to the model's predictive accuracy. To determine the optimal number of features for an RFE-based model, cross-validation is used to evaluate and select the subset with the best performance. Hence, RFE can select the best features for a specific learning algorithm. RFE is considered to be computationally expensive as it traverses through all the features one after the other [41]. Weighted Kernels [49] and regularization methods [52], like Lasso, Ridge and Elastic net, have also gained prominence.

7.2.4 Hybrid and Ensemble Feature Selection

Hybrid methods for feature selection involve combining at least two different methods and applying them, usually in succession. Hybrid methods attempt to take advantage of the benefits of the constituent methods while leveraging their strengths. In

Table 7.1 A summary of feature selection techniques

Methods	Description	Strengths	Weaknesses	Examples
Filter	• Rank features using a criterion calculated based on the data properties	• Fast, computationally inexpensive, and as such, can be applied to higher dimensions of data • Multivariate methods take the relationship between features into consideration	• Univariate methods ignore feature dependencies • Insensitive to the learner's heuristics • Deciding on the best threshold when selecting from ranked features is not deterministic	• Information gain • Chi-square test • Fisher score • Correlation coefficient • Variance threshold
Wrapper	• Use search strategies to generate feature subsets which are then evaluated by a learner	• Dependencies between features in a subset are considered • Interaction with the learner results in better performance than filter	• Features are learner specific • Interaction with the learner increases the likelihood of overfitting • Computationally expensive	• Sequential feature selection or elimination (e.g. RFE) • Genetic algorithm • Simulated annealing
Embedded	• Are learning algorithms that can weigh the contribution of each feature to its performance	• Interacts with the learner but is less prone to overfitting • Computationally less expensive than wrapper and has better performance than filter • Dependencies between features are inherently considered	• Features selected are learning algorithm specific	• LASSO • Ridge Regression • Elastic Net • Decision Trees
Hybrid	• Combines other methods to achieve the accuracy of wrappers and the efficiency of filters	• Better performance than filters and less computationally demanding than wrappers	• The setbacks of the filter and wrapper methods are not eliminated, they are reduced. The features remain specific to the learning algorithm	• Filter followed by embedded methods • Hybrid genetic algorithms
Ensemble	• Aggregates the output of different feature selection methods or subsets	• Ensures stable and robust feature selection	• Depending on the constituent methods, it could be computationally expensive and difficult to understand	• Could be made up of multiple feature selection methods

the literature, the most reported is the combination of filter and wrapper methods. Their use has been widely reported for biomedical data [35]. Hsu et al. [49] separately filtered two sets of features using F-score or information gain as the filtering criterion. The resulting features were combined and further treated with wrappers (Fig. 7.1d). They reported improved predictions in comparison with using filters alone and a decreased computational time compared to using wrappers only. Reddy et al. [53] applied a hybrid GA-based descriptor optimization technique for consistently selecting descriptor subsets that represented the whole initial descriptor space. The weights of the selected subsets were analyzed to understand the contribution of each feature to the prediction of HIV protease inhibitors, revealing the role of hydrophobic interactions. This implies the interpretability of the method.

Ensemble methods represent the application of a feature selection method on different subsets of features obtained by using subsampling strategies like bootstrapping. The resulting features from each of the subsets are aggregated using mean, weights, or simple linear aggregation [38, 39] (Fig. 7.1e). This method is often used to deal with the challenges of perturbation and instability experienced by most feature selection methods. Seijo-Pardo et al. [39] provided an in-depth discussion of ensemble methods of feature selection. Dutta et al. [54] proposed an ensemble descriptor selection that searches for descriptor subsets using a genetic algorithm whose objective function is a linear combination of the root-mean-square deviation (RMSE) of all the models in the ensemble. They reported an improvement and found that the resulting model had good performance on the PDGFR and COX-2 datasets. A 96% reduction in noise and an improvement in performance was reported by Zhu et al. [55], using a recursive random forest to rule out a quarter of the least important descriptors at each iteration. This performed better than the least absolute shrinkage and selection operator (LASSO). The authors highlighted that the difference between the prediction performance of random forest and LASSO mainly resulted from the use of variables selected by different strategies, rather than from differences between the learning algorithms.

We have summarized the characteristics, strengths, and weaknesses of the five classes of feature selection methods described above in Table 7.1 in order to assist a user in choosing the appropriate tool based on user-specific requirements and/or goals.

7.3 Feature Extraction

The algorithms employed for mathematical representation of molecular descriptors and fingerprints are independent of the size of molecules, allowing the generation of a fixed length set of descriptors for every molecule regardless of size [7]. The generation of fixed length vectors can introduce redundant descriptors for certain molecules within a library. An optimized feature set achieved by feature extraction can minimize redundancy, noise, correlation between descriptors, and consequently generate classifiers with improved prediction accuracy [20].

A mathematical description of feature extraction is as follows: Considering a descriptor space, $x \in R^n$, find a mapping $y = f(x)$ to obtain transformed feature vector y, where $y \in R^k$ and $k < n$. The vector y should preserve the majority of molecular information in R^n. The goal is to achieve a reduction in dimension without negatively impacting the prediction performance. An optimal mapping, $y = f(x)$, is one that minimizes the prediction error.

Feature extraction transforms the initial feature space to a new, lower dimension feature space by combining the features in the original space. As a result, it is difficult to associate the new features with the old. Further analysis, such as feature importance explanation, becomes very difficult as there is no physical meaning for the newly mapped features that are obtained from feature extraction. Here we discuss some commonly used feature extraction techniques.

7.3.1 Principal Component Analysis

Principal component analysis (PCA) is a multivariate, nonparametric method employed for dimensionality reduction [56, 57]. It works by performing a linear combination of the features, also referred to as the principal components, to achieve the maximum variance. At its core, PCA is centered on determining the eigenvectors of the input data's covariance matrix. This linear transformation can minimize redundancy and reduce the number of features, which increases the information in the resulting features. Each of the resulting features, called principal components, is a combination of several original features. These principal components are also highly uncorrelated because the first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible [26]. A detailed discussion on the different applications of PCA in SAR modeling was provided in [57]. Klepsch et al. [58] applied PCA to a curated P-glycoprotein inhibitors data set of 1608 compounds, where the first two principal components were reported to explain 71.7% of the variance in the dataset. This approach was applied to classification and an analysis into the effect of the initial descriptors on these two components showed that hydrophobic information, such as the number of aromatic bonds and the partition coefficient, was the major contributor to the principal components. According to [59], 2-aryl-1,3,4-Thiadiazole derivatives were classified into distinct clusters of active or inactive molecules when PCA was performed instead of using all of the descriptors calculated.

Considering that principal components are combinations of the original features, all the original features are still available within the components. This is useful for interpretation of models because knowing the original features that contribute to a component can reveal the types of features that are closely related. A key challenge with PCA is that it is unable to handle data with complicated structures that may not be represented in a linear subspace [60]. Kernel PCA (KPCA) [61, 62] was designed to serve as the nonlinear form of PCA. KPCA is based on kernel functions that

intrinsically perform a nonlinear mapping of the input space to a feature space followed by performing linear PCA in this feature space. KPCA generated vectors have been used to train SVM models [59], and it was shown that KPCA is efficient over a wide range of virtual screening dataset inputs using MACCS and ECFP fingerprints. It was also observed that the KPCA embedding largely depended on the properties of the underlying representation as its performance on the ECFP fingerprint varied with the hashing employed.

7.3.2 Autoencoder

Autoencoders [63, 64] are unsupervised neural networks with an odd number of hidden layers that can be applied for nonlinear feature extraction. They employ the backpropagation algorithm to try to create a set of output values which are equal to the input by minimizing the error between the output and the input layer. The network architecture can be designed such that the middle layer is smaller, i.e., has fewer nodes than the input and output layers (Fig. 7.2). In that case, the network is forced to learn a compact representation (embedding) of the input data [65]. In an early work, Hinton et al. [17] demonstrated that autoencoders generated embeddings of images that were used to reconstruct images. A major drawback of autoencoders is that physical meaning for theoretical insight will be lost. They are also complex to train because they typically require a large amount of training data and a search through many possible hyperparameter values. Blaschke et al. [66] employed generative autoencoders to design new molecules in silico based on

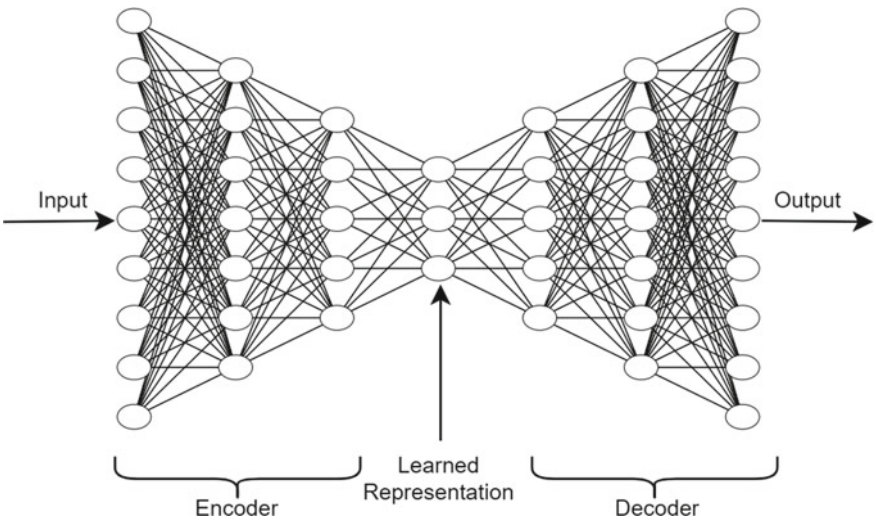

Fig. 7.2 An autoencoder indicating the reduced dimension in the middle layer

the recreated output layer. Burgoon [67] used autoencoders to screen chemicals for potential estrogenic activity by projecting the two neurons in the middle layer into a Cartesian plane. The application of autoencoders for toxicity prediction has not been widely reported, especially for feature extraction. This provides an opportunity for a future area of research.

7.3.3 Linear Discriminant Analysis

Like PCA, linear discriminant analysis (LDA) [65, 68] is a linear transformation technique commonly used for dimensionality reduction. However, LDA is supervised since the discrimination power of the features is taken into consideration. LDA computes an optimal transformation (projection) of the input data on to a line such that classes are separated as clusters. The goal of the projection is to ensure maximum class discrimination by minimizing the within-class distance while maximizing the between-class distance [26]. A weakness of LDA is that if the distribution of a dataset is significantly non-Gaussian, the LDA projections will not be able to preserve any complex structure of the data [69]. Thus, the resulting features may not have good discriminative power. Features extracted with LDA were used by Ren et al. [70] in a stepwise forward manner from a combined pool of experimental data, and chemical structure-based descriptors were employed for predicting aquatic toxicity mode of action. In this work, logistic regression was shown to have a better predictive performance than LDA using the extracted features, with a 7.3% improvement over previously reported classification rates.

In addition to the above-mentioned nonlinear dimensionality reduction techniques, there are also spectral and manifold learning methods, such as t-distributed Stochastic Neighbor Embedding (t-SNE) [71], multi-dimensional scaling (MDS) [72], spectral embedding [73], and isomap [74]. Manifold learning, a class of unsupervised nonlinear algorithms, assumes that the dimensionality of a datasets is only artificially high and thus attempts to uncover the intrinsic low dimensionality. Typically, these algorithms work by computing the similarities between points to find a nearest-neighbor, and then an eigen problem for embedding high-dimensional points into a lower dimensional space [75].

7.4 Miscellaneous

7.4.1 Feature Stability

It is common to use the performance of a model as the metric to evaluate the suitability of a feature reduction algorithm. Therefore, it is an obvious choice to optimize the selection process to obtain the best prediction power possible. However, the stability

or degree of variance of feature selection methods becomes a crucial challenge when the task at hand goes beyond optimizing prediction accuracy to include improving interpretability. A simple scenario may be the case for using substructure-based descriptors for SAR modeling. It is common to consider a substructure that is very relevant for prediction as a major contributor to the activity of that molecule, implying a potential research target. However, many feature selection algorithms tend to be unstable and would yield a different subset if a little perturbation is applied (i.e., when new training samples are added or when some training samples are removed). If every perturbation results in wide variation in the selected subset, then it is difficult to conclude that a feature may be important to the molecule's activity.

Kalousis et al. [76] defined the stability of a feature selection algorithm as "the robustness of the feature subset the algorithm produces in the presence of perturbations in training sets drawn from the same generating distribution." Essentially, stability quantifies how different training sets affect the variation in the selected feature subset. Hence, a similarity measure is often employed to measure the stability of feature selection algorithms. A reliable algorithm should produce the same or similar subset for any perturbations in the training data. Alelyani et al. [77] performed experiments to investigate the causes of instability and reported that dimension, sample size, and the distribution of the training data influenced stability. Larger sample size translated to improved stability, while larger dimensions caused negative effects. Thus, researchers should pay attention to the characteristics of a training dataset. Certain algorithms are also more prone to instability than others. *ReliefF*-based feature selection is affected by the order of samples in a training set, while stochastic search algorithms like GA that use random initialization parameters tend to yield subsets that are unstable [78, 79]. Various metrics for measuring stability have been proposed [78]. To overcome the stability challenge, it has been suggested to employ ensemble selection algorithms based on the technicalities of the selection algorithm in use [78, 80, 81]. Some of these algorithms include Bootstrap sampling, random data partitioning, parameter randomization, or the combination of several of these. Developing algorithms for feature selection that are stable and possess high predictive power is still an open and challenging area. SAR-based toxicity prediction stands to gain a lot from such techniques that can improve speed and accuracy of predictions for regulatory as well as lead optimization purposes.

7.4.2 Validation of Feature Selection

In selecting the optimal feature subset, it is common to evaluate the performance of a learner based on its prediction error. A very common and overlooked mistake is to select features using the entire dataset as a preprocessing step. While this appears to be obviously wrong, it has been reported that many researchers, especially in the biomedical fields, continue to make this mistake and successfully publish in top-ranking journals [82, 83]. If a test set is to be used to evaluate the performance of a feature set, it must not be involved in the feature selection step as that will result in a

selection bias that will yield overly optimistic performance estimates. This is because the features used will have an unfair advantage since they were chosen based on all of the samples. As a result, the model would have gained insight into the features which are more important in the test set. This challenge is more common with wrapper methods [83].

In many practical cases of SAR-based toxicity modeling, there are rarely a large number of compounds across the different end points to be predicted. This makes it difficult to set aside a reasonable batch of data for evaluation purposes. Methods such as cross-validation and bootstrap sampling can be used to avoid sampling bias [34, 82, 83]. Cross-validation techniques like leave-one-out cross-validation (LOOCV) and the *k*-fold method were suggested. Feature selection is to be done in the inner loop of the cross-validation procedure; hence, the algorithm takes the following form for a *k*-fold technique [82]:

(i) Randomly shuffle the data set.
(ii) Randomly split the dataset into K folds.
(iii) For each fold $k = 1, 2,\ldots, K$.

   a. Perform feature selection to obtain an optimal subset with good univariate correlation with the desired end point using all the data except the *k*th fold.
   b. Use the selected features and build a multivariate model with all data except the *k*th fold.
   c. Perform an evaluation using the *k*th fold.

(iv) Aggregate the performance across all K folds to get an unbiased evaluation.

7.5 Summary

QSAR-based predictive toxicity modeling methods are faced with input spaces of thousands of features. To improve the ability of a learner to find a generalizable relationship between molecular descriptors and the toxicity end point of interest, it is expedient to provide the learning algorithm with the minimum number of descriptors while ensuring that the resulting model is interpretable and computationally inexpensive to build. The relevance of a descriptor is assessed by its ability to discriminate between classes in qualitative classification or its correlation to a scalar in quantitative prediction.

In this review, we have discussed different feature selection and extraction methods applicable to SAR-based toxicity modeling. The strengths and weaknesses of each method are highlighted. The choice of which to use should largely depend on the available dataset, and we suggest beginning a new task with a few baseline performance values from a number of methods since no single approach is universally superior. Where the importance of descriptors is sought, feature selection methods such as *filter*, *wrapper*, *embedded* or their combinations (*hybrid* and *ensemble*) may apply. Feature extraction methods transform the features into a lower dimension while

altering the physical meaning of the features. More analysis may be required to interpret the selected features. The stability of selected features and proper feature subset validation methods are often overlooked. Feature selection bias can be avoided by embedding the feature selection process within the inner loop of a cross-validation process to avoid an overly optimistic performance value. Although dimensionality reduction has been shown to improve model performance, there is still room for improvement when it comes to evaluating and validating feature selection and extraction methods and their stability. For the sake of reproducibility, researchers are encouraged to publish important parameters for feature selection or extraction methods they employed, such as the threshold for a variance score. Regardless of the choice of features (molecular descriptors, fingerprints or a combination) used for modeling, SAR models can benefit from dimensionality reduction techniques.

References

1. Lavecchia A (2015) Machine-learning approaches in drug discovery: methods and applications. Drug Discov Today 20(3):318–331
2. Raies AB, Bajic VB (2016) In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley Interdiscip Rev Comput Mol Sci 6(2):147–172
3. Greene N, Pennie W (2015) Computational toxicology, friend or foe? Toxicol Res 4(5):1159–1172
4. Kruhlak NL, Benz RD, Zhou H, Colatsky TJ (2012) (Q)SAR modeling and safety assessment in regulatory review. Clin Pharmacol Ther 91(3):529–534
5. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Mol Inform 29(6–7):476–488
6. Yang H, Sun L, Li W, Liu G, Tang Y (2018) In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. Front Chem 6:30. https://doi.org/10.3389/fchem.2018.00030
7. Danishuddin Khan AU (2016) Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discov Today 21(8):1291–1302
8. Leach AR, Gillet VJ (2007) Molecular descriptors. An introduction to chemoinformatics. Springer, Dordrecht, pp 53–74
9. Todeschini R, Consonni V (2000) Handbook of molecular descriptors. Wiley-VCH, Weinheim
10. Duan J, Dixon SL, Lowrie JF, Sherman W (2010) Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. J Mol Graph Model 29(2):157–170
11. National Institutes of Health (2009) PubChem substructure fingerprint. ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt. Accessed 10 Oct 2018
12. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50(5):742–754
13. Huang R, Xia M, Nguyen D-T et al (2016) Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. Front Environ Sci 3:85. https://doi.org/10.3389/fenvs.2015.00085
14. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. Front Environ Sci 3:80. https://doi.org/10.3389/fenvs.2015.00080
15. Subramanian J, Simon R (2013) Overfitting in prediction models—Is it a problem only in high dimensions? Contemp Clin Trials 36(2):636–641

16. Clarke R, Ressom HW, Wang A et al (2008) The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat Rev Cancer 8(1):37–49
17. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
18. Ang JC, Mirzal A, Haron H, Hamed HNA (2016) Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. IEEE/ACM Trans Comput Biol Bioinform 13(5):971–989
19. Merkwirth C, Mauser H, Schulz-Gasch T, Roche O, Martin Stahl A, Lengauer T (2004) Ensemble methods for classification in cheminformatics. J Chem Inf Comput Sci 44(6):1971–1978
20. Venkatraman V, Dalby AR, Yang ZR (2004) Evaluation of mutual information and genetic programming for feature selection in QSAR. J Chem Inf Comput Sci 44(5):1686–1692
21. Bajorath J (2001) Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. J Chem Inf Comput Sci 41(2):233–245
22. Goodarzi M, Dejaegher B, Heyden YV (2012) Feature selection methods in QSAR studies. J AOAC Int 95(3):636–651
23. Shahlaei M (2013) Descriptor selection methods in quantitative structure—activity relationship studies: a review study. Chem Rev 113(10):8093–8103
24. Bellman R (2016) Adaptive control processes: a guided tour. Princeton University Press, New Jersey
25. Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16–28
26. Van Der Maaten L, Postma E, Van Den Herik J (2009) Dimensionality reduction: a comparative review. J Mach Learn Res 10:66–71
27. Cai J, Luo J, Wang S, Yang S (2018) Feature selection in machine learning: a new perspective. Neurocomputing 300:70–79
28. Tang J, Alelyani S, Liu H (2014) Feature selection for classification: a review. In: Aggarwal CC (ed) Data classification: algorithms and applications, 1st edn. CRC Press, Boca Raton, pp 37–64
29. Johnstone IM, Titterington DM (2009) Statistical challenges of high-dimensional data. Philos Trans A Math Phys Eng Sci 367(1906):4237–4253
30. Zhu X, Wu X (2004) Class noise versus attribute noise: a quantitative study. Artif Intell Rev 22(3): 177 –210
31. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43(1):59–69
32. Sheikhpour R, Sarram MA, Gharaghani S, Chahooki MAZ (2017) A survey on semi-supervised feature selection methods. Pattern Recognit 64:141–158
33. Dy JG, Brodley CE (2004) Feature selection for unsupervised learning. J Mach Learn Res 5:845–889
34. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
35. Solorio-Fernandez S, Martinez-Trinidad JF, Carrasco-Ochoa JA, and Zhang Y-Q (2012) Hybrid feature selection method for biomedical datasets. In: 2012 IEEE symposium on computational intelligence in bioinformatics and computational biology (CIBCB), San Diego, 9–12 May 2012
36. Hsu H-H, Hsieh C-W, Lu M-D (2011) Hybrid feature selection by combining filters and wrappers. Expert Syst Appl 38(7):8144–8150
37. Guan D, Yuan W, Lee YK, Najeebullah K, Rasel MK (2014) A review of ensemble learning based feature selection. IETE Tech Rev 31(3):190–198
38. Brahim AB, Limam M (2017) Ensemble feature selection for high dimensional data: a new method and a comparative study. Adv Data Anal Classif 12(4):937–952
39. Seijo-Pardo B, Porto-Díaz I, Bolón-Canedo V, Alonso-Betanzos A (2017) Ensemble feature selection: homogeneous and heterogeneous approaches. Knowl Based Syst 118:124–139
40. Janecek A, Gansterer W, Demel M, Ecker G (2008) On the relationship between feature selection and classification accuracy. Proc Mach Learn Res 4:90–105

41. Hira ZM, Gillies DF (2015) A review of feature selection and feature extraction methods applied on microarray data. Adv Bioinformatics. https://doi.org/10.1155/2015/198363
42. Rajarshi G, Jurs PC (2004) Development of linear, ensemble, and nonlinear models for the prediction and interpretation of the biological activity of a set of PDGFR inhibitors. J Chem Inf Comput Sci 44(6):2179–2189
43. Guo G, Neagu D, Cronin MTD (2005) A study on feature selection for toxicity prediction. In: Wang L, Jin Y (eds) Fuzzy systems and knowledge discovery. Springer, Heidelberg, pp 31–34
44. Newby D, Freitas AA, Ghafourian T (2012) Pre-processing feature selection for improved C&RT models for oral absorption. J Chem Inf Model 53(10):2730–2742
45. Pudil P, Novovičová J, Kittler J (1994) Floating search methods in feature selection. Pattern Recognit Lett 15(11):1119–1125
46. Brendel M, Zaccarelli R, Devillers L (2010) A quick sequential forward floating feature selection algorithm for emotion detection from speech. In: INTERSPEECH-2010, Chiba, 26–30 September 2010
47. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95—international conference on neural networks, Perth, 27 November–1 December 1995
48. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Longman Publishing Co., Inc, Boston
49. Revathy N, Balasubramanian R (2012) GA-SVM Wrapper approach for gene banking and classificaiton using expressions of very few genes. J Theor Appl Inf Technol 40(2):113–119
50. Shen Q, Jiang J-H, Tao J et al (2005) Modified ant colony optimization algorithm for variable selection in QSAR modeling: QSAR studies of cyclooxygenase inhibitors. J Chem Inf Model 45(4):1024–1029
51. Jain D, Singh V (2018) Feature selection and classification systems for chronic disease prediction: a review. Egypt Informatics J 19(3):179–189
52. Osman H, Ghafari M, Nierstrasz O (2017) Automatic feature selection by regularization to improve bug prediction accuracy. In: 2017 IEEE workshop on machine learning techniques for software quality evaluation (MaLTeSQuE), Klagenfurt, 21 February 2017
53. Reddy AS, Kumar S, Garg R (2010) Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of tipranavir analogs for HIV protease inhibition. J Mol Graph Model 28(8):852–862
54. Dutta D, Guha R, Wild D, Chen T (2007) Ensemble feature selection: consistent descriptor subsets for multiple QSAR models. J Chem Inf Model 47(3):989–997
55. Zhu X-W, Xin Y-J, Ge H-L (2015) Recursive random forests enable better predictive performance and model interpretation than variable selection by LASSO. J Chem Inf Model 55(4):736–746
56. Lauria A, Ippolito M, Almerico AM. (2009) Combined use of PCA and QSAR/QSPR to predict the drugs mechanism of action. An application to the NCI ACAM database. QSAR Comb Sci 28(4):387–395
57. Yoo C, Shahlaei M (2018) The applications of PCA in QSAR studies: a case study on CCR5 antagonists. Chem Biol Drug Des 91(1):137–152
58. Klepsch F, Vasanthanathan P, Ecker GF (2014) Ligand and structure-based classification models for prediction of P-glycoprotein inhibitors. J Chem Inf Model 54(1):218–229
59. Hemmateenejad B, Miri R, Jafarpour M, Tabarzad M, Foroumadi A (2006) Multiple linear regression and principal component analysis-based prediction of the anti-tuberculosis activity of some 2-aryl-1,3,4-Thiadiazole derivatives. QSAR Comb Sci 25(1):56–66
60. Manikandan G, Abirami S (2018) A survey on feature selection and extraction techniques for high-dimensional microarray datasets. In: Anouncia SM, Wiil UK (eds) Knowledge computing and its applications. Springer, Singapore, pp 311–333
61. Reverter F, Vegas E, Oller JM (2014) Kernel-PCA data integration with enhanced interpretability. BMC Syst Biol 8(2):S6
62. Wang Q (2012) Kernel principal component analysis and its applications in face recognition and active shape models. https://arxiv.org/abs/1207.3538. Accessed 10 October 2018

63. Baldi P (2012) Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of ICML workshop on unsupervised and transfer learning, Bellevue, 2 July 2012
64. Goh GB, Hodas NO, Vishnu A (2017) Deep learning for computational chemistry. J Comput Chem 38(16):1291–1307
65. Chandra B, Sharma RK (2015) Exploring autoencoders for unsupervised feature selection. In: 2015 international joint conference on neural networks (IJCNN), Killarney, 12–17 July 2015
66. Blaschke T, Olivecrona M, Engkvist O, Bajorath J, Chen H (2018) Application of generative autoencoder in de novo molecular design. Mol Inform 37(1–2):1700123
67. Burgoon LD (2017) Autoencoder predicting estrogenic chemical substances (APECS): an improved approach for screening potentially estrogenic chemicals using in vitro assays and deep learning. Comput Toxicol 2:45–49
68. Ye J, Ji S (2009) Discriminant analysis for dimensionality reduction: an overview of recent developments. In: Boulgouris NV, Plataniotis KN, Micheli-Tzanakou E (eds) Biometrics: theory, methods, and applications. IEEE Press, Piscataway, pp 1–20
69. Yan H, Dai Y (2011) The comparison of five discriminant methods. In: 2011 International conference on management and service science, Wuhan, 12–14 August
70. Ren YY, Zhou LC, Yang L, Liu PY, Zhao BW, Liu HX (2016) Predicting the aquatic toxicity mode of action using logistic regression and linear discriminant analysis. SAR QSAR Environ Res 27(9):721–746
71. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
72. Borg I, Groenen PJF (2005) Modern Multidimensional Scaling, 2nd edn. Springer Science + Business Media Inc, New York
73. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15(6):1373–1396
74. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323
75. Izenman AJ (2012) Introduction to manifold learning. Wiley Interdiscip Rev Comput Stat 4(5):439–446
76. Kalousis A, Prados J, Hilario M (2007) Stability of feature selection algorithms: a study on high-dimensional spaces. Knowl Inf Syst 12(1):95–116
77. Alelyani S, Liu H, Wang L (2011) The effect of the characteristics of the dataset on the selection stability. In: 2011 IEEE 23rd international conference on tools with artificial intelligence, Boca Raton, 7–9 November 2011
78. Yang P, Zhou BB, Yang JY-H, Zomaya AY (2013) Stability of feature selection algorithms and ensemble feature selection methods in bioinformatics. In: Elloumi M, Zomaya AY (eds) Biological knowledge discovery handbook: preprocessing, mining, and postprocessing of biological data. John Wiley & Sons Inc, Hoboken, pp 333–352
79. Yang P, Ho JW, Yang Y, Zhou BB (2011) Gene-gene interaction filtering with ensemble of filters. BMC Bioinformatics 12:S10. https://doi.org/10.1186/1471-2105-12-S1-S10
80. Yang F, Mao KZ (2011) Robust feature selection for microarray data based on multicriterion fusion. IEEE/ACM Trans Comput Biol Bioinforma 8(4):1080–1092
81. Abeel T, Helleputte T, Van de Peer Y, Dupont P, Saeys Y (2010) Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinformatics 26(3):392–398
82. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning, 2nd edn. Springer-Verlag, New York
83. Ambroise C, McLachlan GJ (2002) Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 99(10):6562–6566

Gabriel Idakwo is Ph.D. student and research assistant in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi. His research has focused on machine learning applied to problems in the field of computational chemistry, including Chemical Structure-Activity-Relationship Modeling, toxicity prediction, and feature selection. Gabriel received his BS from Ahmadu Bello University in 2008 and his MS from the University of Southern Mississippi in 2014.

Joseph Luttrell IV is Ph.D. student and research assistant in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi. His research has focused on machine learning and its application to problems in the areas of bioinformatics and image processing, including protein residue–residue contact prediction, face recognition, and object detection. Joseph received his BS in computer science from the University of Southern Mississippi in 2016.

Minjun Chen is Principal Investigator working at the Division of Bioinformatics and Biostatics of the US FDA's National Center for Toxicological Research and serves as the adjunct faculty and mentor for the bioinformatics program offered jointly by the University of Arkansas at Little Rock (UALR) and the University of Arkansas for Medical Sciences (UAMS). He received the FDA award for outstanding junior investigator (2012), the NCTR scientific achievement award (2014) and the FDA scientific achievement award (2017). He has been co-chairing the FDA Liver Toxicity Working Group since 2014 and edited the book titled "Drug-induced Liver Toxicity" that was published by the Springer in 2018. His primary research interests encompass drug-induced liver injury, drug safety, bioinformatics, and personalized medicine. He has authored or co-authored more than 90 original publications and book chapters.

Huixiao Hong is Chief of Bioinformatics Branch, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research (NCTR), US Food and Drug Administration (FDA), working on the scientific bases for regulatory applications of bioinformatics. Before joining the FDA, he was Manager of Bioinformatics Division of Z-Tech, an ICFI company. He was a research scientist at Sumitomo Chemical Company in Japan and a visiting scientist at National Cancer Institute at National Institutes of Health. He was also Associate Professor and the Director of Laboratory of Computational Chemistry at Nanjing University in China. Dr. Hong is a member of the steering committee of OpenTox, a member of board directors of US MidSouth Computational Biology and Bioinformatics Society, and is in the leadership circle of US FDA modeling and simulation working group. He published more than 180 scientific papers and served as Editor-in-Chief and Editorial Board member for multiple peer-reviewed journals. He received his Ph.D. from Nanjing University in China in 1990 and conducted research in Leeds University in England in 1990–1992.

Ping Gong is Principal Investigator with multidisciplinary expertise in environmental genomics, bioinformatics, bioengineering, mechanistic and predictive toxicology, and molecular modeling. He earned a Bachelor's degree in Environmental Biology and Ecology from Peking University and a Ph.D. in Environmental Toxicology from the Institute of Applied Ecology, Chinese Academy of Sciences. He completed his postdoctoral studies in the Technology University of Berlin (TU Berlin), Swedish University of Agricultural Sciences (SLU), and Biotechnology Research Institute of National Research Council of Canada. As a research biologist, he has led a number of multi-year projects that have greatly improved the fundamental and mechanistic understanding underlying the observed toxicities of military unique and other environmental contaminants. His current research interests focus on mode of action-guided predictive toxicology, genetic variations-conferred herbicide resistance, epigenetics-driven transgenerational inheritance of phenotypic traits, and synthetic biology-based bioengineering and biocontrol of invasive species. He has advised more than 18 graduate students and published over 75 peer-reviewed research articles and book chapters. He also served on the Board of Directors of MidSouth Computational Biology

and Bioinformatics Society (MCBIOS) as well as the editorial boards of Frontiers in Genetics and Environmental Toxicology and Chemistry.

Chaoyang Zhang is Professor in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi (USM). He served as the Director of School of Computing at USM from 2008 to 2014. His research interests include data mining, machines learning, big data analytics, and their applications to the interdisciplinary areas of bioinformatics, health care and toxicity analysis. His research has been supported by the National Science Foundation (NSF), Department of Defense (DOD), National Academy of Sciences (NAS), and National Institute of Health (NIH). Dr. Zhang received his MS in computer science and Ph.D. in computational analysis and modeling from Louisiana Tech University. Prior to joining USM in 2003, he was a research assistant professor in the Department of Computer Science at the University of Vermont.

Chapter 8
An Overview of National Toxicology Program's Toxicogenomic Applications: DrugMatrix and ToxFX

Daniel L. Svoboda, Trey Saddler and Scott S. Auerbach

Abstract DrugMatrix and its automated toxicogenomics reporting system, ToxFX are the scientific communities' largest molecular toxicology reference database and informatics systems. DrugMatrix consists of the comprehensive results of thousands of highly controlled and standardized toxicological experiments where rats or primary rat hepatocytes were systematically treated with more than 600 therapeutic, industrial, or environmental chemicals at both non-toxic and toxic doses. Following administration in vivo, comprehensive studies of the effects of these compounds were carried out after multiple durations of exposure, and in multiple target organs. Study types included pharmacology, clinical chemistry, hematology, histology, body and organ weights, and clinical observations. Additionally, a curation team extracted all relevant information on the compounds from the literature, the Physicians' Desk Reference, package inserts, and other relevant sources. At the heart of the DrugMatrix database are thousands of gene expression data sets generated by extracting RNA from the toxicologically relevant organs and tissues and analyzing these RNAs using the GE Codelink rat array, and the Affymetrix whole-genome 230 2.0 rat GeneChip array systems. Additionally, the database contains 148 scorable genomic signatures, covering 96 distinct phenotypes derive from mining the DrugMatrix gene expression data. The signatures are informative of organ-specific pathology (e.g., hepatic steatosis), and mode of toxicological action (e.g., PXR activation in the liver). The phenotypes cover several common target tissues in toxicity testing (liver, kidney, heart, bone marrow, spleen, and skeletal muscle). Taken as a whole, DrugMatrix

D. L. Svoboda
Sciome LLC, Durham, NC, USA
e-mail: daniel.svoboda@sciome.com

T. Saddler
Kelly Government Solutions, Durham, NC, USA
e-mail: trey.saddler@nih.gov

S. S. Auerbach (✉)
Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Durham, NC, USA
e-mail: auerbachs@niehs.nih.gov

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_8

enables a toxicologist to formulate a comprehensive picture of toxicity with greater efficiency than traditional methods.

Keywords ToxFX · DrugMatrix · Signature · Toxicogenomics · Database · National Toxicology Program · In vivo · In vitro

Abbreviations

API	Application Program Interface
BDE	Brominated Diphenyl Ethers
CEBS	Chemical Effects in Biological Systems
FDA	US Food and Drug Administration
GD	Gestation Day
GUI	Graphical User Interface
NIEHS	National Institute of Environmental Health Sciences
NTP	National Toxicology Program
PND	Postnatal Day

8.1 Introduction and History

The power of transcriptomics for understanding biology relates to the technology's ability to provide a researcher not only the ability to view an entire landscape of molecular alterations but to perform high-dimensional relational analysis that can quantify relative similarities of different biological states. Consistent with this idea, at the inception of the toxicogenomics methodology, it was apparent that in order to achieve contextualization of the complex findings, anchoring of the studies to reference data sets of traditional toxicity metrics would be needed. This realization led to the creation of Iconix Biosciences in the early 2000s [1]. Their mission was to generate a toxicologist friendly resource that integrated pathology, clinical chemistry, and transcriptomics metric all derived from a large set of comprehensive, well-documented short-term in vivo rat toxicity studies [2]. To further broaden the context of the information, targeted high throughput screening assays were performed, rat specific annotated pathways were generated, signatures of pathology were derived from the data, copious pharmacology curation was performed, and toxicologically relevant ontologies were built and curated on several hundred chemicals and drugs. To integrate and provide ease of access to toxicologists, all this information was organized in a relational database, and a sophisticated java-based web application was constructed. A web-browser-based GUI provided simple, end-user query access to the data. It enabled uploading and integration of end-user data, together with data visualization, to identify deeper patterns in the data. This application named DrugMatrix. Due to its robustness of design and detailed consideration of the target

audience, it remains a valuable resource for toxicologists and researchers. It continues to be a, perhaps unsurpassed, example of the power of data integration for genomic toxicology.

DrugMatrix is an exceptional tool for data exploration, and it provides a vast landscape of information that can be challenging to formulate into a single report. Feedback from end users in relation to this issue led Iconix to create ToxFX. By combining the highly curated data in DrugMatrix with templated text processing, ToxFX carries out rapid and automated interpretation and reporting of toxicogenomic study results. In less than 5 min, ToxFX transforms normalized differential gene expression data files, along with simple user-provided data annotations to a detailed report covering numerous levels of interpretation along with supporting data tables, histopathology section images, and other visualizations.

The National Toxicology Program (NTP) took ownership of DrugMatrix and ToxFX and all associated assets in 2010 after it was briefly owned by Entelos. The goal of the acquisition was to make the resources freely available to the research community. Since this time the data, in particular the transcriptomics data, has been extensively reused and broadly circulated. As of this writing, NTP has made the entire contents of the DrugMatrix Database available in multiple forms and has made DrugMatrix and ToxFX available via an open-source distribution to enable users to deploy their own installation of the software.

Throughout this book chapter, different components of DrugMatrix will be referred to. When referring to the DrugMatrix database, we are indicating the relational database (PostgreSql) that holds all the DrugMatrix data that may be downloaded elsewhere (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrixPostgreSqlDatabase.tar.gz). The DrugMatrix-browser-based GUI that allows end users to query the database via a server-based application, will be called the DrugMatrix GUI. These two taken together, along with other associated assets will be referred to as DrugMatrix.

8.2 DrugMatrix Data

The DrugMatrix Database is a highly integrated rat toxicogenomics database containing one of the largest, systematically created collections of data of interest to toxicologists. It is built using the results of short-term toxicogenomic studies on >600 reference drugs/chemicals (433 US Food and Drug Administration (FDA) approved drugs, 63 drugs approved outside the USA, 54 withdrawn or discontinued drugs, 15 standard biochemicals, and 72 standard toxicants) in male Sprague Dawley rats. It is composed of the results of over 5000 Affymetrix 230 2.0 and 12,000 GE Codelink microarray studies performed on samples from liver, kidney, heart, thigh muscle, bone marrow, spleen, brain, intestine, and primary hepatocytes. In addition, 127,000 histopathology, and ~100,000 hematology and clinical chemistry measurements were taken to enable cross-referencing gene expression results in established toxicity metrics. In addition to the data from the animal studies, the database also

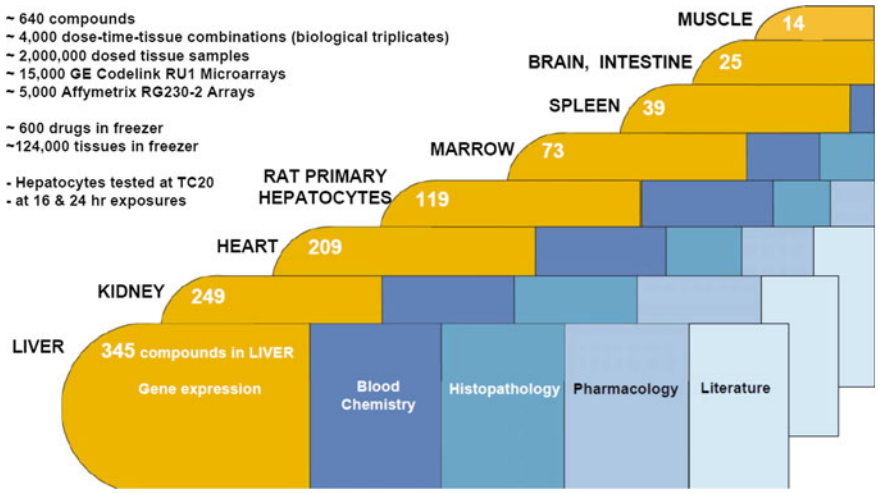

Fig. 8.1 Contents of DrugMatrix. A summary of the contents of the DrugMatrix database is shown. Gene expression studies involving up to 345 compounds (liver) over numerous tissues in primary rat hepatocytes were performed in combination with clinical chemistry, histopathology, in vitro pharmacological assessment, and literature curation. Tissue samples from many of the studies have been retained and available for further investigation

contains ~8000 chemical structures, of which ~2000 have baseline literature curation. There are in vitro molecular pharmacology results for 867 compounds in 130 assays, 900 compounds with detailed literature curation (clinical pharmacology, toxicology, indications, etc), 137 hand-curated signaling pathways, and hundreds of algorithmically derived gene expression signatures (Fig. 8.1).

Figure 8.2 outlines how studies were carried out, the details of which are available in the standard operating procedure manual (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/SOP%20v3%208_NTP.pdf; originally formulated by Iconix Pharmaceuticals). For each in vivo expression study, literature review was undertaken in order to identify a minimum of 3 chemicals that are associated with a similar SAR (e.g. HMGCoA Reductase Inhibitor). Dose-levels and target organs were also identified. A repeated daily dose range finding study was then performed in male Charles River Sprague Dawley rats, identifying maximum tolerated doses for 5-day studies. After reviewing the results of the range finding studies, and considering the literature review, the maximum tolerated dose and fully effective pharmacological dose level for each chemical was identified. Detailed narratives on dose selection processes, and considerations for each study can be found in the Expression Study Domain of DrugMatrix. Using the identified dose levels, an "Array Study" was performed; tissues were collect for transcriptomics, clinical chemistry/hematology and histopathology of organ samples. Detailed histopathology, clinical chemistry, and hematology findings for individual rats can be found in the Chemical Effects in Biological Systems (CEBS) (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix_ClinicalChemistry_Hematology.xlsx and ftp://anonftp.

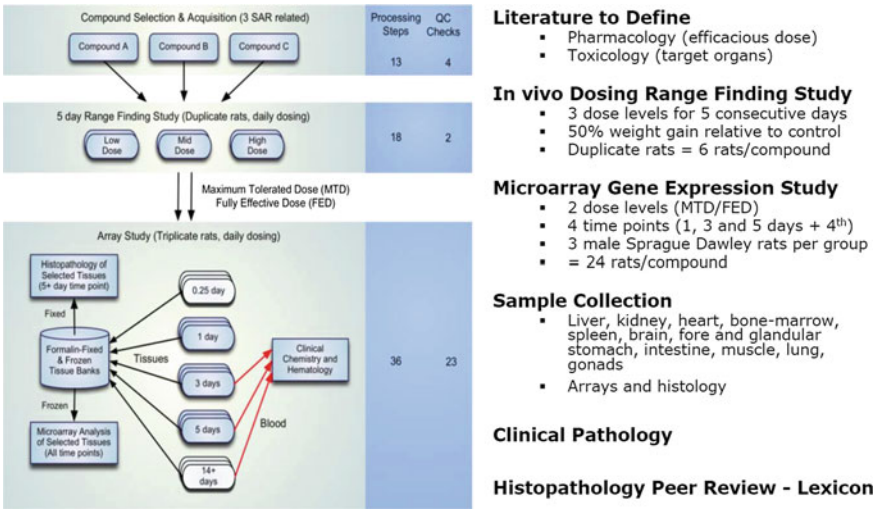

Fig. 8.2 Rat in vivo experimental protocol used for DrugMatrix data generation

niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix_Histopathology.xlsx). In order understand the temporal aspects of toxicity, samples were collected with durations of exposure ranging from 0.25 days to 14 or more days (dose administered daily). For expression studies in rat hepatocytes, a single dose reflective of 24-h "toxic concentration 20" (dose where there is a 20% reduction in cell viability) was dispensed. Hepatocyte gene expression was measured 16 and 24-h post dose delivery. NTP has retained a collection of biological samples from the in vivo studies including snap-frozen liver, heart, kidney, thigh muscle, whole blood and/or plasma that are available upon request for further investigation. A complete list of the available tissue and RNA samples can be found elsewhere (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix_Aliquot%20Information.xlsx and ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix_RNAMicroarray.xlsx).

Initially, GE Codelink microarrays representing approximately 10,000 genes were used to quantify gene expression. In this first phase, >12,000 arrays were run. Due to the evolution of array technology platforms, and ubiquitous usage in the research community, a subset of the RNA samples (>5000) that provided the most informative data were re-analyzed using Affymetrix 230 2.0 microarrays. All raw microarray data is available in the Gene Expression Omnibus database (GEO Data Sets: GSE59913, GSE59923, GSE59894, GSE59895, GSE59905, GSE59906, GSE59907, GSE59925, GSE59926, GSE57800, GSE57805, GSE57811, GSE57815, GSE57816). Data from both platforms are systematically integrated and available through the DrugMatrix Database. All individual treatment transcriptomic signatures, i.e., single dose/duration) from the database are available in CEBS (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/Differential%20Gene%20Expression%20Data/).

Chemical and drug structures are the data type of greatest abundance housed in the DrugMatrix database. Approximately, 8000 chemical structures are curated in the database (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/Drugmatrix_Curated_Chemicals_With_SMILES.txt), with ~2000 of these having some degree of baseline curation and ~800 with full curation. Curation is defined as a collection of facts (e.g., pharmacokinetics, toxicity, pharmacology) describing each compound, recorded in the literature. An integral part of the curation process was the creation of an ontology consistent with terms and logic used in the field of toxicology describing chemical and biological properties. Chemical mapping to ontology terms can be found elsewhere (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/Compound%20Literature%20Annotations/COMPOUND_ANNOTATIONS.txt).

To allow users of DrugMatrix to explore associations between gene expression and effects of drugs and chemicals, a select set of 130 commercially available pharmacological targets were screened using competitive binding assays. The results of these studies can be found elsewhere (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DM_invitro_assay_data.xlsx). These data have also been integrated in number of public HTS databases including ChEMBL and Pubchem (search "DrugMatrix" in these databases to find results).

Distinct species can respond to a chemical challenge in different ways. Most public resources have focused on pathway curation in human or mouse (most common non-human model system for academic studies). Hence, to most effectively interpret rat toxicogenomics data, 137 pathways were curated focusing on a rat biology review of the literature. Detailed citations are provided for the inclusion of genes/proteins, and their linkage to other components of the pathways. All the curated pathways are available for download (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix%20Pathways.zip).

One of the primary goals of acquiring the assets associated with DrugMatrix was to make all data freely available to the research community for mining from a variety of perspectives, using a diversity of computational and bioinformatic approaches. All data resources noted above in addition to additional DrugMatrix data, resources, and information can be found elsewhere (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/).

8.3 DrugMatrix Database

A highly-integrated relational database holds the algorithmically extracted experimental data described above (Fig. 8.3). A detailed description of the database can be found in CEBS (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrixDataWarehouse.pdf). The DrugMatrix data warehouse architecture is a highly denormalized, modified star-schema. The "hubs" of the schema are the six main information domains, or schema dimensions: GENE, COMPOUND, EXPRESSION EXPERIMENT, EXPRESSION STUDY, PATHWAY, and ASSAY (Table 8.1). These hubs represent the main information domains in the DrugMatrix user interface.

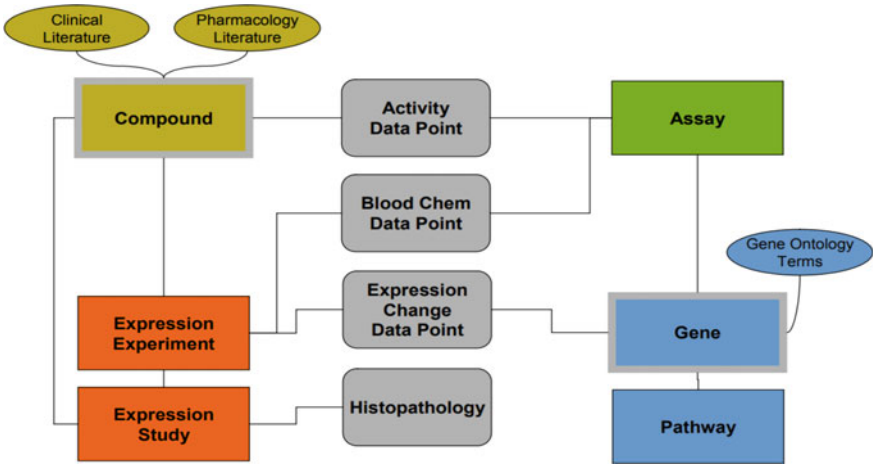

Fig. 8.3 A high-level overview of the main information items in the DrugMatrix data warehouse and user interface. The major information domains used by the DrugMatrix GUI are shown in gold, green, orange, and blue rectangles. This represents a highly simplified in that it represents only 13 of the 70+ tables in the warehouse

The DrugMatrix relational database is available for download (PostgreSql scripts; see below) and will be accessible soon through a database client. A Readme will be available through CEBS which will provide details on how to connect to DrugMatrix PostgreSql Database.

The two primary data domains, GENE and COMPOUND, are connected via annotation of chemogenomic-based secondary domains: ASSAY activities (a direct measure of interaction between a compound and a target) and EXPRESSION array profiles (a measurement of the indirect effects of a compound on the genome). Two secondary domains, PATHWAY and EXPRESSION STUDY, provide extensions of these domains. An additional secondary domain, SIGNATURE, provides patterns of gene expression that can be used to classify compounds, or a set of common experiments reliably compared to the experiment population, or a defined set of control experiments.

The DrugMatrix database contains several types of data that require different statistical modeling and handling, including normalization, ratio calculation, and denominator groupings. While some of the data handling will be discussed herein, a separate document comprehensively covers this topic (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix_Calculations.pdf). This document specifically addresses the key computations currently used for different types of data modeling. A detailed description of raw data handling, such as removal of no-data spots, exclusion of outlier data points, and data normalization is provided. Gene expression changes of treated samples compared to controls, together with the supporting statistics to determine their significance are summarized in detail. The rationale for using different denominators for ratio calculation of gene expression,

Table 8.1 DrugMatrix GUI data domains

Data domains	Contents
Gene	Presents all relevant information from the database for a queried gene, such as annotation, genes that exhibit similar expression patterns, treatments that induce or repress the gene, molecular pharmacology results and activities reported in the literature
Compound	Presents all relevant information from the database for a queried compound, such as chemicals with similar structures, genes that are induced or repressed by the compound, experiments where the chemical was used, molecular pharmacology results for the compound and literature annotations
Assay	Presents all relevant information from the database for a queried molecular pharmacology, clinical chemistry or hematology assay, such as which chemicals were active and details on the assay and its associated gene(s)
Expression	Presents all relevant information from the database for a queried experiment (a single chemical, time point, dose, tissue, species, and array type) such as experiments with similar expression patterns, genes that were induced or repressed in an experiment, and the clinical chemistry, hematology and histopathology results for the experiment
Pathway	Provides information on the hand-curated biochemical, metabolic and signal transduction pathways annotated in DrugMatrix
Expression Study	Presents a list of experiments related to the queried study. A study is a collection of data representing a multi-dose and time point toxicity assessment. Information about a study such as time course, dose justification, pathlab report, and dose justification is found in this domain
Motif	Presents all relevant information from the database for a queried signature motif. A Motif is a pathway-centric gene expression signature that classifies a toxicity endpoint. Information about a motif such as similarity to other motifs, genes contained in the motif, and experiments that is active
Signature	Presents all relevant information from the database for a queried DrugMatrix signature. Signatures are similar to motifs with the exception that the signatures are not limited to pathway genes. Information contained in this domain parallels that for the Motif domain

blood chemistry, and histopathology data is described. The algorithms and computations that form the basis of the drug signatures are elaborated in detail, as well as the analytical tools that are utilized to assist data visualization and interpretation in the DrugMatrix GUI. In addition, compound curation, pathway curation, and gene annotation processes are described in this document.

8.4 DrugMatrix GUI

The DrugMatrix GUI (Fig. 8.4) consists of: (1) A set of sophisticated querying tools that enables queries that filter by expression profile, chemical structure, gene names, compound names, or by the presence of any attribute (e.g., a compound that

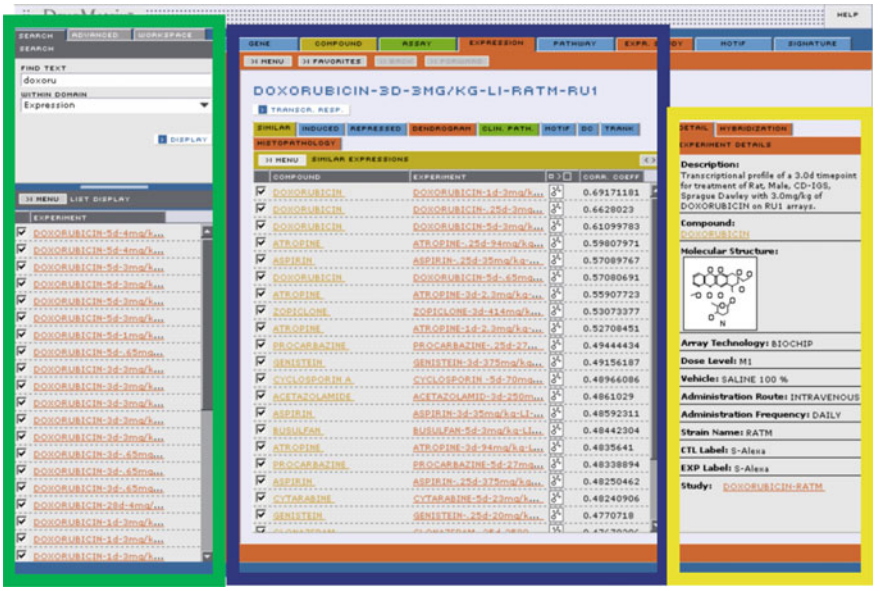

Fig. 8.4 The DrugMatrix GUI. Shown in the figure is an image of the DrugMatrix GUI. Outlined in the green box is the search and list management tools. The Search and List area of the DrugMatrix GUI consists of tools for generating simple searches and advanced searches (i.e., layered queries) of the database as well as a workspace with tools for the management of lists generated by these searches. The workspace section contains the Toolbox which provides several advanced analysis and visualization tools. Selection experiment from the list shown in the lower section of search and list management tools causes the Domain Reports to be populated (dark blue box). Domain reports are organized into seven domains (gene, compound, assay, expression, pathway, expression study, and signature). Each Domain Report contains a subset of Chemogenomic Data Reports and Detail Reports (yellow box). The tabs and structure in the Chemogenomic Data Reports vary depending on the data being queried as does the details report

induces a specific gene and has a specific bio-activity), (2) Comprehensive reporting tools that deliver reports comprising integrated information, including automatically generated links to information from other data domains. For example, the COMPOUND Domain Report provides, among other things, links to expression experiments (and the associated array technology) in which the query compound has been tested, with links to all genes (and related assays) that have been significantly impacted by the query compound, (3) Data visualizations providing extensive detail about query items. The DrugMatrix application uses predefined database searches to extract, and graphically present detailed information relevant to the query item. A document entitled DrugMatrix Reference Guide detailing the DrugMatrix GUI is available through the CEBS database (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix_Ref_Guide.pdf).

Information, extracted from the database, is organized into seven Domain Reports in the GUI that represent each of the DrugMatrix data domains (Table 8.1). Each

Domain Report contains a subset of Chemogenomic Data Reports and Detail Reports that present cross-referenced experimental data from the DrugMatrix database, and curated information from the scientific literature, reference material, and online databases. As operations are executed within the DrugMatrix framework (for example, conduct queries, save lists, or access previously saved lists) Server Status Messages appear at the bottom right of the User Interface, indicating progress of retrieval of the requested operation.

The Search and List area of the DrugMatrix GUI consists of tools for generating Simple Searches and Advanced Searches of the database, as well as a user-owned Workspace with tools for the management of lists generated by these searches. Lists that have been generated by database searches, previously saved user lists that have been retrieved, Data Objects, and Favorite items are all available for use in the Workspace area.

In addition to the data content that can be accessed through the DrugMatrix GUI, several analysis tools are contained within the toolbox component of the interface which enables analysis of uploaded data or data contained in the database (Table 8.2). The TOOLBOX button, located on the WORKSPACE panel, opens a window containing these tools. A list of these tools can be found in Table 8.2. The analysis tools are further described in the reference manual and in the materials and methods of white papers in the help section of DrugMatrix.

A user can perform a wide-ranging number of tasks using the DrugMatrix GUI such as: Upload your own data for analysis or mine the DrugMatrix data; find similar expression profiles; determine significantly up- and down-regulated genes; visualize expression profiles on pathways; construct expression patterns for putative biomarker sets; perform gene ontology analysis of perturbed genes; score gene expression signatures; test performance of a biomarker set for detecting phenotypes; perform hierarchical clustering; find consistently changed genes; identify enriched literature annotations in groups of expression profiles and mine the literature. A detailed tutorial with multiple detailed examples on how to use the DrugMatrix GUI can be found elsewhere (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrix_Tutorial.pdf).

8.5 ToxFX

ToxFX is an automated toxicogenomics analysis and reporting tool. It employs DrugMatrix gene expression data, pathway information, and toxicity signatures to formulate detailed reports, and supplementary results files from user-provided gene expression data sets. Each report contains an executive summary, study description, details of the quality control metrics, relative impact on transcription, DM signature scores, relative response of pathways, cytochrome P450 changes, and most consistent gene expression changes. Less than 5 min is typically required from upload of normalized data to output of a report.

8.6 Example Analysis of dE-71 Gene Expression Data Using DrugMatrix and ToxFX

To demonstrate the utility of DrugMatrix, we present here an analysis of a toxicogenomic study of a mixture of low molecular weight-brominated flame retardants (DE-71) that was demonstrated to be carcinogenic in rat liver [3]. Study results have been published elsewhere, and the reader has referred this manuscript for a more detailed account of the results [4]. In short, pregnant female Wistar Han rats were administered 50 mg/kg DE-71 starting at Gestation Day (GD) 6 of the fetus through Postnatal Day (PND) 12, and then the pups were directly dosed from PND12 to PND21. The pups were euthanized on PND22, 24 h after the last dose, and livers were taken at necropsy of the male pups for microarray studies. Of note is that the dosing paradigm, and the strain and age of sacrifice of rats were different from experiments recorded in the DrugMatrix reference data set. Further, DE-71 treatments are not part of the DrugMatrix reference data set. The gene expression data was generated using Affymetrix whole-genome 230 2.0 microarrays. A total of 10 microarrays were run (5 vehicle and 5 chemical treated; all samples were independent biological replicates).

Table 8.2 List of data analysis tools available in DrugMatrix

Tool	Functionality
List editing tools	Subtract a list (genes, compounds, etc) or group of lists from a list, combine lists, or generate a list of common items between lists
Data translation tools	Convert one type of list of items to another type of list (e.g., a list of genes to a list of associated assays)
Pattern creator	Allows the user to derive a correlation-based signature from an experiment list and gene set
Data import/export tools	Export or import previously saved lists enabling the user archive and/or share lists
Hypergeometric analysis	Performs chemical ontology enrichment analysis of a list of experiments
Expression experiment matrix	Generates a graphical matrix, with a heat-map-type display, to compare expression ratios for a specified list of genes in a specified list of experiments
Bio-activity matrix	Creates a graphical matrix, with a heat-map-type display, to compare the bio-activities from the specified list of assays and compounds
Pathway impact matrix	Examines the impact of compounds, in a specified expression experiment list, on gene pathways
Blood chemistry matrix	Creates a heat map display to examine the perturbation of blood chemistry assays, in a specified expression experiment list

(continued)

Table 8.2 (continued)

Tool	Functionality
Compound compare	Presents a table of two-dimensional images of the molecular structures of the compounds in the specified list of compounds
Pathway visualization	Enables the user to see the impact of compounds on the genes in a pathway of interest at various doses and times
GO data query	Allows the user to analyze a specified gene list using Gene Ontology annotations to find significantly over-represented terms
Significant gene finder	Extracts the most significant set of genes for a list of experiments
Drug signature heat-map	Generates a heat-map view of the signature hits for the experiments
Drug signature histogram	Allows the user to generate a view of SVM signatures displayed as a histogram for the query experiment and the control list
Pattern assessment	Assesses the accuracy of the signature derived from the "Pattern Creator" tool

The data was first normalized using the Affymetrix Expression Console using the Plier algorithm. The normalized CHP files were uploaded into DrugMatrix using the DrugMatrix Study Builder. An initial assessment of the data in the Similarity tab of Expression domain indicated that DE-71 elicited a gene expression pattern that was most similar to Phenobarbital-like inducers, such as Phenobarbital and Dypyrone, consistent with previous observations with other brominated diphenyl ethers (BDEs). A Hypergeometric Analysis (Chemical/drug annotation enrichment analysis) of the 25 most similar reference experiments in DrugMatrix indicated an enrichment of aromatase inhibitors suggesting that DE-71 may be interfering with steroid metabolism, an observation consistent with other studies [5].

A review of the top induced genes presented on the Induced tab indicated induction of Cyp2b1 and Cyp1a1. This is consistent with the PB-like induction properties of the BDEs and the AhR-activating properties of contaminating brominated dioxins/furans, respectively [6]. In addition, there was a striking induction of the urinary protein, Rup2, estrogen sulfotransferase and Sult1e1. A review of the top–down-regulated genes on the Repressed tab, a down-regulation of Cyp17 which in combination with the effects on Rup2 and Sult1e1 suggests a perturbation of the sex steroid signaling cascade and the potential for endocrine disruption which is consistent with the hypogeometric analysis and published findings [5]. Fgf21, a growth factor with antidiabetic properties was down-regulated. In combination with observed induction of Lep this observation suggests the potential for metabolic perturbations associated with DE-71 exposure [7]. Down-regulation of Lrp10 and Abcg8 suggest a potential alteration in cholesterol homeostasis which is consistent with the clinical chemistry results observed in this and other studies [8].

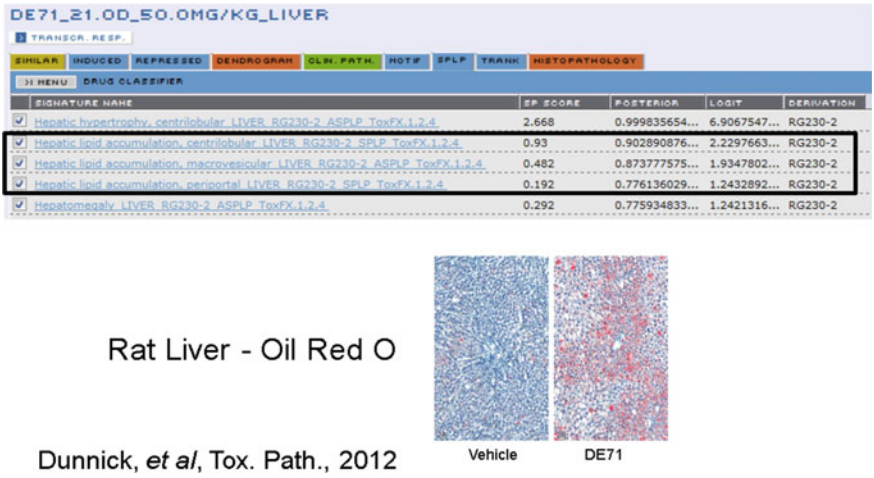

Fig. 8.5 Pathology Signatures Scoring of DE-71 Gene Expression. Scoring of the liver gene expression data using the DrugMatrix signatures identified two possible pathological processes following DE-71 exposure, hepatic lipid accumulation (in black box) and hepatocyte hypertrophy. Follow-up pathological assessment using Oil Red O staining of the DE-71-treated liver revealed clear lipid accumulation in the hepatocytes

A signature scoring analysis produced hits for liver hypertrophy and lipid accumulation. Both effects were observed in the published study, and effects on cholesterol have been documented in an independent assessment of DE-71 toxicity (Fig. 8.5).

A Pathway Impact Matrix analysis (Settings: $P < 0.02$, both up- and down-regulated genes, score type: % changed) indicated Cholesterol Biosynthesis was the most affected pathway. A review of DE-71's effect on the Cholesterol Biosynthesis pathway using the Pathway Visualization tool indicated most of the genes in this pathway were down-regulated by DE-71. The finding related to cholesterol is consistent with other findings this study that are noted above and in independent investigations [8].

For purposes of illustration and comparison with DrugMatrix, an analysis of the DE-71 expression data was performed using ToxFX. In short, ToxFX yielded similar results, suggesting potential effects on liver hypertrophy and steatosis, along with pathway level perturbations of xenobiotic metabolism, AhR signaling, and Cholesterol Biosynthesis. Most notable is the complete analysis, report and supplementary results files were generated in less than five minutes after upload of the normalized CHP files.

8.7 Accessing the DrugMatrix Database

NTP no longer hosts DrugMatrix or ToxFX applications for the public, however, the complete DrugMatrix database can be downloaded (ftp://anonftp.niehs.nih.gov/ntp-cebs/datatype/Drug_Matrix/DrugMatrixPostgreSqlDatabase.tar.gz) and deployed into a local PostgreSql installation using a dump of the PostgreSQL database. Note that the uncompressed database is larger than 10 GB.

8.8 Deploying the DrugMatrix and ToxFX Applications for Local Use

Source code for the DrugMatrix and ToxFX applications have been posted to Github (DrugMatrix: https://github.com/NIEHS/DrugMatrix; ToxFX: https://github.com/NIEHS/ToxFX).

Expertise with Java, Tomcat, and database administration will be required to deploy the applications to a local system. A detailed description on how to deploy the application within your organization can be found at the above links.

8.9 DrugMatrix Data Reuse

There have been numerous publications that have employed the DrugMatrix data to glean insight into toxicological processes. These publications are illustrative of what is possible with a large, well-annotated data set. The publications can be grouped into toxicity signature derivation [9–20], mechanism of action studies [21–25], data mining and biological network creation and analysis [14, 20, 26–30] and technology evaluation [31]. In addition, various components of the DrugMatrix database have been incorporated into other databases and resources [32–40].

8.10 Conclusion and Future Directions

Members of the National Toxicology Program and NIEHS view the acquisition, public-access implementation, and development of the DrugMatrix and ToxFX resources as a step forward in bringing genomics into the arena of regulatory toxicology and furthering the goals of the Tox21 effort to revolutionize toxicology testing. No doubt further work needs to be done to develop additional tools, and a framework for the utilization of toxicogenomics data. Having DrugMatrix and ToxFX, together with their associated background data in the public domain will allow open and crit-

ical assessment of these approaches, which is necessary to vet any new approach to toxicological assessment.

Future plans for DrugMatrix include regular updating of gene, pathway and ontology annotations, integration of additional pathways from additional sources, and broadening of the chemical space of the reference genomic data. Integration of available public data sets, expansion of the number of data platforms and species that can be directly loaded into DrugMatrix for analysis and implementation of APIs to allow for ease of interaction with other popular genomics analysis tools are also on the horizon. Recommendations from end users will of significant value. Send any recommendations to Drugmatrix@nih.gov. Where relevant, the data and functionalities added to DrugMatrix will be implemented in ToxFX.

Acknowledgements The authors would like to acknowledge the large group of former employees of Iconix Biosciences, in particular, Alan Roter for the original data generation and development of DrugMatrix and ToxFX. We would also like to acknowledge the ongoing support of the Sciome LLC and the NTP CEBS database team for support and maintenance of the tools and for help in disseminating the contents of the database. Finally, the authors would like to thank B. Alex Merrick for review of the book chapter.

References

1. Fielden MR, Halbert DN (2007) Iconix biosciences, Inc. Pharmacogenomics 8(4):401–405
2. Ganter B, Tugendreich S, Pearson CI, Ayanoglu E, Baumhueter S, Bostian KA, Brady L, Browne LJ, Calvin JT, Day GJ, Breckenridge N, Dunlea S, Eynon BP, Furness LM, Ferng J, Fielden MR, Fujimoto SY, Gong L, Hu C, Idury R, Judo MS, Kolaja KL, Lee MD, McSorley C, Minor JM, Nair RV, Natsoulis G, Nguyen P, Nicholson SM, Pham H, Roter AH, Sun D, Tan S, Thode S, Tolley AM, Vladimirova A, Yang J, Zhou Z, Jarnagin K (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. J Biotechnol 119(3):219–244
3. Dunnick JK, Pandiri AR, Merrick BA, Kissling GE, Cunny H, Mutlu E, Waidyanatha S, Sills R, Hong HL, Ton TV, Maynor T, Recio L, Phillips SL, Devito MJ, Brix A (2018) Carcinogenic activity of pentabrominated diphenyl ether mixture (DE-71) in rats and mice. Toxicol Rep 5:615–624
4. Dunnick JK, Brix A, Cunny H, Vallant M, Shockley KR (2012) Characterization of polybrominated diphenyl ether toxicity in Wistar Han rats and use of liver microarray data for predicting disease susceptibilities. Toxicol Pathol 40(1):93–106
5. Song R, He Y, Murphy MB, Yeung LW, Yu RM, Lam MH, Lam PK, Hecker M, Giesy JP, Wu RS, Zhang W, Sheng G, Fu J (2008) Effects of fifteen PBDE metabolites, DE71, DE79 and TBBPA on steroidogenesis in the H295R cell line. Chemosphere 71(10):1888–1894
6. Sanders JM, Burka LT, Smith CS, Black W, James R, Cunningham ML (2005) Differential expression of CYP1A, 2B, and 3A genes in the F344 rat following exposure to a polybrominated diphenyl ether mixture or individual components. Toxicol Sci 88(1):127–133
7. Lin X, Liu YB, Hu H (2017) Metabolic role of fibroblast growth factor 21 in liver, adipose and nervous system tissues. Biomed Rep 6(5):495–502
8. van der Ven LT, van de Kuil T, Verhoef A, Leonards PE, Slob W, Canton RF, Germer S, Hamers T, Visser TJ, Litens S, Hakansson H, Fery Y, Schrenk D, van den Berg M, Piersma AH, Vos JG (2008) A 28-day oral dose toxicity study enhanced to detect endocrine effects of a purified technical pentabromodiphenyl ether (pentaBDE) mixture in Wistar rats. Toxicology 245(1–2):109–122

9. Natsoulis G, El Ghaoui L, Lanckriet GR, Tolley AM, Leroy F, Dunlea S, Eynon BP, Pearson CI, Tugendreich S, Jarnagin K (2005) Classification of a large microarray data set: algorithm comparison and analysis of drug signatures. Genome Res 15(5):724–736
10. Natsoulis G, Pearson CI, Gollub J, Eynon BP, Ferng J, Nair R, Idury R, Lee MD, Fielden MR, Brennan RJ, Roter AH, Jarnagin K (2008) The liver pharmacological and xenobiotic gene response repertoire. Mol Syst Biol 4(1):175
11. AbdulHameed MD, Ippolito DL, Stallings JD, Wallqvist A (2016) Mining kidney toxicogenomic data by using gene co-expression modules. BMC Genom 17(1):790
12. Ippolito DL, AbdulHameed MD, Tawa GJ, Baer CE, Permenter MG, McDyre BC, Dennis WE, Boyle MH, Hobbs CA, Streicker MA, Snowden BS, Lewis JA, Wallqvist A, Stallings JD (2016) Gene expression patterns associated with histopathology in toxic liver fibrosis. Toxicol Sci 149(1):67–88
13. AbdulHameed MD, Tawa GJ, Kumar K, Ippolito DL, Lewis JA, Stallings JD, Wallqvist A (2014) Systems level analysis and identification of pathways and networks associated with liver fibrosis. PLoS ONE 9(11):e112193
14. Tawa GJ, AbdulHameed MD, Yu X, Kumar K, Ippolito DL, Lewis JA, Stallings JD, Wallqvist A (2014) Characterization of chemically induced liver injuries using gene co-expression modules. PLoS ONE 9(9):e107230
15. Gusenleitner D, Auerbach SS, Melia T, Gomez HF, Sherr DH, Monti S (2014) Genomic models of short-term exposure accurately predict long-term chemical carcinogenicity and identify putative mechanisms of action. PLoS ONE 9(7):e102579
16. Adler M, Ramm S, Hafner M, Muhlich JL, Gottwald EM, Weber E, Jaklic A, Ajay AK, Svoboda D, Auerbach S, Kelly EJ, Himmelfarb J, Vaidya VS (2016) A quantitative approach to screen for nephrotoxic compounds in vitro. J Am Soc Nephrol 27(4):1015–1028
17. Fielden MR, Brennan R, Gollub J (2007) A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. Toxicol Sci 99(1):90–100
18. Fielden MR, Eynon BP, Natsoulis G, Jarnagin K, Banas D, Kolaja KL (2005) A gene expression signature that predicts the future onset of drug-induced renal tubular toxicity. Toxicol Pathol 33(6):675–683
19. Perez LO, Gonzalez-Jose R, Garcia PP (2016) Prediction of non-genotoxic carcinogenicity based on genetic profiles of short term exposure assays. Toxicol Res 32(4):289–300
20. Sutherland JJ, Webster YW, Willy JA, Searfoss GH, Goldstein KM, Irizarry AR, Hall DG, Stevens JL (2018) Toxicogenomic module associations with pathogenesis: a network-based approach to understanding drug toxicity. Pharmacogenomics J 18(3):377–390
21. Hu W, Sorrentino C, Denison MS, Kolaja K, Fielden MR (2007) Induction of cyp1a1 is a nonspecific biomarker of aryl hydrocarbon receptor activation: results of large scale screening of pharmaceuticals and toxicants in vivo and in vitro. Mol Pharmacol 71(6):1475–1486
22. Nault R, Kim S, Zacharewski TR (2013) Comparison of TCDD-elicited genome-wide hepatic gene expression in Sprague-Dawley rats and C57BL/6 mice. Toxicol Appl Pharmacol 267(2):184–191
23. Yu K, Gong B, Lee M, Liu Z, Xu J, Perkins R, Tong W (2014) Discovering functional modules by topic modeling RNA-Seq based toxicogenomic data. Chem Res Toxicol 27(9):1528–1536
24. AbdulHameed MD, Ippolito DL, Wallqvist A (2016) Predicting rat and human pregnane X receptor activators using bayesian classification models. Chem Res Toxicol 29(10):1729–1740
25. Tugendreich S, Pearson CI, Sagartz J, Jarnagin K, Kolaja K (2006) NSAID-induced acute phase response is due to increased intestinal permeability and characterized by early and consistent alterations in hepatic gene expression. Toxicol Pathol 34(2):168–179
26. Hu G, Agarwal P (2009) Human disease-drug network based on genomic expression profiles. PLoS ONE 4(8):e6536
27. Iskar M, Zeller G, Blattmann P, Campillos M, Kuhn M, Kaminska KH, Runz H, Gavin AC, Pepperkok R, van Noort V, Bork P (2013) Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. Mol Syst Biol 9:662

28. Mulas F, Li A, Sherr DH, Monti S (2017) Network-based analysis of transcriptional profiles from chemical perturbations experiments. BMC Bioinform 18(Suppl 5):130
29. Taguchi YH (2017) Identification of candidate drugs using tensor-decomposition-based unsupervised feature extraction in integrated analysis of gene expression between diseases and DrugMatrix datasets. Sci Rep 7(1):13733
30. Sutherland JJ, Webster YW, Willy JA, Searfoss GH, Goldstein KM, Irizarry AR, Hall DG, Stevens JL (2017) Toxicogenomic module associations with pathogenesis: a network-based approach to understanding drug toxicity. Pharmacogenomics J
31. Gong B, Wang C, Su Z, Hong H, Thierry-Mieg J, Thierry-Mieg D, Shi L, Auerbach SS, Tong W, Xu J (2014) Transcriptomic profiling of rat liver samples in a comprehensive study design by RNA-Seq. Sci Data 1:140021
32. Romer M, Backert L, Eichner J, Zell A (2014) ToxDBScan: large-scale similarity screening of toxicological databases for drug candidates. Int J Mol Sci 15(10):19037–19055
33. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44(W1):W90–W97
34. Liu L, Tsompana M, Wang Y, Wu D, Zhu L, Zhu R (2016) Connection map for compounds (CMC): a server for combinatorial drug toxicity and efficacy analysis. J Chem Inf Model 56(9):1615–1621
35. Xing L, Wu L, Liu Y, Ai N, Lu X, Fan X (2014) LTMap: a web server for assessing the potential liver toxicity by genome-wide transcriptional expression data. J Appl Toxicol 34(7):805–809
36. Hur J, Danes L, Hsieh JH, McGregor B, Krout D, Auerbach S (2018) Tox21 enricher: web-based chemical/biological functional annotation analysis tool based on Tox21 toxicity screening platform. Mol Inform 37(5):e1700129
37. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrian-Uhalte E, Davies M, Dedman N, Karlsson A, Magarinos MP, Overington JP, Papadatos G, Smit I, Leach AR (2017) The ChEMBL database in 2017. Nucleic Acids Res 45(D1):D945–D954
38. Kupershmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, Shekar M, Wang H, Park J, Cui W, Wall GD, Wisotzkey R, Alag S, Akhtari S, Ronaghi M (2010) Ontology-based meta-analysis of global collections of high-throughput public data. PLoS One 5(9)
39. Kohonen P, Benfenati E, Bower D, Ceder R, Crump M, Cross K, Grafstrom RC, Healy L, Helma C, Jeliazkova N, Jeliazkov V, Maggioni S, Miller S, Myatt G, Rautenberg M, Stacey G, Willighagen E, Wiseman J, Hardy B (2013) The toxbank data warehouse: supporting the replacement of in vivo repeated dose systemic toxicity testing. Mol Inform 32(1):47–63
40. Lea IA, Gong H, Paleja A, Rashid A, Fostel J (2017) CEBS: a comprehensive annotated database of toxicological data. Nucleic Acids Res 45(D1):D964–D971

Daniel L. Svoboda is a Data Scientist at Sciome LLC. He primarily collaborates with the Division of the National Toxicology Program, National Institute of Environmental Health Sciences. His work is focused on the application of molecular and high-dimensional data to toxicological assessment, and the development and maintenance of pertinent software applications.

Trey Saddler is a contract Data Scientist in the Office of Data Science at the National Toxicology Program. Trey mainly writes web applications and data analysis pipelines in R and Python to analyze genomic data.

Scott S. Auerbach is a Group Leader in Toxicoinformatics at the National Toxicology Program. His research focuses on application of molecular and high-dimensional data to toxicological assessment.

Chapter 9
A Pair Ranking (PRank) Method for Assessing Assay Transferability Among the Toxicogenomics Testing Systems

Zhichao Liu, Brian Delavan, Liyuan Zhu, Ruth Robert and Weida Tong

Abstract The use of animal models for risk assessment is not a reliable and satisfying paradigm. Accompanying the strategic planned shift by regulatory agencies, more and more advocating campaigns such as the 3Rs in Europe and Tox21/ToxCast in the USA were proposed to develop in silico and in vitro approaches to eliminate animal use. To effectively implement non-animal models in risk assessment, novel approaches are urgently needed for investigating the concordance between testing systems to facilitate the selection of the fit-for-purpose assay. In this chapter, we introduce a Pair Ranking (PRank) method for the quantitative evaluation of assay transferability among the different toxicogenomics (TGx) testing systems. First, we will summarize the critical issues of TGx related to its success in risk assessment. Second, we will elucidate the application of proposed PRank method for addressing key questions in TGx. Finally, we will suggest some potential use of the PRank method for advancing risk assessment.

Z. Liu (✉) · B. Delavan · L. Zhu · W. Tong
Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR 72079, USA
e-mail: zhichao.liu@fda.hhs.gov

B. Delavan
e-mail: bsdelavan@ualr.edu

L. Zhu
e-mail: Liyuan.Zhu@fda.hhs.gov

W. Tong
e-mail: Weida.Tong@fda.hhs.gov

B. Delavan
Department of Biosciences, University of Little Rock, Little Rock, AR, USA

R. Robert
ApconiX, BioHub at Alderley Park, Alderley Edge, UK
e-mail: Ruth.Roberts@apconix.com

Department of Biosciences, University of Birmingham, Birmingham, UK

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_9

Keywords IVIVE · Toxicogenomics · 3Rs · Animal models · Risk assessment

Abbreviations

3Rs	Refine, Reduce and Replace
ALT	Alanine Aminotransferase
AOPs	Adverse Outcome Pathways
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
CTD	Comparative Toxicogenomics Database
DILI	Drug-Induced Liver Injury
ECFP	Extended-Connectivity Fingerprints
EPA	United States Environmental Protection Agency
GLP	Good Laboratory Practice
hTERT	Human Telomerase Reverse Transcriptase
ICH	The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
iPSC	Induced Pluripotent Stem Cell
IVIVE	In Vitro-to-In Vivo Extrapolation
LDH	Lactate Dehydrogenase
LINCS	The Library of Integrated Network-Based Cellular Signatures
MAQC	Microarray Quality Control
OECD	Organisation for Economic Co-operation and Development
POP	Percentage of Overlapped Pathways
PRank	Pair Ranking
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
ROC	Receiver Operating Characteristic
SIDER	Side Effect Resources
TG-GATEs	Toxicogenomic Project—Genomics Assisted Toxicity Evaluation System
TGx	Toxicogenomics

9.1 Introduction

The use of animal models in risk assessment is based on the presumption that the biological response of animals mimics that of humans [1]. However, the divergence between the species causes limited extrapolation power from animal to human [2]. Suboptimal transferability between animal and human stimulated rethinking and reevaluation of innovative preclinical testing systems. Consequently, the strategic plan in risk assessment has been a shift to develop alternative approaches for testing toxicity. In Europe, the “refine, reduce and replace” (3Rs) has been advocated

and promoted by REACH legislation [3] and the 7th Amendment of the Cosmetics Directive [4]. In the USA, the Tox21 program is led by the EPA [5–7] and ToxCast was initialized by a cross-agency effort [8] to advancing regulatory science [9]. This program was proposed by the US Food and Drug Administration (FDA) to promote in vitro and in silico approaches for enhancing risk assessment and eliminating animal use.

Toxicogenomics (TGx) is a sub-discipline of pharmacology that offers a groundbreaking addition to conventional toxicology approaches [10]. TGx approaches have been widely applied in addressing different toxicological questions at the molecular level [11]. For example, Fielden et al. [12] employed a five-day rat in vivo TGx model to predict the nongenotoxic carcinogenicity, which generated better prediction performance with a mechanistic assessment of underlying mechanism. In the following studies, researchers expanded the studies for both size and treatment duration to achieve a better prediction performance with accuracies in the range 75–80% [13–16]. Furthermore, some studies utilized cell-based in vitro (HepG2 cells) TGx assay to uncover different mechanisms between genotoxic and nongenotoxic hepatocarcinogens [17] and further developed cell-based in vitro TGx prediction models [18]. Herwig et al. [18] developed human in vitro TGx models in hepatoma-derived cells and hTERT-immortalized renal proximal tubule epithelial cells to predict nongenotoxic carcinogens. Huang et al. [19] employed a serum-based TGx assay to predict drug-induced liver injury, which yielded a 92.1% accuracy with several important pathways including Toll-like receptor signaling, apoptosis and mitochondrial damage-related DILI mechanism enriched. Advances in emerging technologies such as next-generation sequencing and bioengineering including iPSC cell culture have also been integrated into TGx field and provide a promising approach for risk assessment [20, 21].

Unlike decades ago, several large public available TGx data sets such as open TG-GATEs [22, 23], DrugMatrix [24] and PredTox [25] have been generated and currently provide tremendous opportunities for formulating hypotheses to advance toxicology researches. Considering the different assay types, species, and genomics technologies in the preclinical setting, a comprehensive assessment among different TGx assay systems in the preclinical setting is urgently needed for selecting the fit-for-purpose approach.

To fill the gap, we developed a Pair Ranking (PRank) method to assess the transferability among the different TGx assays and utilized our method to address several key questions in the TGx field. In this chapter, we will first lay out the key questions for promoting in vitro TGx systems in risk assessment. Then, we will elaborate on the PRank method with a few case studies. Final, we will summarize the roadmap for further positioning the proposed PRank method toward potential regulatory applications.

9.2 Key Questions in Toxicogenomics

To better apply the TGx assay systems in testing different toxicity endpoints and deciphering the underlying mechanisms of toxicology, we provided a landscape of the internal relationship among the preclinical testing systems and human toxicity (Fig. 9.1).

9.2.1 In Vivo Animal Models

The toxicology community is now questioning the use of animal models in risk assessment. Despite their widespread utilization, using animal models has many disadvantages. For chemical toxicology studies, animal models can take an average of two years to perform for each chemical and cost millions of dollar per compound. The lack of laboratories capable of carrying out animal studies is also a significant issue. Only about 15 chemicals per year are tested using animal models. The REACH files in Finland contains over 30,000 chemicals that could be tested using animal models

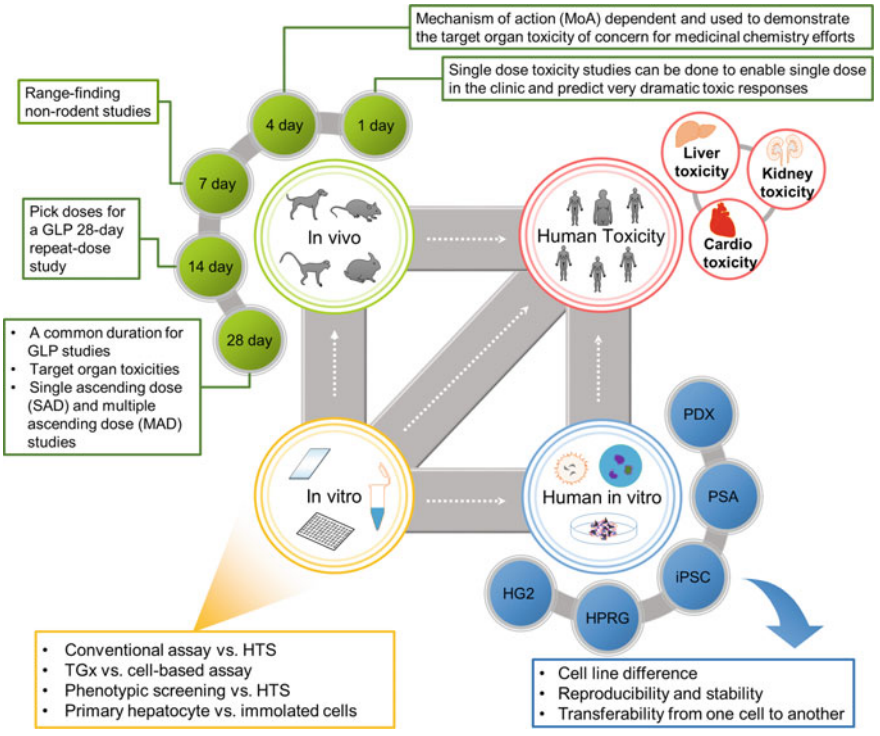

Fig. 9.1 Interior relationship among the preclinical testing systems and human toxicity

[26]. TGx methods allow for the screening of drugs, providing accurate results at a much lower cost than traditional drug development methods. TGx produces results faster than traditional drug development. However, species differences and the impact of multiple time/dosage effect still need to be optimized. Here, we summarized the aims and purpose of an in vitro experiment with different treatment duration to explain the complexity:

(1) One-day experiment: single dose toxicity studies can be done to establish the dosage in the clinic and applied to predict diverse toxic response.
(2) Four-day experiment: can demonstrate target-organ toxicity concerns for medicinal chemistry efforts under the specific mechanism of action (MoA).
(3) Seven-day experiment: mainly used to establish the dosage range for non-rodent animal studies.
(4) Fourteen-day experiment: prerequisite stage to pick the dose for a GLP twenty-eight-day repeated dose study.
(5) Twenty-eight-day experiment: a golden standard GLP study to assess the target organ toxicities and study single ascending dose (SAD) and multiple ascending doses (MAD) for seeking for first human dose in a clinical trial.

9.2.2 In Vitro Testing Systems

The in vitro assay systems derived from animal or human tissues provide an opportunity to complement long-term animal-based testing procedures, which has been marked as uncertainty and with limited transferability power [27–29]. High throughput approaches for transcriptional profiling such as in vitro gene expression assays have been widely applied and intensively explored for its potential capability in both toxicology and pharmacology fields [30–32]. The evolving nature of genomic technology provides multiple choices on TGx experiment designs for addressing different toxicological questions. Meanwhile, concerns are also raised on the reproducibility of genomic technologies due to the complex nature and aspects regarding cell types, genomics platforms, intro/across laboratory, and data analysis methodologies [33].

9.2.3 Human Toxicity

The objective of either in vitro or in vivo testing systems is to truly reflect the biological response of humans. To select the right testing assay systems for assessing the toxicity endpoint in human, we need better understanding the etiology and natural history of human toxicity. Several knowledge bases have been developed to increase our understanding of toxicity and provide a "one-stop" solution for prediction model development and new hypothesis generation. For example, NCTR scientists led the effort to develop a liver toxicity knowledge base (LTKB) to provide different drug

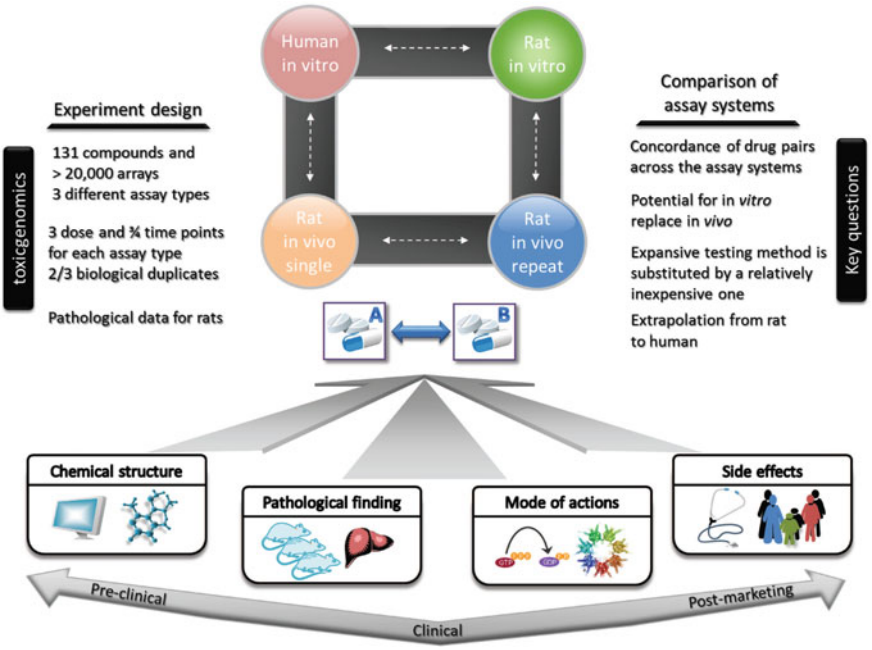

Fig. 9.2 Key questions in the toxicogenomics field

properties and host information on DILI [ref]. More importantly, a reproducible DILI classification scheme was developed for facilitating DILI prediction model development [34, 35]. Furthermore, Wen et al. [36] developed a drug-induced rhabdomyolysis atlas (DIRA) that mainly provides three folds of drug-induced rhabdomyolysis related information including a classification scheme for drugs' potential to rhabdomyolysis, post-marketing surveillance data of drug-induced rhabdomyolysis and drug property information.

From a translational toxicology perspective, it is of great importance to comprehensively explore assay transferability among the testing systems and provide useful information for preclinical toxicity screening. Our study looks to use TGx techniques computationally to answer the following five questions from a translational science point of view (Fig. 9.2):

1. Can in vitro analysis predict in vivo results?
2. Can short-term assays (as short as one day) replace long-term assays (such as twenty-eight-day assays)?
3. Is the extrapolation between assay systems endpoint dependent?
4. Is the extrapolation between assay systems adverse outcome pathways (AOPs) related?
5. Can a TGx system provide extra value to Read-Across?

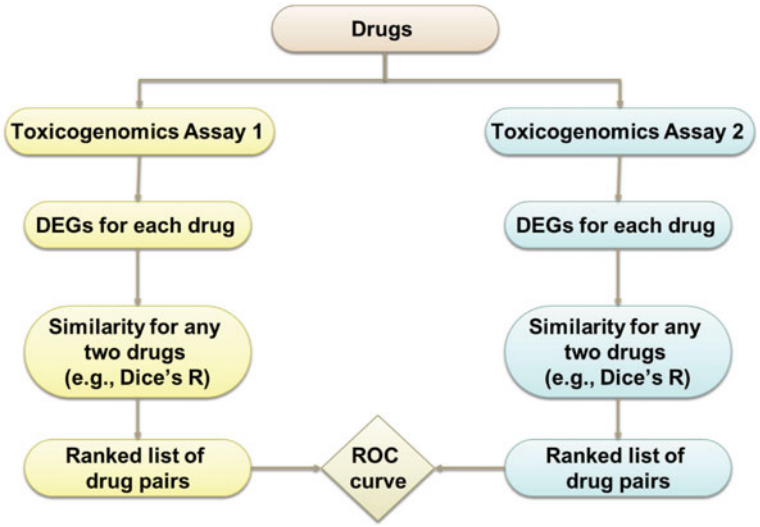

Fig. 9.3 Workflow of the Pair Ranking (PRank) method

9.3 Pair Ranking (PRank) Method

We developed a pair ranking (PRank) method to quantitatively assess assay transferability with a PRank score, which is based on the preservation of the order of similarity rankings of compound pairs between the testing systems using a receiver operating characteristic (ROC) curve analysis [37, 38]. The PRank method is a framework including the following four steps (Fig. 9.3):

(1) Gene signature generation: The gene signatures of each compound in each testing system were generated based on ranking fold change values (treatment group vs. control group). In PRank, we choose top/down number of N genes as the signature for each compound.

(2) Compound pairwise similarity calculation: In each TGx testing system, the compound pairwise similarity was calculated by using Dice's correlation coefficient with Eq. (9.1)

$$\text{Dice's coefficient} = \frac{2|S_A \cap S_B|}{|S_A| + |S_B|} \tag{9.1}$$

where $|S_A|$ and $|S_B|$ are the number of significant genes of compound A and B. $|S_A \cap S_B|$ is the number of overlapping genes between compound A and B.

(3) Cut-off value determination: In PRank, the transferability between the assay systems is directional. For example, if you want to assess the transferability from assays A and B, the binary values should be assigned based on the distribution

of ranked compound pairwise similarity list. Here, we used the 0.95 quantile value as a cut-off to transform the ranked compound pairwise similarity list into 0 and 1.

(4) ROC curve analysis: finally, the PRank score between the two testing assay platforms can be calculated based on ROC curve analysis. Consequently, the area under the curve (AOC) value was considered as the PRank score to assess the assay transferability quantitively.

This proposed PRank method aims to provide a novel framework to assess assay transferability. Therefore, detailed strategies in each step could be modified and updated. For example, alternative strategies to generate the compound gene signature such as the fold change $+p$ value criteria suggested by the MAQC consortium [39] was also fit for PRank. Furthermore, significant genes could be domain-specific. For instance, a lot of toxicogenomics annotation resources such as the Comparative Toxicogenomics Database (CTD) could be employed to further limit the significant gene list into different adverse related pathways (AOPs) or toxicity-related gene sets [40]. In the current version, Dice's correlation coefficient was used to calculate the compound pairwise similarity. Other similarity measures such as Tanimoto similarity or KL divergence based on topic similarity are also suggested for the future version. Lastly, we used ROC curve analysis to calculate the PRank score to represent assay transferability quantitatively. Other strategies for comparing two lists are also worth testing.

In the conventional approach for querying assay transferability, the compound is tested in the different assay testing systems. If the gene expression patterns of the compound are similar in assay system A and B, we consider the compound could reflect the same biology in both testing systems. However, this approach always suffers from interior assay difference and batch effect in experimental design, which fails to provide a global view on the assay transferability. In the PRank, "similar" compounds have implied similar toxicity profiles in toxicogenomic space. A highly similar pair of compounds should be observed in both testing TGx assays. In another word, if the pairwise similarity of the two compounds is consistently ranked on the top of all the compound pairs under the testing assays, these two compounds can be a highly likely similar pair. The same concept can be extrapolated to assess the transferability of any two testing systems where if two assays could produce the same ranking resolution, we consider the two assays interchangeable.

9.4 Toxicogenomics Data and Annotation Resources

9.4.1 Open TG-GATEs

To explore transferability among TGx testing systems using the proposed PRank method, a large scale of TGx data set from the Open TG-GATEs was employed [31]. TG-GATEs is an acronym for Toxicogenomic Project—Genomics Assisted

Table 9.1 Open TG-GATEs data used to investigate assay transferability

Species	Human	Sprague–dawley rat (6 weeks old)	
Experiment	In vitro	In vitro	In vivo
Tissues	Liver	Liver	Liver
#Compounds	170		
Treatment	2, 8, 24 h (2 h is omitted in some cases)	2, 8, 24 h	• 3, 6, 9, and 24 h after a single administration • 24 h after the last dose of repeated administration for 3, 7, 14 and 28 days
Dose	Low, middle, high (1:5:25, low is omitted in some cases)	Low, middle, high (1:5:25)	Low, middle, and high (mainly 1:3:10)
Clinical information	Cell viability (LDH release and DNA contents)	Cell viability (LDH release and DNA contents)	Histopathology: liver and kidney, body weight, organ weight (liver and kidney), food consumption, hematology, and blood biochemistry

Toxicity Evaluation System. The Open TG-GATEs is an open TGx database led by the Japanese Toxicogenomic Project consortium (TGP), which is accessed at http://toxico.nibio.go.jp/english/index.html. The Open TG-GATEs database contains four types of TGx assays including two in vitro assays (i.e., rat and human primary hepatocytes) and two in vivo assays (i.e., rat liver single dose and repeated dose) that have been exposed to 170 compounds at different dosages and time points.

Table 9.1 listed the information on the four different TGx assays. More details about the experimental design of each assay can be found elsewhere [22, 31]. The microarray data in each TGx assay were preprocessed by using Factor Analysis for Robust Microarray Summarization (FARMS) [41] with custom chip definition files (CDFs) from Brainarray [42]. The fold change values for each compound were calculated by compared treatment group versus the matched controls. More detail on microarray data preprocessing was described in our previous studies [37, 38]. In this chapter, we only employed data from high concentration/dosage and the longest treatment duration for carrying out the assay comparison analysis.

9.4.2 Drug-Induced Liver Injury (DILI) Annotations

To further investigate whether the extrapolation between assay systems is endpoint dependent, we used well-established drug-induced liver injury (DILI) annotation data sets. DILI is a major concern for drug developers and contributes to many drug attritions [43]. In this chapter, three different ways to classify DILI were utilized: NCTR DILI classification scheme [34], Xu's label [44], and Sakatis' label [45]. The NCTR DILI classification scheme was based on FDA approved drug labeling information, which classified the drugs into three categories including Most-DILI concern, Less-DILI concern and No-DILI concern. In this chapter, we used drugs belonging to Most-DILI concern. Xu's data set was refined by using images of the cells to assess toxicity endpoints in human liver cultures. Sakatis' data adopted in vitro biological activation to detect DILI potential in over two hundred compounds. Only the drugs with positive DILI from Xu and Sakatis' data were mapped to the Open TG-GATEs list.

Aside from different DILI annotation data sets, we also employed different hepatotoxic manifestation data for further investigation of our PRank method. Specifically, the SIDER database (http://sideeffects.embl.de/), which stands for Side Effect Resource, is an online database listing reported side effects for both over-the-counter and prescription drugs. The side effect terms in the SIDER database were standardized with preferred terms (PTs) by using the Medical Dictionary for Regulatory Activities (MedDRA) ontology. The hepatotoxic-related side effects were extracted by mapping the PT terms onto the Society of organ (SOC) level. The PTs with primary SOC *hepatobiliary* were considered as hepatotoxic-related side effects. The extracted hepatotoxic-related side effects were further classified by the domain experts and generated the following five categories: liver hepatobiliary abnormality, transaminase elevations, histologic findings, liver injury patterns, and the severity of liver injury [37].

9.4.3 Therapeutic Categories

We also examined concordance among TGx assay systems for different therapeutic categories. The WHO Anatomical Therapeutic Chemical (ATC) Classification System was used to categorize the compounds in different therapeutic classes. ATC consists of a five-level ontological structure. The second level representing the therapeutic subgroup was used in this study.

9.4.4 Gene Sets Related to Adverse Outcome Pathways (AOPs)

To investigate whether the assay transferability is AOP-specific, the gene sets related to different AOP were extracted from the Comparative Toxicogenomics Database (CTD) [40]. CTD aims to provide a comprehensive resource for better understanding the interrelationship among genes, chemicals, and diseases and advancing the public health. In this chapter, we used the gene-pathway association table in CTD, which was downloaded from http://ctdbase.org/downloads/. Only the pathways with more than 200 genes were kept for further analysis.

9.4.5 Code Availability

The source code of PRank, the processed microarray data, and the annotation data used in this chapter can be downloaded from GitHub at the link https://github.com/iguana128/Frontier-source-codes.

9.5 Case Studies

9.5.1 In Vitro to in Vivo Extrapolation (IVIVE) in TGx

Efforts have been made to develop in vitro systems tailored to address toxicologically relevant mechanisms and enhance risk assessment. Also, drug failures either in preclinical or clinical development often require pharmaceutical companies to go back into lead optimization and select new molecules without unwanted properties. In this context, ICH-M3 (http://www.ich.org/products/guidelines/safety/safety-single/article/guidance-on-nonclinical-safety-studies-for-the-conduct-of-human-clinical-trials-and-marketing-author.html) highlights, "additional nonclinical studies to provide mechanistic understanding can be useful." In this field, in vitro to in vivo extrapolation (IVIVE) has emerged as one of the powerful twenty-first-century methodologies for risk assessment and safety evaluation. IVIVE is defined as, "The qualitative or quantitative transposition of experimental results or observations made in vitro to predict phenomena in vivo, on full living organisms." A comprehensive assessment of IVIVE potential in TGx assay systems is of great importance for promoting the 3Rs principle and improving risk assessment power.

The proposed PRank method was employed to investigate the IVIVE potential based on the Open TG-GATEs data sets. Figure 9.4a illustrates the concordances among three TGx assays including rat in vitro, human in vitro, and rat in vivo repeated dosing. The highest PRank score 0.77 was obtained between two in vitro assay systems, highlighting the testing systems was dominated by the divergence among

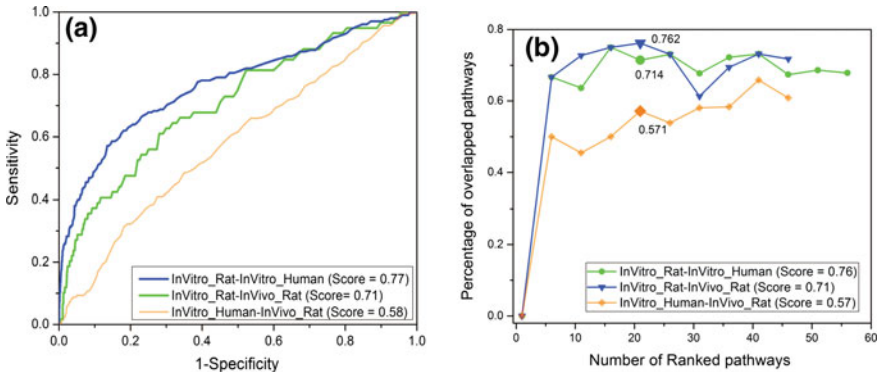

Fig. 9.4 In vitro to in vivo extrapolation (IVIVE) potential: **a** PRank method; **b** the percentage of overlapping pathways (POPs) based on enriched KEGG pathways with an adjusted *p*-value less than 0.05 using Fisher's exact test

the assay systems. Furthermore, a high IVIVE potential (i.e., PRank score = 0.70) was observed between rat in vitro and rat in vivo repeated dose, indicating non-animal approaches could detect the similar biological response derived from animal models. However, the poor transferability was obtained between human in vitro and rat in vivo repeated dose with a PRank score of 0.58, suggesting the complexity of species differences under the different testing systems. Figure 9.4b shows the concordance among three TGx assay systems in the pathway level. The percentage of overlapped pathways (POPs) between any two TGx assays was calculated. The same trend was observed in the pathway level compared to that of gene level, which further demonstrated the proposed PRank method could generate the reliable results for assay transferability assessment.

9.5.2 Short-Term Assays Show the Potential to Replace Long-Term Assays

Long-term rodent assays are still adopted in the preclinical setting for detecting toxicity such as carcinogenicity. A fierce debate was ongoing on whether animal testing could sufficiently provide hard proof for potential carcinogenic risk to humans. Consequently, community efforts have been made to develop a short-term animal study with minimal treatment time and a single dose design with both a reduced time and a lower cost of assessment regarding animal resources and workforce to replace the two-year assay. In our previous study, we conducted a comparative analysis of a predictive model for nongenotoxic carcinogenicity and suggested a short-term five-day TGx animal model has a great potential to predict the long-term endpoint [13]. In the TGx setting, the twenty-eight-day repeated dose assay is considered as a golden standard assay to establish the target organ toxicity. Here, we used the PRank method

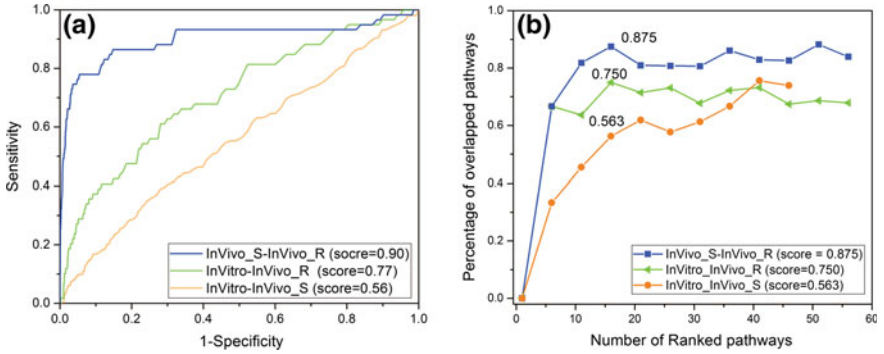

Fig. 9.5 Transferability among the three rat toxicogenomics assays: **a** PRank method. **b** The percentage of overlapping pathways (POPs) based on enriched KEGG pathways with an adjusted *p*-value less than 0.05 using Fisher's exact test

to examine whether a short in vivo single TGx assay (1 day) could have a good correlation with a twenty-eight-day in vivo repeated dose testing.

We conducted an assay comparison among the three preclinical TGx testing assay systems (i.e., Rat in vitro—InVitro, Rat in vivo single dose—InVivo_S, and Rat in vivo repeated dose—InVivo_R) using PRank. A very high PRank score, 0.90, was found for the one-day in vivo single dose and the twenty-eight-day in vivo repeated dose (Fig. 9.5a). The high PRank score between these two assay systems indicates a strong probability of using the shorter, less expensive 24-h single dose in place of the longer, more expensive twenty-eight-day repeated dose without loss of any predictive power. However, we did not see the good concordance (i.e., PRank score = 0.56) between rat in vitro and rat in vivo single dose. Similarly, the POP analysis was implemented to verify further the results derived from PRank (Fig. 9.5b). The same pattern was observed at pathway level as well with a decreasing order of POP values 0.875, 0.750, and 0.563 for InVivo_S- InVivo_R, InVitro- InVivo_R, and InVitro-InVivo-S, respectively.

9.5.3 TGx Assay Transferability Is Endpoint Dependent

In the past decade, DILI prediction models have been developed considerably by using various machine learning technologies with different complexity of data profiles. However, it seems that the prediction performance is still suboptimal [46]. One key question for preclinical DILI model development is how to choose a "fit-for-purpose" in vitro assay for assessing different DILI endpoints.

To further investigate whether TGx assay transferability is endpoint dependent, we carried out PRank analysis by limiting the compounds that belong to different hepatotoxic-related endpoints and different therapeutic categories. To measure any

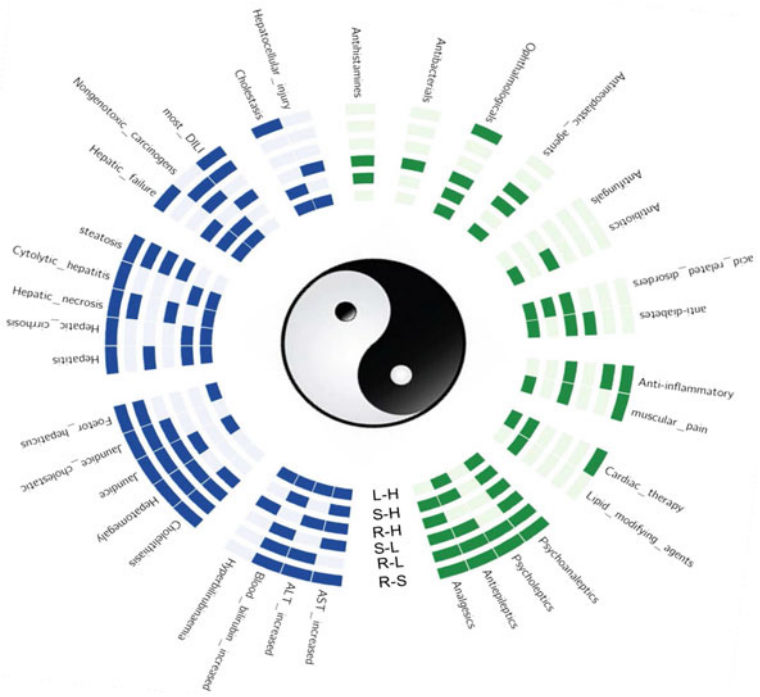

Fig. 9.6 Circos plot of PRank scores across different hepatotoxic-related endpoints (blue color) and different therapeutic categories based on the Anatomical Therapeutic Chemical Classification (ATC) System: S—in vivo single dose; L—in vivo repeated dose; R—rat in vitro; and H—human in vitro

improvement of assay transferability in different hepatotoxic-related endpoints, we calculated the PRank scores for four groups of compounds (Table 9.2 and blue color in Fig. 9.6). Two of these groups had an endpoint of "most DILI concern" and "hepatic failure." The two other groups used the endpoint of general DILI. Overall, the PRank score for *InVitro_Rat-InVivo_Rat* increased by 7% for all four groups of DILI endpoints studied. When the PRank scores for InVitro_Human-InVivo_Rat were computed using the four DILI groups, the PRank scores increased in three of the four DILI endpoints. We wanted to make sure these increases were not the result of chance. To safeguard the results from chance, we carried out a permutation test by selecting an equal number of compounds from the universe of compounds from each of the four DILI endpoints. Each selected compound was analyzed by the PRank method. The analysis was conducted 100,000 times, removing any potential bias when the compounds were selected. Similarity, the assay transferability also varied in different therapeutic categories (green color in Fig. 9.6). For example, compounds in *psychanaleptics* have an excellent transferability among the different TGx assay systems, indicating the in vitro assay could be sufficient for testing the compounds regarding different toxicities.

Table 9.2 Assay transferability for different hepatotoxic endpoints

Categories	Number of compounds	InVitro_Rat-InVivo	InVitro_Human-InVivo	InVitro_Rat-InVitro_Human
All the compounds	120	0.71	0.58	0.77
Drug-induced liver injury (DILI)				
Most-DILI concern	46	0.76	0.62	0.74
Xu's label	47	0.82	0.53	0.73
Sakatis's label	51	0.77	0.63	0.73
Hepatic failure	24	0.77	0.72	0.82
Other hepatotoxic endpoints				
Biochemical parameters				
AST increased	13	0.69	0.62	0.78
ALT increased	9	0.81	0.52	0.88
Hepatic enzyme increased	8	0.71	0.68	0.80
Blood bilirubin increased	6	0.81	–	–
Liver injury patterns				
Cholestasis	7	0.95	0.71	0.70
Hepatocellular injury	16	0.75	0.57	0.89
Hepatobiliary abnormality				
Cholelithiasis	6	0.86	–	–
Foetor hepaticus	7	0.81	0.62	0.96
Hepatomegaly	5	0.89	0.59	0.50
Jaundice	49	0.69	0.63	0.76
Jaundice cholestatic	21	0.40	0.61	0.74
Histologic findings				
Hyperbilirubinaemia	7	0.84	0.68	0.78
Hepatitis	43	0.70	0.61	0.73
Hepatic cirrhosis	7	0.80	0.44	0.89
Liver disorder	8	0.82	0.68	0.96
Hepatic function abnormal	33	0.77	0.62	0.71
Steatosis	13	0.67	0.70	0.85
Hepatic necrosis	13	0.64	0.63	0.68
Cytolytic hepatitis	10	–	0.61	0.87
Total (%) of IVIVE score increased		78.3	73.9	43.5

9.5.4 *Concordance Among TGx Assays is Adverse-Outcome-Pathway-(AOP)-Specific*

The concept of AOP is well established in risk assessment, which aims to provide a structured representation of biological events and their causing adverse effects [47]. An AOP usually includes a molecular initiating event (MIE), key events (KE), and an adverse outcome (AO). The development of an AOP is very time-consuming and domain-expert-knowledge-dependent. Some computational methodologies have been developed to enrich conceptual AOP description to eliminate manual curation [48, 49]. The AOP concept has been widely used to assess different toxicity endpoints and uncover their underlying mechanism [50, 51]. To understand the assay transferability in different AOPs is of great importance to further implement the AOP concept for the preclinical screening purpose.

We further studied the concordance when limiting the analysis to selecting genes along specific AOPs. The gene-pathway associations from the Comparative Toxicogenomic Database (CTD) were used. The investigation eventually used 106 gene sets from computational constructed AOPs, from which at least 200 genes for each AOP existed. When computed, the concordance between the three assay systems was similar with the findings at the whole gene/pathway level: the PRank score for the InVivo_S-InVivo_R was the highest, followed by InVitro-InVivo_R, then last was InVitro-InVivo_S (Fig. 9.7a). Furthermore, some AOPs such as fatty-acid-related pathways were with high PRank scores among all the assay comparisons (Fig. 9.7b).

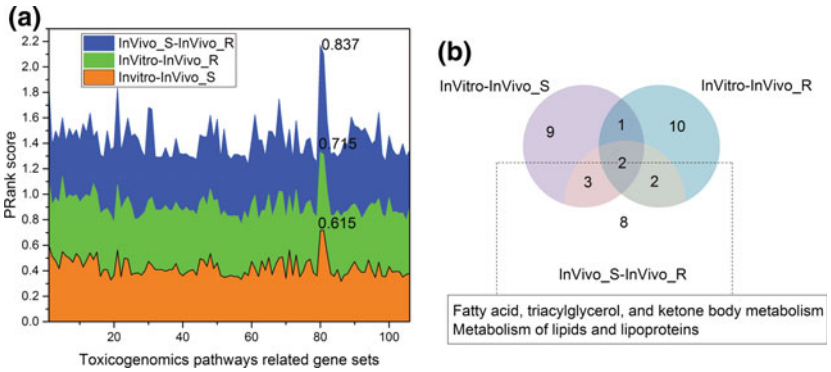

Fig. 9.7 Assay transferability among the three rat toxicogenomics assays for different adverse outcome pathways (AOPs): **a** stacked plots of PRank scores for different AOPs in the three TGx assays; **b** a venn diagram of the top 15 AOPs ranked by the PRank score in each assay system

9.5.5 *Toward Biological Data-Based Read-Across*

Read-across is a data gap filling strategy to assess the toxicity of untested compounds based on the analog to similar chemicals, for which toxicity data are available. Read-across approaches have been widely adopted by regulatory programs such as OECD and REACH [52]. Besides structure and physicochemical properties, the biological data-based similarity may enhance this process [53].

We compared compound similarity between chemical space and three preclinical TGx assays (Fig. 9.8). We observed low Pearson's correlation coefficients of 0.30, 0.20, and 0.21 for chemical space versus InVitro, InVivo_S, and InVivo_R, respectively. The divergence between chemical space and toxicogenomic space indicated that toxicogenomics data might provide extra value to improve chemical-based Read-across.

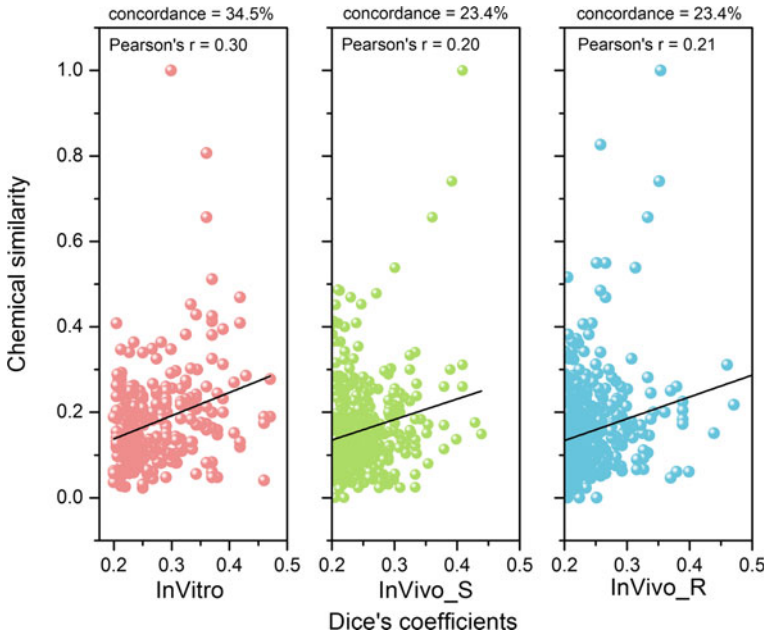

Fig. 9.8 Concordance between chemical space and TGx space: the compound pairwise chemical similarity was calculated based on ECFP-4 fingerprints by using Pipeline Pilot. Pearson's correlation coefficients were calculated between chemical space and three TGx assays, respectively

9.6 Closing Remarks

The critical assumption in using animal models was that the findings of the animal model would correlate to how the compound would behave in humans. Recent researchers have shown this correlation to be poor, leading researchers to determine if in vitro cell-based assays in combination with in silico approaches could enhance 3Rs principles. These researchers generated many novel techniques, and non-animal testing means to tackle the issue. Toxicogenomics (TGx) is one of these techniques that shows excellent promise to forecast toxicity in drug compounds while meeting the 3Rs goal. Toward studying how TGx could be used in replacing animal models, we developed a computational method called PRank, to address the transferability among different TGx testing assays and promote the in vitro TGx.

In our study, we used data culled from the TG-GATEs database. TG-GATEs is a large database for toxicogenomics, but there are many compounds not in the database. Other drug transcriptional databases with more compounds were recently generated and could be utilized to further probe PRank's potential for assay transferability. The most promising of these databases is the LINCS database [54]. The LINCS database, which is publicly available, expands significantly upon the Connectively Map (CMap). LINCS consists of more than 20 k compounds, and transcriptomic profiles were generated across more than 400 different cell lines. The LINCS database shows how genes, drugs, and diseases are associated with common gene-expression signatures [55]. Rather than use the entire human genome, the LINCS database uses 1000 genes as "landmark" genes as representative genes and uses these landmark genes to extrapolate to the whole human genome. The LINCS database can be accessed at https://clue.io. With the LINCS data set, we can further apply our PRank method to address another question such as repurposing the transcriptomic profiles from immortalized cell lines for toxicity assessment. Furthermore, the transferability between the assays in some novel cell culture such as iPSC and traditional cell cultures could also be assessed.

The use of the PRank system is not confined to the areas of toxicogenomics. These studies show that the PRank computational methodology can be successfully applied to other types of data sets. For example, the PRank can be used to high throughput screening assays from the Tox21 project or ToxCast. The National Institute of Environmental Sciences (NIEHS), part of the National Institute of Health (NIH), has a multitude of open data sets, such as the Environmental Genome Project, that PRank can be utilized to explore the strength of relationships.

As animal models begin to fall out of favor among compound safety studies, new systems must be ready to assess the potential toxicity of new compounds in humans. Utilizing already existing and open toxicogenomic databases present an opportunity to develop novel in vitro and in silico strategies to assess this potential toxicity. The PRank computational tool is a promising approach to bridge the gap between the available data and to gain insight into how new compounds may present their toxicity to humans, and thus answer some of the most significant questions in the toxicology field.

Disclaimer This article reflects the views of the authors and should not be construed to represent FDA's views or policies.

References

1. Justice MJ, Dhillon P (2016) Using the mouse to model human disease: increasing validity and reproducibility. Dis Models Mech 9(2):101–103
2. Perrin S (2014) Make mouse studies work. Nature 507(7493):423–425
3. Williams ES et al (2009) The European Union's REACH regulation: a review of its history and requirements. Crit Rev Toxicol 39(7):553–575
4. Schumann R (2002) The seventh amendment to the cosmetics directive: what does DG enterprise want from ECVAM? ATLA Altern Lab Anim 30:213–214
5. Collins FS et al (2008) Transforming environmental health protection. Science 319(5865):906–907
6. Kavlock RJ et al (2009) Toxicity testing in the 21st century: implications for human health risk assessment. Risk Anal 29(4):485–487
7. Tice Raymond R et al (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121(7):756–765
8. Dix DJ et al (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5–12
9. Hamburg MA (2011) Advancing regulatory science. Science 331(6020):987–987
10. Chen M et al (2012) A decade of toxicogenomic research and its contribution to toxicological science. Toxicol Sci 130(2):217–228
11. Qin C et al (2016) Toxicogenomics in drug development: a match made in heaven? Expert Opinion Drug Metab Toxicol 12(8):847–849
12. Fielden MR et al (2007) A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. Toxicol Sci 99(1):90–100
13. Liu Z et al (2011) Comparative analysis of predictive models for nongenotoxic hepatocarcinogenicity using both toxicogenomics and quantitative structure-activity relationships. Chem Res Toxicol 24(7):1062–1070
14. Ellinger-Ziegelbauer H et al (2008) Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. Mutat Res/Fundam Mol Mech Mutagen 637(1–2):23–39
15. Uehara T et al (2008) A toxicogenomics approach for early assessment of potential non-genotoxic hepatocarcinogenicity of chemicals in rats. Toxicology 250(1):15–26
16. Gusenleitner D et al (2014) Genomic models of short-term exposure accurately predict long-term chemical carcinogenicity and identify putative mechanisms of action. PLoS ONE 9(7):e102579
17. Lee WJ et al (2014) Investigating the different mechanisms of genotoxic and non-genotoxic carcinogens by a gene set analysis. PLoS ONE 9(1):e86700
18. Herwig R et al (2016) Inter-laboratory study of human in vitro toxicogenomics-based tests as alternative methods for evaluating chemical carcinogenicity: a bioinformatics perspective. Arch Toxicol 90(9):2215–2229
19. Huang J et al (2010) Genomic indicators in the blood predict drug-induced liver injury. Pharmacogenomics J 10(4):267–277
20. Liu Z et al (2016) Mechanistically linked serum miRNAs distinguish between drug induced and fatty liver disease of different grades. Sci Rep 6:23709
21. Ruden DM et al (2017) Frontiers in toxicogenomics in the twenty-first century—the grand challenge: to understand how the genome and epigenome interact with the toxic environment at the single-cell, whole-organism, and multi-generational level. Front Genet 8:173

22. Uehara T et al (2010) The Japanese toxicogenomics project: application of toxicogenomics. Mol Nutr Food Res 54(2):218–227
23. Igarashi Y et al (2015) Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res 43:D921–D927 (Database issue)
24. Ganter B et al (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. J Biotechnol 119(3):219–244
25. Suter L et al (2011) EU framework 6 project: predictive toxicology (PredTox)-overview and outcome. Toxicol Appl Pharmacol 252(2):73–84
26. Rowan AN (2015) Ending the use of animals in toxicity testing and risk evaluation. Camb Q Healthc Ethics 24(4):448–458
27. Kirkland D et al (2005) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens—I. Sensitivity, specificity and relative predictivity. Mutat Res Genet Toxicol Environ Mutagen 584(1–2):1–256
28. Nel A et al (2013) Nanomaterial toxicity testing in the 21st century: use of a predictive toxicological approach and high-throughput screening. Acc Chem Res 46(3):607–621
29. Sonneveld E et al (2006) Comparison of in vitro and in vivo screening models for androgenic and estrogenic activities. Toxicol Sci 89(1):173–187
30. Lamb J et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795):1929–1935
31. Igarashi Y et al (2015) Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res 43(D1):D921–D927
32. Ganter B et al (2006) Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix® database. Pharmacogenomics 7(7):1025–1044
33. Sugden LA et al (2013) Assessing the validity and reproducibility of genome-scale predictions. Bioinformatics 29(22):2844–2851
34. Chen M et al (2011) FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today 16(15):697–703
35. Chen M et al (2016) DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. Drug Discov Today 21(4):648–653
36. Wen Z et al (2018) Drug-induced rhabdomyolysis atlas (DIRA) for idiosyncratic adverse drug reaction management. Drug Discov Today 24(1):9–15
37. Liu ZC et al (2017) In vitro to in vivo extrapolation for drug-induced liver injury using a pair ranking method. ALTEX-Altern Anim Experimentation 34(3):399–408
38. Liu Z et al (2018) Transcriptional responses reveal similarities between preclinical rat liver testing systems. Front Genet 9:74
39. Consortium M (2006) The microarray quality control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 24(9):1151–1161
40. Davis AP et al (2017) The comparative toxicogenomics database: update 2017. Nucleic Acids Res 45:D972–D978 (Database issue)
41. Hochreiter S et al (2006) A new summarization method for affymetrix probe level data. Bioinformatics 22(8):943–949
42. Dai M et al (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33(20):e175–e175
43. Lee WM (2003) Drug-induced hepatotoxicity. N Engl J Med 349(5):474–485
44. Xu JJ et al (2008) Cellular imaging predictions of clinical drug-induced liver injury. Toxicol Sci 105(1):97–105
45. Sakatis MZ et al (2012) Preclinical strategy to reduce clinical hepatotoxicity using in vitro bioactivation data for >200 compounds. Chem Res Toxicol 25(10):2067–2082
46. Chen M et al (2014) Toward predictive models for drug-induced liver injury in humans: are we there yet? Biomark Med 8(2):201–213

47. Ankley GT et al (2009) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741
48. Nymark P et al (2018) A data fusion pipeline for generating and enriching adverse outcome pathway descriptions. Toxicol Sci 162(1):264–275
49. Bell SM et al (2016) Integrating publicly available data to generate computationally predicted adverse outcome pathways for fatty liver. Toxicol Sci 150(2):510–520
50. Bal-Price A, Meek ME (2017) Adverse outcome pathways: application to enhance mechanistic understanding of neurotoxicity. Pharmacol Ther 179:84–95
51. Perkins EJ et al (2015) Adverse outcome pathways for regulatory applications: examination of four case studies with different degrees of completeness and scientific confidence. Toxicol Sci 148(1):14–25
52. Patlewicz G et al (2014) Food for thought: read-across approaches—misconceptions, promises and challenges ahead. ALTEX-Altern Anim Experimentation 31(4):387–396
53. Zhu H et al (2016) t(4) report: supporting read-across using biological data. Altex 33(2):167–182
54. Subramanian A et al (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 171(6):1437–1452.e1417
55. Koleti A et al (2018) Data portal for the library of integrated network-based cellular signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. Nucleic Acids Res 46(D1):D558–D566

Zhichao Liu got his Ph.D. from Nankai University in China. He is currently a principal investigator at NCTR, FDA. Dr. Liu's background is across Chemistry, Biology and Computer Science. In past 15 years, he led and took part in many cutting-edge projects from both industry and academia. Specifically, he developed innovative approaches for in silico drug repositioning framework to seek for novel opportunities for rare disease treatment development. The goals aim at providing the standard in silico pipeline for drug repositioning and early drug safety detection by retrieving, integrating, and organizing the information from chemical, biological, and clinical spaces, which help industry to seek for the optimal route to accelerate the drug development efficacy from advanced regulatory sciences perspective. His interests also widely cover (1) development of flexible and integrative risk prediction systems for drug safety evaluation; (2) development and application of standard pipeline for genomic data (e.g., microarray/NGS) analysis for drug safety and efficacy questions. Dr. Liu has published more than 60 papers in the prestigious journals.

Brian Delavan is a Ph.D. candidate in College of Public Health at University of Arkansas for Medical Sciences. His research has focused on developing computational approaches for drug repositioning of rare diseases and bio-imaging analysis. Brian received his Master of Public Health (MPH) at University of Arkansas for Medical Sciences in 2016.

Liyuan Zhu graduated from Tsinghua University with a bachelor's degree in Biology. After that, he studied Molecular Virology and Microbiology at Baylor College of Medicine, where he discovered his interests in bioinformatics. He then got a master's degree at Columbia University in Biomedical Informatics. He now works at NCTR, FDA on toxicogenomics research.

Ruth Robert is chair of drug discovery at Birmingham University, UK and Cofounder of ApconiX, an integrated toxicology and ion channel company that brings together a team of world-renowned nonclinical safety experts with over 400 years of drug discovery and development experience. Previously, Ruth was Global Head of Regulatory Safety at AstraZeneca and Director of Toxicology for Aventis. Ruth is an established science professional bringing rigorous expert thinking to toxicology, drug discovery and drug development. With >140 publications in peer-reviewed

journals, she is focused on reducing attrition attributable to safety and toxicity. A former president of the British Toxicology Society, former president of EUROTOX, former secretary to SOT and president elect of the Academy of Toxicological Sciences, Ruth was the recipient of the SOT Achievement award in 2002, the EUROTOX Bo Holmstedt award in 2009 and the SOT Founders award in 2018 for outstanding leadership in toxicology.

Weida Tong is Director of Division of Bioinformatics and Biostatistics at FDA's National Center for Toxicological Research (NCTR/FDA). He has served as a science advisory board member for several large projects involving multiple institutes in Europe and the USA. He also holds several adjunct positions at universities in the USA and China. His division at FDA works to develop bioinformatic methodologies and standards to support FDA research and regulation and to advance regulatory science and personalized medicine. The most visible projects from his group are (1) leading the Microarray Quality Control (MAQC) consortium to develop standard analysis protocols and quality control metrics for emerging technologies to support regulatory science and precision medicine; (2) development of liver toxicity knowledge base (LTKB) for drug safety; (3) in silico drug repositioning for the enhanced treatment of rare diseases; and (4) development of the FDA bioinformatics system, ArrayTrackTM suite, to support FDA review and research on pharmacogenomics. In addition, his group also specializes in molecular modeling and QSARs with specific interest in estrogen, androgen, and endocrine disruptor. Dr. Tong has published more than 230 papers and book chapters.

Chapter 10
Applications of Molecular Dynamics Simulations in Computational Toxicology

Sugunadevi Sakkiah, Rebecca Kusko, Weida Tong and Huixiao Hong

Abstract Computational toxicology is a discipline seeking to computationally model and predict toxicity of chemicals including drugs, food additives, and other environmental chemicals. Risk assessment of chemicals using current in vitro or in vivo experimental methods is at best time-consuming and expensive. Computational toxicology seeks to accelerate this process and decrease the cost by predicting the risk of chemicals to humans and animals. Molecular dynamics (MD) simulation, an emerging computational toxicology technique, characterizes the interactions of chemicals with biomolecules such as proteins and nucleic acids. This chapter will give a brief review both of available software tools for MD simulations and also how to apply these software tools to computational toxicology challenges. We also summarize key protocols to run MD simulations.

Keywords MD simulations · Computational toxicology · In silico toxicology · AMBER · GROMACS · NAMD · Desmond

S. Sakkiah · W. Tong · H. Hong (✉)
National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA
e-mail: Huixiao.hong@fda.hhs.gov

S. Sakkiah
e-mail: Suguna.Sakkiah@fda.hhs.gov

W. Tong
e-mail: Weida.Tong@fda.hhs.gov

R. Kusko
Immuneering Corporation, Cambridge, MA 02142, USA
e-mail: bkusko@immuneering.com

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_10

Abbreviations

AFP	α-fetoprotein
AMBER	Assisted Model Building with Energy Refinement
AR	Androgen receptor
CHARMM	Chemistry at HARvard Macromolecular Mechanics
DPPC	Dipalmitoylphosphatidylcholine
EDCs	Endocrine disrupting chemicals
GROMACS	GROningen MAchine for Chemical Simulations
IFD	Induced fit docking
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator
MD	Molecular dynamics
NAMD	NAnoscale Molecular Dynamics
NPT	Constant number (N), pressure (P), and temperature (T)
NVE	Constant number (N), volume (V), and energy (E)
NVT	Constant number (N), volume (V), and temperature (T)
nAChRs	Neuronal acetylcholine receptor
PDB	Protein Data Bank
Pmemd	Particle Mesh Ewald Molecular Dynamics
QSAR	Quantitative structure–activity relationship
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
SPC	Simple point-charge water model
VMD	Visual molecular dynamics
WT	Wild type

10.1 Introduction

Toxicans are defined as any chemical substance or a mixture of various chemical substances which completely or partially damages an organism. The toxicity of a compound or substance is dose-dependent. A toxic chemical entering the human body is absorbed and interacts with other chemical substances already present in the body, leading to harm. A driving imperative in the toxicology field is twofold: (1) Determine the toxic dosage level of compounds, and (2) identify their effects on the environment and living organisms such as animals, plants, and humans. Both experimental toxicology and computational toxicology predict the toxicity and dosage of chemicals. Experimental toxicology leverages in vivo and in vitro methods. In vivo methods are slow to yield compound toxicity predictions (http://alttox.org/mapp/toxicity-testing-overview Accessed March 19, 2018). The utility of in vitro methods is constrained by time and cost. Computational toxicology, also known as "in silico toxicology," generates predictive models via computation to understand the interactions and adverse health effects of various chemicals present in the air, food, water,

etc. Predictive models can be developed using different computational methods such as pharmacophore modeling [1–4], molecular docking [5–9], and machine learning methods [10–22]. These computational methods reduce animal model use, cost, and time while improving safety evaluation and risk assessment of chemicals. The emerging field of computational toxicology predicts and examines toxicity of chemicals [23, 24]. The following tools are mainstays of computational toxicology prediction:

(1) Databases to store chemical data including chemical properties and toxicity.
(2) Software tools to generate molecular descriptors for chemicals [25].
(3) Programs to run molecular dynamics (MD) simulations.
(4) Algorithms for molecular modeling.
(5) Statistical packages to generate two-dimensional (2D) and three-dimensional (3D) quantitative structure–activity relationship (QSAR) models [26–28].
(6) Web servers or standalone applications to predict toxicity of chemicals using pre-built models.
(7) Software tools to visualize prediction models.

Among various algorithms and tools used in computational toxicology, MD simulations are increasing in prevalence. MD simulation is a well-known technique in other fields including drug design, descriptor generation, structural biology, protein analysis, identification of hot spot residues, and more. In the field of toxicology, MD simulation can link structural biology with chemical toxicity information. MD simulation informs physical movement of the atoms or molecules in a molecular system and the structural changes of the protein in a time-dependent manner. This computational technique also illuminates conformational changes and protein fluctuations induced by binding of various compounds or chemicals. Different algorithms exist for MD simulations, addressing various aspects of computational toxicology. This chapter reviews MD simulations as applied to computational toxicology.

10.2 History of MD Simulations

MD is a 100-year-old technique, but only gained traction with the scientific community during the twentieth century [29]. Table 10.1 contains a brief history of MD simulation. In the mid-50s, Fermi, Pasta, Ulam, and Tsingou successfully developed the Monte Carlo simulation method [30]. MD simulations build on these statistical methods. In 1957, Alder and Wainwright studied the interaction of a hard sphere using MD simulations. The results revealed many key learnings on simple liquid behavior [31, 32]. The next milestone was achieved by Rahman in 1964, when the first realistic liquid argon simulation occurred [33]. Next, in the late 70s, the technique of MD simulations was further improved by simulating several hundreds of atoms up to biological systems [34, 35], i.e., immersing the whole protein in solution, embedded the protein in a lipid layer, or macromolecular complexes [36, 37]. Rahman and Stillinger created the first realistic simulation of liquid water in 1974 [38]. In 1977, McCammon et al. simulated the first bovine pancreatic trypsin protein inhibitor [34].

Table 10.1 History of MD simulation

Year	Authors	Techniques
1929	P. M. Morse and J. E. Lennard-Jones	Model vibrational excitations: atomic potentials
1937		London dispersion forces due to polarization
1946		Molecular Mechanics: use of Newton's equations and force fields for the characterization of molecular conformations
1953	Metropolis Monte Carlo (MC) by Metropolis, Rosenbluth, Rosenbluth, Teller & Teller	Simulation of a dense liquid of 2D spheres
1955	Fermi, Pasta, and Ulam	Simulation of anharmonic 1D crystal [30]
1957	Alder and Wainwright	MD (MD) simulation of hard spheres [32]
1960	Gibson et al.	Simulation of damaged Cu crystal [76]
1964	Rahman	MD simulation of liquid Ar [33]
1969	Barker and Watts	Monte Carlo simulation of water
1971	Rahman and Stillinger	MD simulation of water
1972	I. R. McDonald	NPT simulation using Monte Carlo
1976	Woodcock, Angell, Soules	First simulations of SiO2 (silica) using a BMH potential. Achieves tetrahedral coordination
1980	H. C. Andersen	MD method for NPH, NVT, NPT ensembles
1980	M. Parrinello and A. Rahman	Parrinello–Rahman method for study of crystal structure transformation with corrections from S. Yashonath
1986	R. Car and M. Parrinello	Ab initio (includes electronic degrees of freedom)
1990s	Stillinger-Weber, Vashista, Finney, Ciccotti	Improvement of interaction potentials
1992		Transfer of the CECAM (Centre Europeen de Calcul Atomique et Moleculaire) to Lyon (Fr). Promotion and tutorial of advanced computational methods in material sciences
1998	Duan and Kollman	1 ms MD simulations of the folding of the villin headpiece in explicit solvent
1999	Sugita and Okamoto	Replica exchange MD
2000s		Improvement and massive diffusion of CPMD techniques among the community need of massively parallel computing (MPI, OpenMPI)

In the 1980s, MD simulation algorithmic improvements allowed for simulation and calculation of the free energy of macromolecules and protein complexes [39–41]. In the 1990s, high-temperature MD simulations successfully characterized the dynamic changes of macromolecules and captured the ensemble of structures based on the time scale [42]. Thanks to improvement in algorithms and computational power, MD simulations were additionally employed to study intermolecular interactions. Among various theoretical techniques, MD simulation is one of the widely used techniques across various research fields and is thus highly cited [43]. Presently, the number and scale of simulation techniques have expanded greatly. Using modern tools, a researcher can run a microsecond-scale simulation of solvated protein and protein complexes such as protein-nucleic acid, protein–protein, protein–small molecules, and protein or protein complexes immersed in a lipid layer. These simulations allow the field to answer questions around binding modes and thermodynamics of a small molecule, as well as folding and structural changes of a protein.

10.3 Types of MD Simulations

10.3.1 All-Atom Simulations

All-atom stimulations are run with simulation packages including NAMD, GROMACS, and AMBER. These packages determine molecular structure, conformational analysis, and the dynamic property of proteins, protein complexes, lipid layers, and polymers. The main drawback or limitation of all-atom simulations is the time and computational cost. All-atom MD simulations explicitly represent each atom in solutes and solvents, which drives up the computational cost. To overcome this limitation, coarse-grained simulation methods were developed.

10.3.2 Coarse-Grained Simulations

A coarse-grained simulation models large-scale protein or protein complex models and enables longer time-scale simulations by increasing the order of magnitude compared with all-atom simulations. Therefore, coarse-grained simulation is an alternative and effective method when compared to all-atom stimulation. Coarse-grained simulation can be run using a common MD simulation package by applying a specific coarse-grained force field. In this kind of simulation, a small group of residues or atoms in the simulated system are treated as a single particle rather than individual atoms. By leveraging this speed-up of fewer degrees of freedom, coarse-grained simulations are faster than all-atom simulations. The running time of an MD simulation depends on the frequencies of motion including bond stretching, side chain and loop motion, and angle bending. Water molecules are not explicitly simulated here.

Due to the above mentioned reasons, coarse-grained simulation is best employed for long time-scale simulations. The drawbacks of this approach include: (i) loss of atomic resolution, (ii) less accurate energetic estimation compared with atomistic approaches, (iii) limited availability of force fields, and (iv) need for additional stimulations to obtain atomistic trajectory details.

10.3.3 Classical MD

Nowadays, classical MD simulations are used to investigate many properties of molecular systems. Classical MD simulations calculate the movement of particles from an initial input structure throughout the simulation. Interactions of each particle in the simulation system with all other particles are calculated by the total force acting on a given particle. The new position of each atom after a specific time interval is determined by acceleration, previous position, and velocity. Classical MD simulations are best applied to understand protein structure and function or to propose hypotheses based on experimental data. The electronic distribution from classical MD is relatively like the coarse-grained MD method. If the interaction site has assigned fixed partial charges and an approximate model for polarization effects, the motion of electrons is not dominated by a time scale. The main limitation of classical MD simulation is the system size and time scale.

10.4 MD Simulation Software

A panoply of MD simulation packages has been developed over the years. The most popular packages include GROningen MAchine for Chemical Simulations (GROMACS) [44], Assisted Model Building with Energy Refinement (AMBER) [45], NAnoscale Molecular Dynamics (NAMD) [46], TINKER [47], CHARMM [48], LAMMPS [49], DL_POLY [50], MOLDY [51], and Desmond [52]. Some MD simulation packages have their own force fields, while others only provide MD simulation algorithms and require suitable force fields as input. The GROMACS, CHARMM, NAMD, GROMACS, and AMBER packages predominate biomolecular simulations. AMBER and CHARMM have their own force fields and provide various scripts to run and analyze simulations [53, 54]. GROMACS and NAMD show remarkable efficiency for large-scale biomolecular simulations. GROMACS [55], AMBER [54], and CHARMM [53] have coarse-grained force fields. NAMD supports only standard force fields such as AMBER and CHARMM. Desmond is a newcomer software package from D.E. Shaw Research which supports force fields including AMBER, CHARMM, and POLS-AA [56]. As examples, Figs. 10.1, 10.2, 10.3 and 10.4 give an MD simulations protocol overview for GROMACS, AMBER, Desmond, and NAMD.

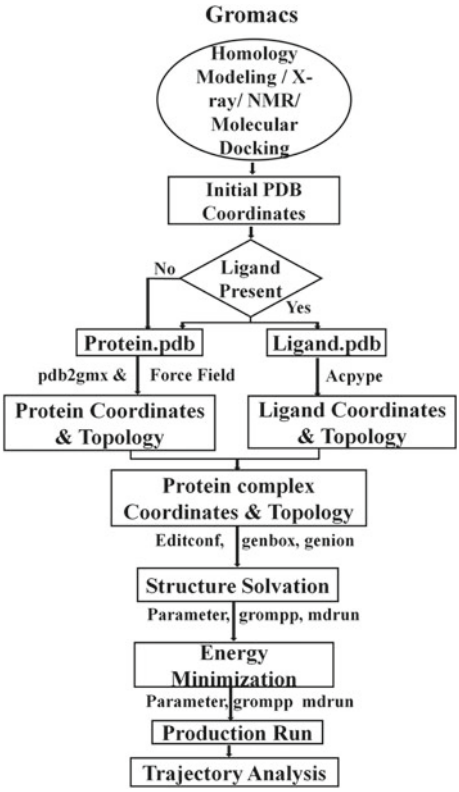

Fig. 10.1 Steps involved in running the molecular dynamics simulations using GROMACS

10.4.1 GROMACS

GROMACS is a free MD simulation package for modeling biomolecules such as DNA, RNA, and lipids. After initial development at the Department of Biophysical and Chemistry, University of Groningen, Netherlands, GROMACS, now is maintained by various universities and research centers worldwide. GROMACS runs on graphics processing units (GPUs) and central processing units (CPUs). Steps to run GROMACS are the following:

1. Protein preparation: Coordinates of a protein 3D structure can be either downloaded from the Protein Data Bank (PDB) (www.rcsb.org) or generated from modeling, such as homology modeling. Visualization tools including VMD, Chimera, and PyMol can strip out water and other small molecules. To generate their topology files, the processed protein and ligand are saved as separate files.
2. Generation of molecular topology files: A topology file is generated for the protein (and for small molecules/ligands if the simulated system is a protein/ligand complex). The topology of a molecular system can be generated only when the

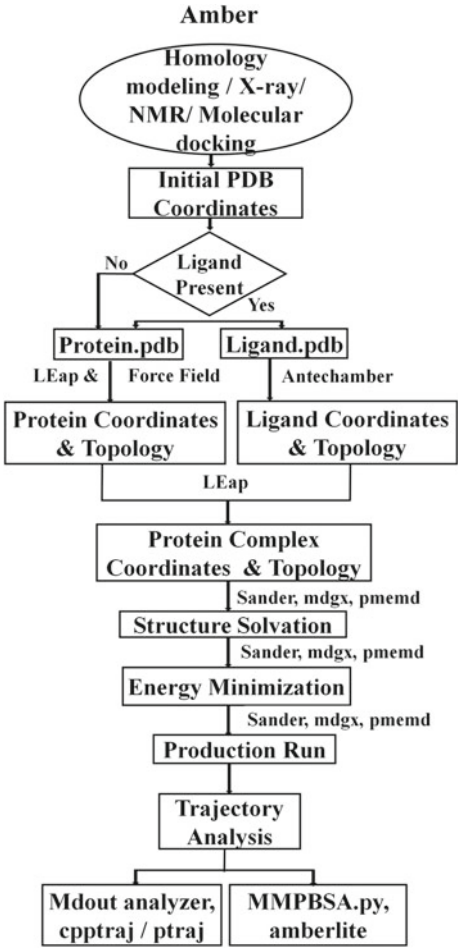

Fig. 10.2 Pipeline to run molecular dynamics simulations in AMBER

force files are available for all building blocks. There are two steps to generate the topology file for the protein complex. GROMACS pdb2gmx (GROMACS command) and external tools generate the topology file for a protein and its ligand, respectively. The command pdb2gmx converts a pdb file into a molecular topology file. The topology file contains a full description of all types of interactions in the protein. Small molecules are not recognized by the force fields used in the GROMACS. Hence, external tools are used to generate a topology file for the ligand based on the force field applied to the protein. If force fields from AMBER, CHARMM, GROMACS, and OPLS are used for the protein, Antechamber/ACPYPE, CGENFF, PRODRG/ATB, and Topolbuild/TopolGen can generate the requisite ligand topology files. The ligand topology file is inserted into the protein topology file ("top file") by adding a line (#include ligand.itp) into the protein.top after the position restraint line.

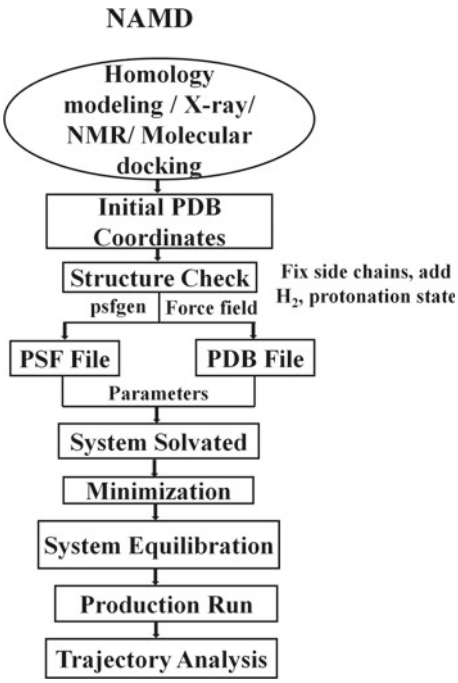

Fig. 10.3 Steps involved in molecular dynamics simulations in NAMD

3. Determination of a solvate box: The commands Editconf and genbox define a unit cell around a protein or protein complex and to fill said unit cell with water molecules, respectively. Before energy minimization, the water molecules are pre-equilibrated to remove water molecule overlap.
4. Energy minimization: A short minimization run removes large forces from structural distortions due to the addition of hydrogens and the broken hydrogen bond network in water. The grompp command minimizes a solvated protein for two reasons: (1) to remove local strains in the protein or protein complex due to the addition of hydrogen atoms and (2) to remove bad van der Waals interactions. This command collects all parameters to run the minimization, topology, and coordinate. It writes all results into a single tpr (portable binary run input) file containing the starting structure, molecular topology, and parameters form simulations.
5. Addition of ions: The pdb2gmx command also shows the charge of the protein. The tpr file is an input for the genion command to add positive or negative ions to neutralize the charged protein in the solvated system. The neutralized system is subject to energy minimization using grompp command to remove severe clashes.
6. Equilibration: The solvated system is equilibrated to avoid unnecessary protein distortions in the simulation. In this step, heavy atoms of the protein are fixed at

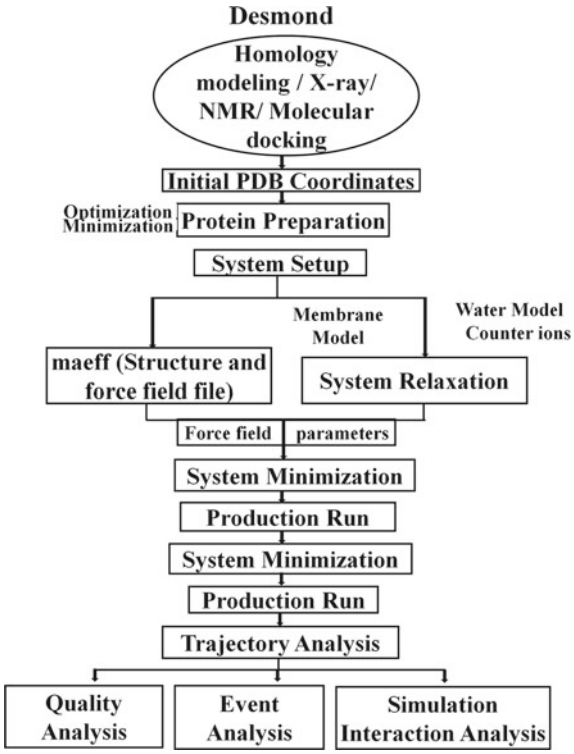

Fig. 10.4 Protocol to run molecular dynamics simulations using Desmond

the starting structure and only water molecules are permitted to move or relax around the protein.

7. Production run: The equilibration and production run differ only subtly. In a production run, the pressure coupling and position restraints are turned off. A production run uses a longer simulation and writes coordinates to a trajectory file.
8. Post-simulation analysis: GROMACS has various commands for analysis of the resultant trajectory file. Root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) are calculated using commands g_rms and g_rmsf, respectively. Commands are available in GROMACS for calculation of gyration of protein, energy, distance between atoms and hydrogen bonds and for secondary structure prediction.

10.4.2 AMBER

AMBER was developed in Peter Kollman's group by combining several force fields and since has been widely applied to various biomolecules. AMBER includes mul-

tiple classical molecular force fields for nucleic acids, small molecules, and amino acids and contains parameters for solvents, lipids, and carbohydrates. AMBER is not a single source MD simulation program but instead provides a set of programs that can run together. The atomic coordinates of the simulation system, the molecular topology file, force fields, and a script with all commands are prerequisites to running.

LEaP, tleap or xleap, is command line programs which generate or modify the parameter files for a new system or the existing systems. These programs convert most of the chemical structure file types (such as mol2 and pdb) to AMBER parameter file types (.lib, .prepi, parm.dat, and .frcmod). The energy minimization and dynamics simulation parameter files contain key information required to run the simulation. These commands also generate topology (.prmtop, .parm7, or .top) and coordination (.inpcrd and .crd) files based on the parameter files.

The pdb4AMBER command converts pdb files from different sources like X-ray, nuclear magnetic resonance, and homology modeling to a format suitable to LEaP.

The parmed command is used to validate and extract parameter information from the parameter–topology file for the simulation system. This command also makes minor modifications to parameter–topology file.

The AntechAMBER command uses the general AMBER force field (GAFF) to generate and modify the parameter files for small molecules and amino acids.

The program sander is used to execute the energy minimization, equilibration, and production runs. Sander interactively relaxes the system by iteratively moving the atoms until a suitable low average gradient is reached. The system configuration files are automatically generated during the production run by integrating Newtonian equations of motion. The final production run generates a configurational space and allows the structure to cross over smaller potential energy barriers than the energy minimization. During dynamics simulations, the configuration of the system is saved at the regular time interval for later analysis. Thermodynamic integration is used to calculate system free energy. Sander module is used to study protein conformational search, protein structure modeling, and structural refinement of proteins.

The program pmemd (Particle Mesh Ewald Molecular Dynamics) is the updated version of sander with increased speed via parallel scaling during the production run. The required pmemd input and output files are very similar to sander.

The program mdgx is a dynamics engine written in C and sorts the atoms to simplify the information flow during force calculation. It adopts select features from pmemd and sander. The most common use of mdgx is redesign of dynamics algorithms or models to support parameters for new models.

The program mdout_analyzer.py is a simple Python script which summarizes the information from pmemd or sander output files. The program Cpptraj, written in C++, analyzes trajectory files such as coordinate extractions, RMSD, RMSF, bond and angle calculations, and hydrogen bond analysis. The Python program pytraj has flexibility to analyze data from trajectory files.

10.4.3 NAMD

NAMD is written in Charm++ [57]. It was developed by the Theoretical and Computational Biophysics Group (TCB) and Parallel Programming Laboratory (PPL) at the University of Illinois at Urbana-Champaign.

The 3D structure of protein atomic coordinates from PDB, protein structure file (psf), force field parameter file, and configuration files are prerequisites to running a NAMD simulation. The initial protein 3D structure downloaded from PDB does not contain hydrogen atoms because X-ray crystallography cannot resolve the hydrogen atoms. Hence, the command psfgen generates the pdb file with hydrogen atom coordinates. The generated pdb file is subject to energy minimization to ensure reasonable atomic coordinates. The minimized protein coordination file is immersed in a water box to adopt a cellular environment for the target protein. The protein can be solvated in two ways for energy minimization and equilibration: (i) vacuum, without periodic boundary conditions, or (ii) water box, with periodic boundary conditions.

The important features of NAMD are given below:

(1) Compatibility with the CHARMM force field to stimulate the systems. This force field was used by other programs such as CHARMM and X-PLOR, hence it is easy to migrate from one program to another as well as to analysis the trajectory files.
(2) Efficiently utilization of the Particle Mesh Ewald algorithm—full electrostatic interactions to reduce the electrostatic complexity.
(3) Leverages the Verlet integration method—multiple time steps are applied to compute local interactions at each time step and to reduce the computational cost for long-range interaction calculations.
(4) Variety of simulation options.
(5) Easy to modify and extend a given run.
(6) Interactive simulations.

10.4.4 Desmond

Desmond was developed by D. E. Shaw Research group using numerical methods [58] and novel parallel algorithm [59] to run high-speed MD simulations for biological systems. It is integrated with the Maestro modeling environment (Schrodinger, Inc), a commercial software package compatible with VMD and analysis tools.

Proteins directly downloaded from PDB are not suitable for dynamics analysis because they lack hydrogen atoms and contain poorly defined bond orders and formal charges. Hence, protein preparation is the initial and a very important step in MD simulations. The basic steps to run MD simulations using Desmond from Maestro are given below:

1. Import the protein 3D structure file either from PDB or from a homology modeling method into Maestro.

2. Prepare the protein structure using the Protein Preparation Wizard from Maestro. Ions and other small molecules are removed, bond orders are corrected, and the missing residues, atoms, and hydrogens are added to the system. Lastly, reliable protonation states of the residues are set.
3. Generate a solvated system using a Solvation tab. This step adds water molecules around a protein or inserts a protein into a lipid layer. The Solvation tab has the following options: (a) none—Do not use the solvent (run the system in a vacuum); (b) predefined—Select one of the solvent models such as SPC, TIP4P, TIP3P, and TIP4PEW and three organic solvents, methanol, octanol, and dimethyl sulfoxide (DMSO); (c) custom—Import the solvent system from a file. The Solvation tab also sets up the periodic boundary box by specifying the shape and size of the box. To proceed with this step, first select the shape of the box from the three basic shapes: cubic, orthorhombic, and triclinic. While selecting, the box size calculation has two options: (i) buffer—Calculate the size using the given buffer distance between the solute and the box boundary and (ii) absolute size—Specify the distance between the solute and the simulation box.
4. Insert the protein into a lipid layer with the setup membrane from system builder panel if the target system is a membrane protein. DPPC, POPC, and POPE are the three membrane models supported by Desmond.
5. Calculate the charge of the system. Positive (Na^+) or negative (Cl^-) ions are added to neutralize the solvated system or to set a desired ionic environment for the protein.
6. Before MD simulations, relax the whole system into a local energy minimum either by minimization or by selecting the panel options. In this step, the system is minimized by the steepest descent method followed by limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithms. Generally, two parameters are set for the minimization: (i) maximum number of iterations and (ii) convergence threshold for the gradient.
7. Utilize the desmond panel to set the simulation parameters such as production run, simulated annealing, or replica exchange for the system. The simulation section is where the user specifies the simulation time (in nanosecond) and the recording time interval (in picosecond) for the energy calculation and between the snapshots in the trajectory file. Additionally, selecting the ensemble type (such as NVT, NPT, NVE, NPγT, and NPAT) to set the temperature, pressure, and surface tension occurs in the simulation Section.
8. Finally, run the simulation. The resultant trajectory is analyzed by the simulation quality analysis or the simulation event analysis panel. The result includes a summary of the simulation and analysis of total and potential energy, temperature, pressure, and volume throughout the simulation.

10.5 MD Simulation Protocol

An overview of the MD simulation protocol is depicted in Fig. 10.5.

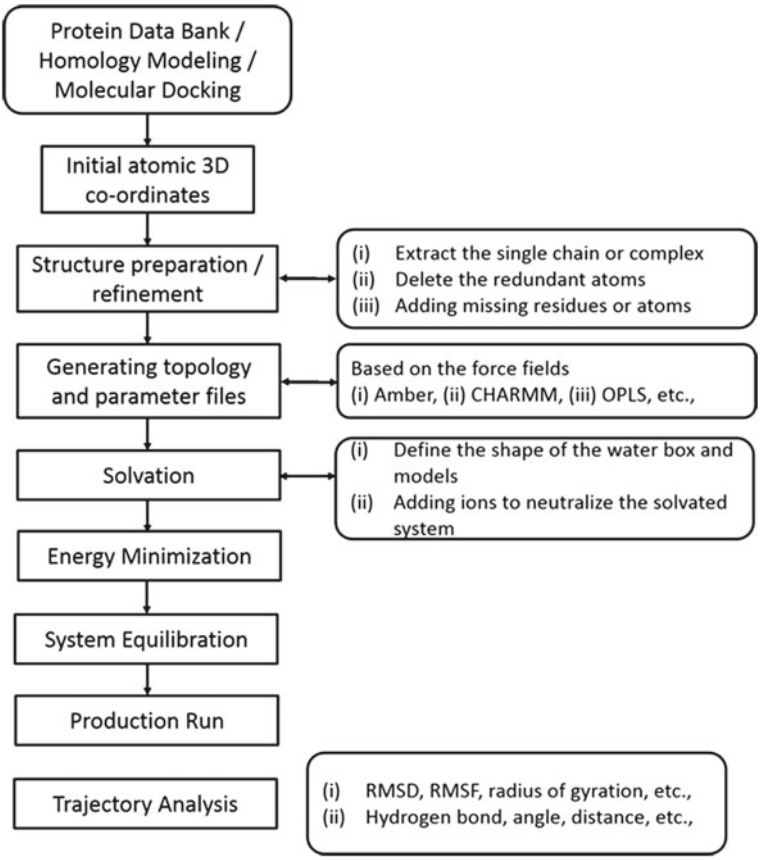

Fig. 10.5 Overview of the MD simulation protocol

10.5.1 Initial Coordinates

The first step in an individual MD simulation is to select or obtain the 3D structure of a protein or a protein complex. X-ray and nuclear magnetic resonance are the two experimental techniques to determine 3D structures deposited in PDB. If the 3D protein structure is not available in PDB, homology modeling can be used to construct a 3D protein structure. The selection of the atomic representation of a protein is an important step in MD because it can influence the dynamics simulation results.

10.5.2 Preparation of the Protein Structure

After the initial 3D structure is obtained, the following steps should be carried out before the energy minimization.

1. Remove the redundant atoms: The experimentally determined 3D structure of a protein might be in the dimeric or multimeric form with stabilizing chemical agents from the structure determination. Here, the monomer of the protein is extracted from the multimeric form of the protein. The small atoms or chemical agents which are not paramount to protein function should be removed.
2. Add missing residues and atoms: Some proteins from PDB might lack residues or atoms in their coordinate files. Hence, it is important to cross-check whether the selected protein structure has any missing residues or atoms. Once missing residues or atoms are found, the missing regions should be rebuilt using computational software. After the missing residues or atoms are rebuilt in the structure, hydrogen atoms should be added and the protonation states should be assigned ionizable residues such as arginine, glutamate, lysine, aspartate, and histidine which play a very important role in most of the protein functions.
3. Replace mutated atoms: During crystallization, some protein–ligand complexes are not able to form a stable structure. Thus, some residues in a protein complex were mutated to obtain a stable complex. Such residues should be replaced by the appropriate native residues.

10.5.3 Generating Topology and Parameter Files

The topological file for a given protein contains all geometric information including angles, bonds, and interactions. Once the protein is prepared, it is next important to generate the topology files for the protein and small molecules by applying the appropriate force field [60–65]. Sometimes, the topology file combines the parameter files (describing the potential energy of the systems) that are generated by applying the chosen force field. Various force fields have been developed for proteins, lipids, nucleic acids, carbohydrates, and small molecules. The force field selection is guided by the nature of the protein and the purpose of the research. There are various utilities to generate the topology files for small molecules which are compatible for most MD simulation programs.

10.5.4 Solvating the System

Explicit and implicit models are two different water models simulating a protein or a protein complex in an aqueous solution. The implicit solvent model is efficient and yields a reasonable description of the behavior of the solvent but fails to provide

details about solute molecule fluctuation in the solvent density. Explicit models such as SPC and TIP3P are used in most MD simulation programs and detail the solvent spatial descriptions. Membrane proteins should be inserted into a pre-equilibrated lipid layer. After system assembly in solute or lipid, new topology files are built to run energy minimization and reduce van der Waals interactions.

10.5.5 Adding Ions

A molecular system should be solvated in a pre-equilibrated water box. The protein in the solvated water box might carry a charge. Hence, the system is neutralized by adding ions including Na^+, Cl^-, and K^+. Positive ions are employed for systems with negative charge and vice versa.

10.5.6 Minimizing Structure

Energy minimization is a crucial step in MD simulations. This step configures the protein system to be stable and removes bad clashes between the solute and solvent, as well as filtering out inappropriate geometry.

10.5.7 Establishing Equilibration Dynamics

A reasonable structure obtained from energy minimization is required to continue the simulation. Equilibration is an important step preceding the production run. The main purpose of equilibration is to reduce non-equilibrium effects and to avoid or reduce structural distortions. The protein (embedded in the box of explicit system solvents) is fixed, and water molecules are permitted to move about the protein. Once the water molecules are equilibrated, the protein constraint is removed and the whole system is equilibrated together. MD simulation begins in earnest only when the system equilibrates the ions and solvents around the protein. There are three ensembles in the equilibration run: (i) NVE, constant number of the particles, volume, and total energy; (ii) NVT, constant number of particles, volume, and temperature; and (iii) NPT, constant number of particles, volume, and temperature. In the NVT ensemble, the thermostat maintains the system temperature to reach a desired value. Next, NPT stabilizes system pressure. The final production run is executed only when the system is temperature and pressure stable.

10.5.8 Establishing Production Dynamics

A well-equilibrated system from the previous step run in the final production MD simulation.

10.5.9 Analyzing Trajectory

The trajectory file obtained from the production run needed to compute the structural, thermodynamics, and dynamic nature of the protein or protein complex by calculating the RMSD, RMSF, radius of gyration, extract a representative structure, and more. The quality of the representative structure from the trajectory can be verified using Ramachandran plots. Visualizing trajectory files using various molecular graphics software packages yields insight into atomic-level protein conformational changes based on the time interval. In addition, detailed quantitative structural information such as hydrogen bonds, inter- and intra-molecular interactions, radius of gyration, bond angles, distance and geometrical quantities is calculated. RMSD and RMSF show the deviation of the protein structures from the initial structure based on the time interval and different flexible region of the protein, respectively. The free energy decomposition can be calculated from the trajectory files using the MM-PB(GB)SA method [66].

10.6 Applications of MD Simulation

Predicting compound toxicity by leveraging computational methods is an emerging yet important field [67]. MD simulation has been applied to predict toxicity. Below are a few examples of recent MD simulation applications in computational toxicology.

10.6.1 Binding Interactions Between Chemicals and Human Nicotinic Acetylcholine Receptor α4β2

Tobacco product addiction is a major global health concern. Among various tobacco constituents, nicotine plays a major role in tobacco addiction by binding to the neuronal nicotinic acetylcholine receptors (nAChRs). Among different types of nAChRs, α4β2 mediates nicotine addiction. Hence, pinpointing detailed interactions between nicotine and the human α4β2 receptor will reveal the mechanism of nicotine addiction. In addition, the interaction can inform a predictive model to screen tobacco constituents. To address these questions, homology modeling, molecular docking,

and dynamics simulation techniques elucidated binding interactions between the human α4β2 and tobacco constituents [68].

Initially, the 3D structure of the human α4β2 extracellular domain was constructed utilizing the crystal structure of Ct-AChBP and homology modeling. MD simulations were used to minimize the constructed human α4β2 3D structure. Eleven components which were both crystallized with nAChRs or AChBP and also have binding data for α4β2 were selected to dock with α4β2 protein. The LigPrep2.0 tool from Schrodinger optimized the 11 ligands (Fig. 10.6) by applying the OPLS_2015 force field (downloaded from PDB). The standard precision (SP) tool from Glide docked the optimized or minimized ligands in the active site of α4β2. The active site of α4β2 consists of residues V96, Y98, S153, W154, Y195, and Y202 from α4 and W57, V111, F119, L121, and F157 from β2. The best α4β2 active site binding pose for each ligand was selected based on the Glide scores. Interaction studies revealed that the hydrophobic and aromatic interactions play a major role in ligand binding in the active site of α4β2. The hydrophobic part of the ligand bounds to the major surface α4β2 binding pocket extended region.

The selected best α4β2 complexes (11 total) obtained from molecular docking was optimized with AMBER. The AMBER ff99SB and general AMBER force field (GAFF) force field were applied to the α4β2 protein and the ligands, respectively. Antechamber assigned the AM1-BCC charges for the ligands. Each α4β2 complex was solvated in a TIP3P water box with 10 Å truncated octahedron box. Added sodium ions neutralized the systems. The SHAKE and Particle Mesh Ewald (PME)

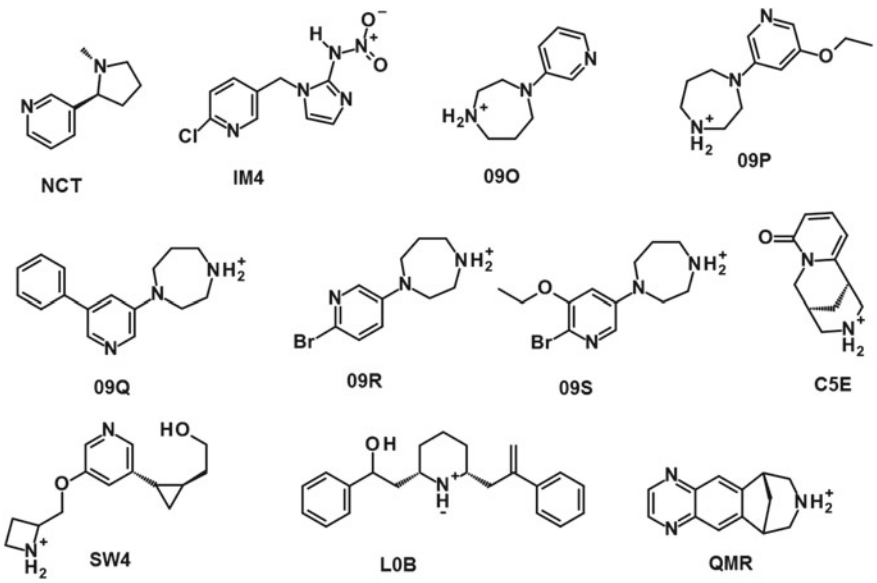

Fig. 10.6 Two-dimensional chemical structures of 11 compounds obtained from PDB. The three letters under the structure represent the ligand ID from PDB

constrained the bonds with hydrogens and applied long-range interactions for the non-bonded interactions. Two-thousand-step steepest descent minimization followed by 8000-step conjugated gradient minimization was carried out for each entire system prior to the final production run. Each system was subjected to 200 ps of equilibration to reach the stable states. The equilibrated system was subsequently subjected to a 10 ns production run at 300 K temperature, 1 atm pressure, and 2 fs of time steps.

The resultant trajectory files were used to calculate the RMSD values, demonstrating that the α4β2 complexes reached stable states. The last snapshot from each trajectory file was selected to elucidate the interactions between the 11 ligands and the critical residues in the active site of α4β2 (Table 10.2).

10.6.2 Identification of the Ligand Binding Mode in α-Fetoprotein

By aid of a transport protein, endocrine disrupting chemicals (EDCs) enter the cell and cause adverse effects by disrupting endocrine receptors including androgen and estrogen receptors. α-fetoprotein (AFP) blocks AR and ER mediated responses by binding with androgens and estrogens to prevent entry into the target cells. Thus, to understand the endocrine disruptive potential of a chemical, it is important to elucidate binding to AFP. In 2012, Hong et al. measured the AFP binding affinity of 125 structurally diverse chemicals and discovered that 53 chemicals are AFP binders and 72 are non-binders. To assess risk of endocrine disrupting chemicals, the binding mode of rat AFP–ligand was studied. Initially, homology modeling was used to constructed a 3D structure of rat AF which was further optimized by an MD simulation. The resultant trajectory file from the MD simulation informed selection of one representative AFP protein 3D structure. This chosen structure was used as a receptor to dock the 13 classes of chemicals in its active site. Finally, the 13 complexes selected from the molecular docking were refined using MD simulations to understand the binding patterns of these chemicals in the active site of rat AFP protein [69].

Molecular docking is one of the well-known techniques to predict ligand pose in the active site of a protein. The major drawback of the rigid molecular docking is its failure to give flexibility to the protein. AFP is a flexible protein which undergoes an active site ligand binding-induced conformational change [70]. Hence, the MD simulation method gave flexibility to the AFP–ligand complex and optimized complexes obtained from molecular docking. The 13 rat AFP–ligand complexes from molecular docking were subjected to MD simulations to provide insight into the conformational changes of rat AFP due to the 13 structurally diverse ligands binding in the active site.

Prior to molecular docking, the 13 diverse ligands were optimized using Gaussian 09 with the basis set of 6-31G. The electrostatic potential of the ligands was calculated with the mechanical method (B3LYP). For MD simulations, GAFF was applied

Table 10.2 Interactions between the 11 ligands and critical residues in human α4β2 nAChR

Ligand ID	Name	H-bond	Pi-Pi	Pi–cation	Pi-Sigma	Hydrophobic
NCT	L-(-)-nicotine	NCT:H15-153:O	Y98-NCT	Y98, W154, Y202-NCT:N2	W57-NCT:H1, H5	V96, L97, F151, L121
IM4	Imidacloprid		Y195-IM4			V111, F119, F172
09O	1-(pyridin-3-yl)-1,4-diazepane	09O:H5-W154:O	Y195-09O	W154-09O:N1 W57-09O:N1		L121
09P	1-(5-ethoxypyridin-3-yl)-1,4-diazepane	09P:H3-W154:O		W154,Y195,Y202-09P:N1		V96, L97, L121
09Q	1-(5-phenylpyridin-3-yl)-1,4-diazepane			Y195-09Q:N2 F172-09Q:N2		M36, L121, L169, F172
09R	1-(6-bromopyridin-3-yl)-1,4-diazepane		W57-09R	Y202-09R:N3		V96, F119, L121, F172
09S	1-(6-bromo-5-ethoxypyridin-3-yl)-1,4-diazepane	Y195:HH-09S:N3	Y195-09S	Y98, W154-9S:N1		V96, L97, F119, L121, F172
LOB	Alpha-lobeline		W154-LOB	W154,Y195, Y202-LOB:N1	LOB-T155:HA	V96, L97, F151, A110, V11, F119, L121
QMR	Varenicline		Y202-QMR W57-QMR	Y195-QMR:N1 W57-QMR:N1	Y195-QMR:H1	V96, L97, F15, F119, L121
SW4	2-[(1R,2S)-2-[5-[[(2S)-azetidin-2-yl]methoxy]pyridin-3-yl] cyclopropyl]ethanol	Y195:HHSW4:O1		F119-SW4:N2		V111, F119, F121
C5E	Cytisine	C5E:H6-S153:O; C5E:H15-W154:O	W57-C5E	W154, Y202-C5E:N2	Y202-C5E:H4	V96, L97, F151, L121

to the optimized 13 AFP-complexes. Before the 60 ns production run, the AFP-complexes were minimized, equilibrated, and heated. During the simulations, the atomic coordinates of the AFP-complexes were saved in the trajectory file every 10 ps.

The RMSD plot for the protein and ligand was the basis for system dynamic stability estimation, and the four representative structures are plotted in Fig. 10.7. The AFP complexes with 2,3,4,5 tetrachloro-4'-biphenylol, DES, and flavanone were stabilized at the last 10 ns. Binding of α-zearalanol and weak binders (chalcone and diethyl phthalate) had considerable fluctuations in last 5 ns.

The diethyl phthalate came out of the binding pocket of AFP at 10 ns, indicating low binding effect. In support of this MD simulation result, Hong et al. experimentally proved that the diethyl phthalate has a weak binding affinity for AFP [71]. RMSD analysis of the AFP complexes trajectory files revealed multiple rotatable bonds in the ligands induced a remarkable conformational change on AFP compared to a rigid ligand. The ligand binding-induced conformational changes in AFP's active site were observed in the last frame of the trajectory files, indicating that MD simulations can capture ligand binding protein structural changes.

10.6.3 Ligand Binding Interactions of Human α7 Nicotinic Acetylcholine Receptor

Despite the highly publicized adverse health effects caused by tobacco use, the addiction associated with tobacco product nicotine has led to difficulty in quitting among users. The neuronal nAChRs play a key role in tobacco addiction. The α7 subtype of nAChRs is a key receptor in addiction mediation. Worryingly, it is unknown whether other tobacco constituents (>8000 exist) are addictive or not. Experimentally determining the binding affinity of ~8000 tobacco constituents would be too expensive and time-consuming to be practical. Instead, in silico methods evaluated the addiction potential of ~8000 tobacco constituents. Due to the absence of an experimental crystal structure, homology modeling was leveraged to build the 3D structure of the human α7 nAChR ligand binding domain. The α7 nAChR chimera (PDB ID:3SQ6) was the template to model human α7 nAChR. When undergoing such modeling, it is essential to consider the flexibility of the protein in the presence of the ligand as the target protein can undergo ligand binding driven structural changes which impact function. Most software packages permit only partial protein flexibility during small molecule docking. Here, a competitive docking model (CDM) overcame this drawback of rigid protein docking and found the poses of compounds when binding to the homology modeled human α7 nAChR. Compounds experimentally tested against human α7 nAChR were used to evaluate the ability of the CDM. The model's predictions of compound binding in the active site of the human α7 nAChR were thus experimentally validated. MD simulation was used to investigate the residues involved in the critical compound interactions. Building on the elucidated key inter-

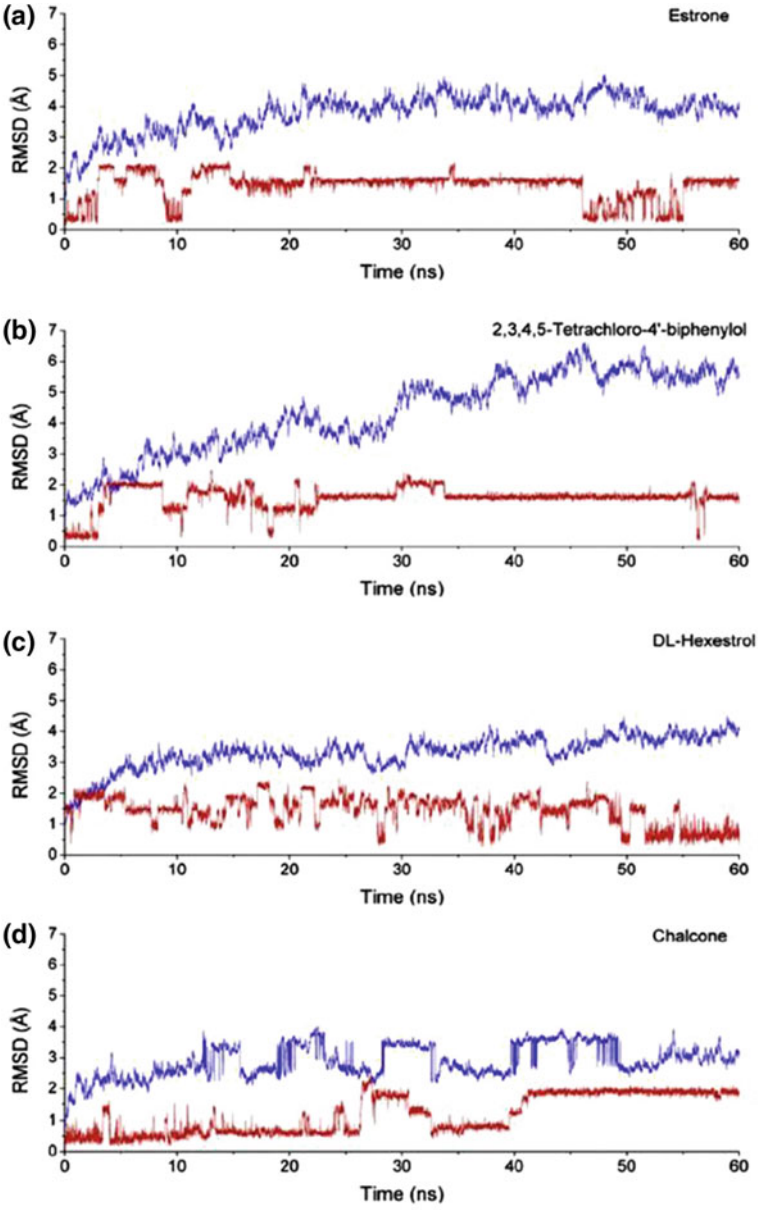

Fig. 10.7 RMSD from the trajectory files of rat AFP complexes. Blue line represents the protein backbone RMSD, and red line represents the ligand RMSD. **a** AFP–estrone, **b** AFP-2,3,4,4-tetrachloro-4′-biphenylol, **c** AFP-DL-hexestrol, and **d** AFP-chalcone

acting residues from MD simulation, a predictive model was developed to predict tobacco constituents binding activity. Training and test set compounds from PDB informed and evaluated the model (Fig. 10.8). The predictive model successfully differentiated α7 nAChR binders from non-binders. Of 12 experimental test compounds, 11 were predicted correctly as nAChR α7 binders by the predictive model. Hence, this predictive model might be helpful to screen tobacco constituents which are responsible for addiction.

10.6.4 Antagonist Binding-Induced AR Structural Changes

Androgen receptor (AR) is as prostate cancer target is gaining traction in the drug discovery field. AR activity is blocked both by non-steroid and steroid antagonists. Prolonged use of AR antagonists causes mutations in AR which paradoxically cause AR antagonists act as AR agonists. Driven by the binding of agonist/antagonist, AR undergoes a considerable conformational change which impacts DNA and co-regulatory protein binding. The X-ray crystal structure of antagonist binding in the ligand binding pocket of wild-type (WT)-AR is required to understand the mechanism. Unfortunately, such a crystal structure does not exist. To address this challenge, induced fit molecular docking and MD simulations were used to characterize AF2 site structural changes and the mechanism of antagonist binding in the ligand binding pocket of WT AR. Induced fit molecular docking and a long-term MD simulation were leveraged to construct the WT-AR-antagonist complex and to analyze WT-AR structural changes, respectively. Three molecular systems (WT-AR bound by agonist R1881, WT-AR bound by antagonist bicalutamide, and Mutant-AR bound by bicalutamide) were utilized to study the agonist binding-induced structural changes in the WT-AR AF2 site. WT-AR bound to agonist R1881 and mutant-AR bound to bicalutamide were selected from PDB. Induced fit docking (IFD) and MD simulations generated the structure of WT-AR bound to antagonist bicalutamide. Twenty-five binding poses of bicalutamide in the ligand binding pocket of WT-AR were obtained from IFD. Among the 25 poses, the best pose of bicalutamide in the ligand binding pocket of WT-AR was selected based on both the IFD scores and also critical residue interactions around the active site. The three systems were subject to 1 microsecond of an MD simulation using AMBER 14 to optimize the three complexes. The RMSD and RMSF plots revealed that the systems were stable throughout the simulations (Fig. 10.9a, b). Five residues had shown a deviation of more than 2 Å. As expected, those residues are present in the loop regions (Fig. 10.9c). The representative structure from the trajectory files was superimposed on the X-ray crystal structures of WT-AR with R1881 and mutant-AR with bicalutamide, which demonstrates the reliability of the MD simulations (Fig. 10.10). The binding of bicalutamide in the ligand binding pocket moved residues V716, K720, Q733, M734, Q738, and E897 and changed the structure of the AF2 site of WT-AR, rendering AF2 unsuitable for co-activator binding. The electrostatic potential map revealed that residues V716/K720/Q733 or M734/Q738/E897 played a vital role in the formation of the positive and negative

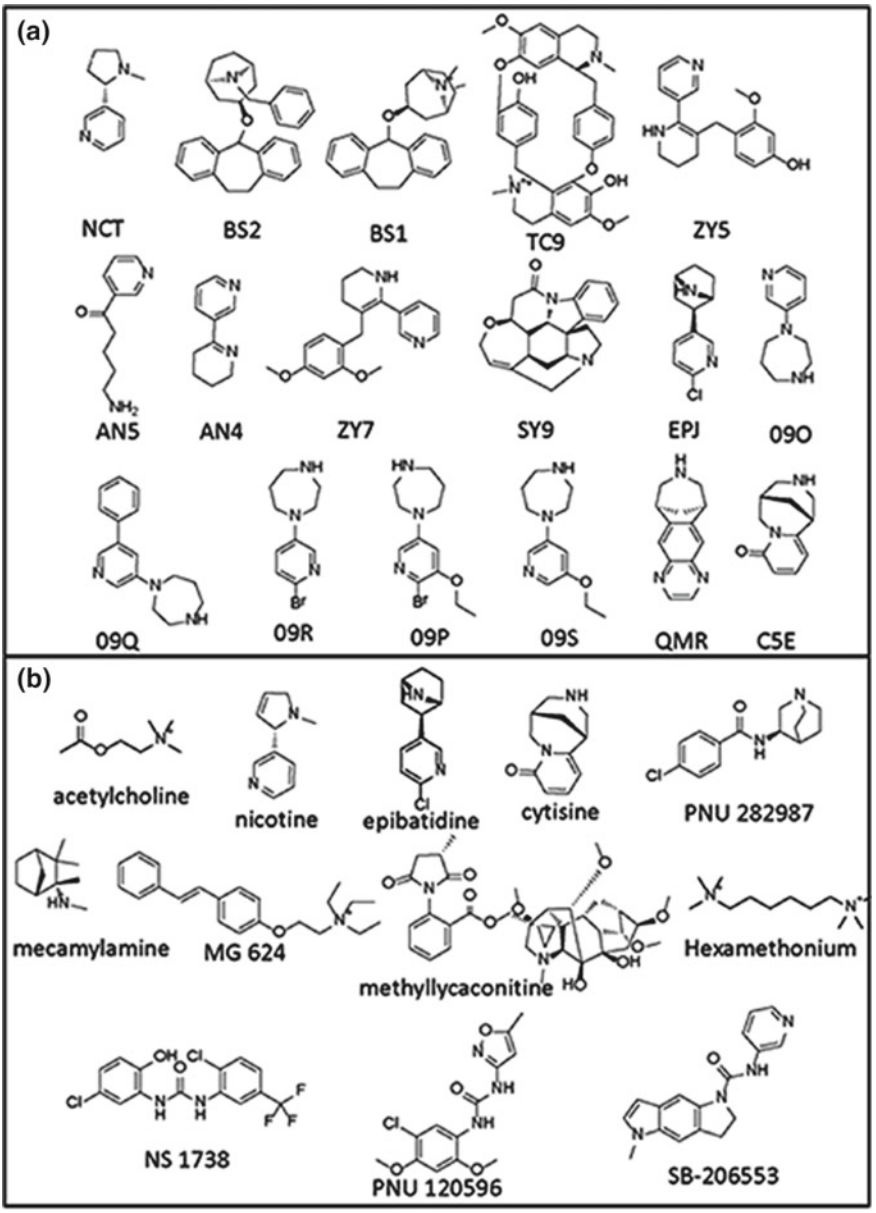

Fig. 10.8 Training **a** and test **b** compounds for developing and validating the predictive model

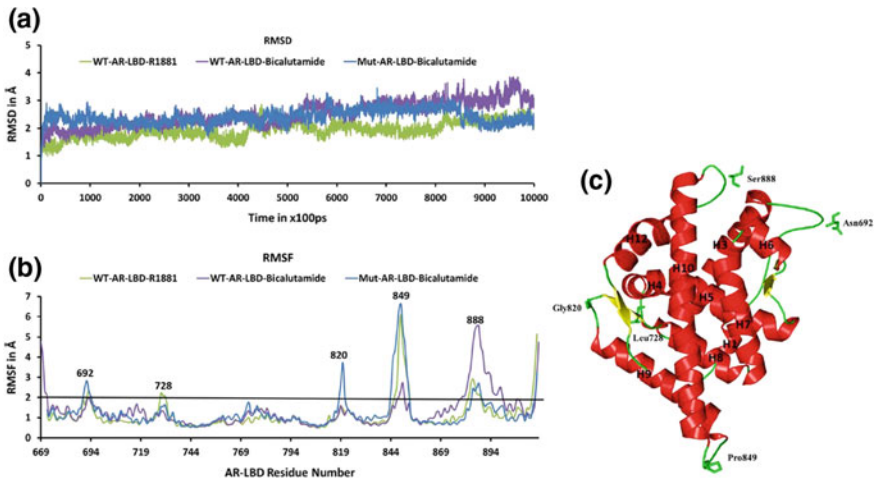

Fig. 10.9 Analysis of 1 μs MD simulation trajectory files. **a** RMSD, **b** RMSF, and **c** WT-AR-R1881 are drawn in the ribbon model, and the five residues which showed a deviation greater than 2 Å are shown in stick model. Green: WT-AR-R1881; purple: WT-AR-bicalutamide; blue: mutant-AR-bicalutamide

charge clump on AF2 site (Fig. 10.11). The charge clumps in AF2 provide a suitable place for the co-regulator protein to bind tightly. To summarize, MD simulation enabled the discovery of the charge clump disruption caused by antagonist binding in the WT-AR ligand binding pocket.

10.7 Future Perspectives

To date, many chemicals lack toxicity information. Understanding chemical properties or classifying a chemical as a toxic or non-toxic by experimental methods is often prohibitively time-consuming and expensive. Seeking to overcome this fundamental limitation, computational toxicology is attractive for chemical toxicity prediction. Computational toxicology integrates data or information from various sources to develop predictive models based on the mathematical computer calculation. Recently, MD simulation has emerged as an attractive computational toxicology technique. Understanding the algorithm behind various MD simulation methods will empower researchers to reveal the solution for various problems. Along with MD simulation algorithms, input data quality and user simulation understanding fundamentally make or break toxicity prediction success.

Most MD simulation results are validated using experimental data to confirm reliability. Protein conformational changes, strong interaction of ligand in the binding pocket of the protein, and hot spot residues are validated through MD simulations. X-ray and nuclear magnetic resolution are experimental methods which pinpoint both

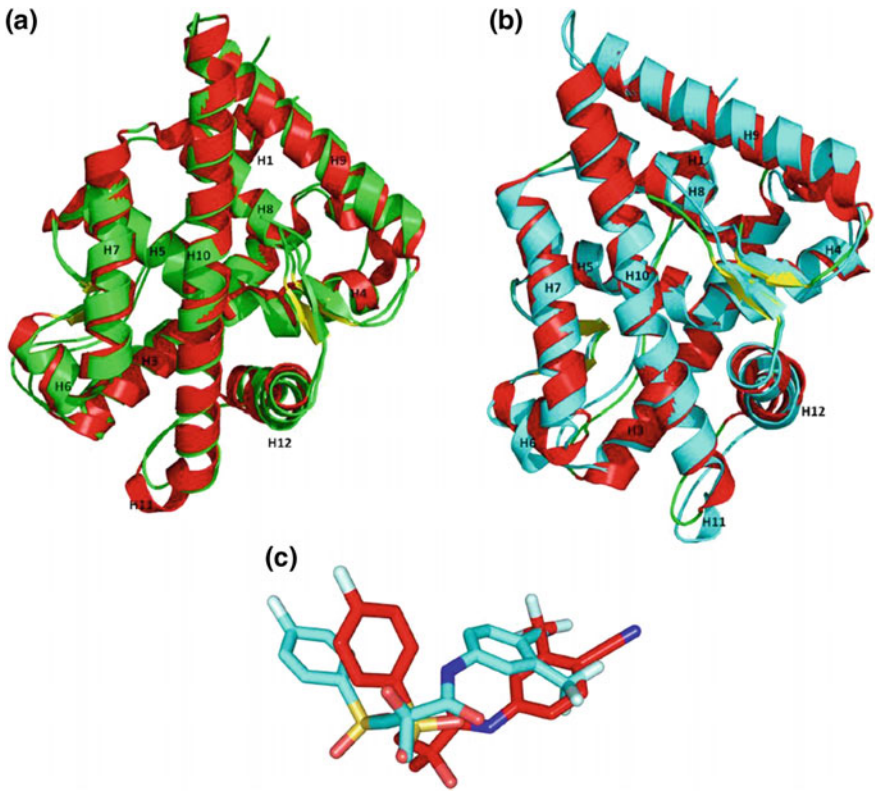

Fig. 10.10 Superimposition of the X-ray crystal structures against the representative structures from the 1 microsecond trajectory files. **a** WT-AR-R1881, **b** mutant-AR-bicalutamide, and **c** superimposition of bicalutamide from PDB and the representative structure of mutant-AR-bicalutamide. Red: X-ray crystal structure; green: WT-AR-R1881; cyan: mutant-AR-bicalutamide

protein conformations and interactions between the small molecules and the proteins. Nuclear magnetic resonance provides multiple conformations of a single protein and is often used to analyze structural changes of proteins including protein recognition and folding, conformation, and dynamic changes. The MD trajectory files can be compared with nuclear magnetic resonance data to check the reliability of the MD simulation result. MD simulations will be applied to understand or study the quality of homology modeled protein structures, to identify the structural or conformational changes of proteins, to elucidate important interactions between a protein and small molecules, and to estimate protein–protein and protein–ligand binding affinity.

A recent advance in MD simulation algorithms and computer hardware will enable microsecond-scale MD simulations for the macromolecules such as protein or protein complexes. Nowadays, many researchers and scientists run microsecond and millisecond MD simulations to characterize protein structures [72–75]. In the future,

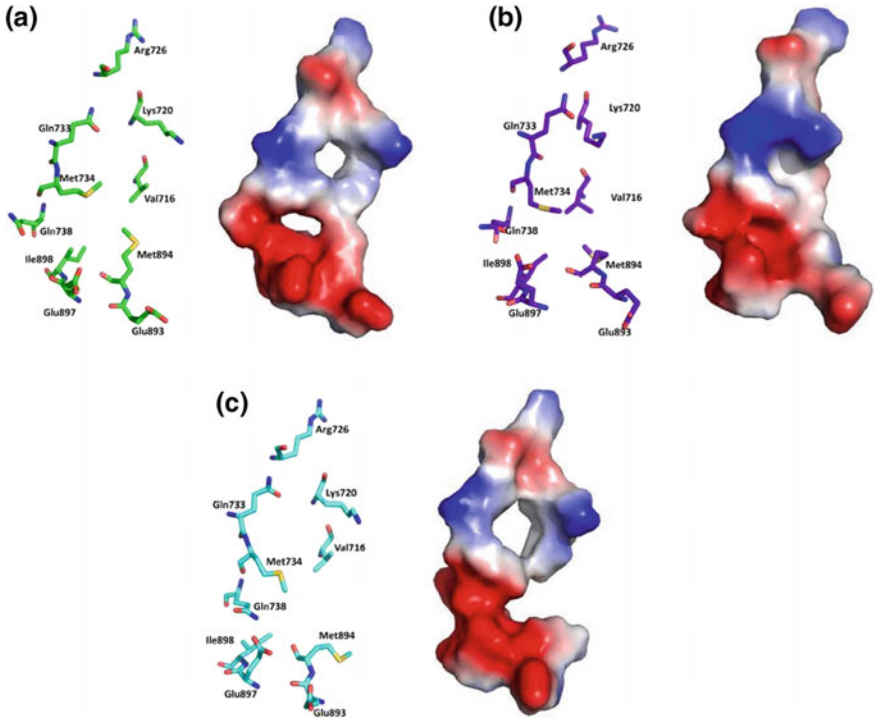

Fig. 10.11 Electrostatic potential map for AF2 site. **a** WT-AR-R1881, **b** WT-AR-bicalutamide, and **c** mutant-AR-bicalutamide

we expect that long-term MD simulation will be a powerful tool to elucidate or capture protein structural changes.

Acknowledgements This chapter is supported in part by an appointment to the Research Participation Program at the National Center for Toxicological Research (Sugunadevi Sakkiah) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the US Food and Drug Administration. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Food and Drugs Administration.

References

1. Neamati N, Hong H, Owen JM, Sunder S, Winslow HE, Christensen JL, Zhao H, Burke TR Jr, Milne GW, Pommier Y (1998) Salicylhydrazine-containing inhibitors of Hiv-1 integrase: implication for a selective chelation in the integrase active site. J Med Chem 41:3202–3209
2. Hong H, Neamati N, Winslow HE, Christensen JL, Orr A, Pommier Y, Milne GW (1998) Identification of Hiv-1 integrase inhibitors based on a four-point pharmacophore. Antiviral Chem Chemother 9:461–472

3. Neamati N, Hong H, Sunder S, Milne GWA, Pommier Y (1997) Potent inhibitors of human immunodeficiency virus type 1 integrase: a novel four-point pharmacophore searching of the NCI 3D database. Mol Pharmacol 52:1041–1055
4. Hong H, Neamati N, Wang S, Nicklaus MC, Mazumder A, Zhao H, Burke TR, Pommier Y, Milne GWA (1997) Discovery of Hiv-1 integrase inhibitors by pharmacophore searching. J Med Chem 40:930–936
5. Luo H, Du T, Zhou P, Yang L, Mei H, Ng HW, Zhang W, Shu M, Tong W, Shi L, Mendrick DL, Hong H (2015) Molecular docking to identify associations between drugs and class I human leukocyte antigens for predicting potential idiosyncratic drug reactions. Comb Chem High Throughput Screen 18(3):296–304
6. Drake RR, Neamati N, Hong H, Pilon A, Sunthankar P, Hume SD, Wilne GWA, Pommier Y (1998) Identification of a mononucleotide binding site in human HIV-1 integrase. Proc Natl Accad Sci USA 98:1495–1500
7. Ng HW, Leggett C, Sakkiah S, Pan B, Ye H, Wu L, Selvaraj C, Tong W, Hong H (2018) Competitive docking model for prediction of the human nicotinic acetylcholine receptor α7 binding of tobacco constituents. Oncotarget 9(24):16899–16916
8. Luo H, Mattes W, Mendrick DL, Hong H (2016) Molecular docking for identification of potential targets for drug repurposing. Curr Top Med Chem 16(30):3636–3645
9. Ng HW, Shu M, Luo H, Ye H, Ge W, Perkins R, Tong W, Hong H (2015) strogenic activity data extraction and in silico prediction show the endocrine disruption potential of bisphenol A replacement compounds. Chem Res Toxicol 28(9):1784–1795
10. Hong H, Hong Q, Perkins R, Shi L, Fang H, Su Z, Dragan Y, Fuscoe JC, Tong W (2009) The accurate prediction of protein family from amino acid sequence by measuring features of sequence fragments. J Comput Biol 16(12):1671–1688
11. Liu J, Mansouri K, Judson R, Martin MT, Hong H, Chen M, Xu X, Thomas R, Shah I (2015) Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem Res Toxicol 28:738–751
12. Luo H, Ye H, Ng HW, Shi L, Tong W, Mendrick DL, Hong H (2015) Machine learning methods for predicting HLA-peptide binding activity. Bioinform Biol Insights 9:21–29
13. Tong W, Hong H, Fang H, Xie Q, Perkins R (2003) Decision forest: combining the predictions of multiple independent decision tree models. J Chem Inf Comput Sci 43:525–531
14. Hong H, Tong W, Xie Q, Fang H, Perkins R (2005) An in silico ensemble method for lead discovery: decision forest. SAR QSAR Environ Res 16:339–347
15. Xie Q, Ratnasinghe LD, Hong H, Perkins R, Tang ZZ, Hu N, Taylor PR, Tong W (2005) Decision forest analysis of 61 single nucleotide polymorphisms in a case-control study of esophageal cancer; a novel method. BMC Bioinformatics 6(Suppl 2):S4
16. Hong H, Tong W, Perkins R, Fang H, Xie Q, Shi L Multiclass decision forest-a novel pattern recognition method for multiclass classification in microarray data analysis. DNA and Cell Biology 23:685–694
17. Hong H, Thakkar S, Chen M, Tong W (2018) Development of decision forest models for prediction of drug-induced liver injury in humans using a large set of fda-approved drugs. Sci Rep 7(1):17311
18. Hong H, Rua D, Sakkiah S, Selvaraj C, Ge W, Tong W (2016) Consensus modeling for prediction of estrogenic activity of ingredients commonly used in sunscreen products. Int J Environ Res Public Health 13(10):E958
19. Hong H, Harvey BG, Palmese GR, Stanzione JF, Ng HW, Sakkiah S, Tong W, Sadler JM (2016) Experimental data extraction and in silico prediction of the estrogenic activity of renewable replacements for bisphenol A. Int J Environ Res Public Health 13(7):E705
20. Hong H, Shen J, Ng HW, Sakkiah S, Ye H, Ge W, Gong P, Xiao W, Tong W (2016) A rat α-fetoprotein binding activity prediction model to facilitate assessment of the endocrine disruption potential of environmental chemicals. Int J Environ Res Public Health 13(4):372
21. Ng HW, Doughty SW, Luo H, Ye H, Ge W, Tong W, Hong H (2015) Development and validation of decision forest model for estrogen receptor binding prediction of chemicals using large data sets. Chem Res Toxicol 28(12):2343–2351

22. Luo H, Ye H, Ng HW, Shi L, Tong W, Mendrick DL, Hong H (2015) Machine learning methods for predicting HLA-peptide binding activity. Bioinform Biol Insights 9(Suppl 3):21–29
23. Kavlock RJ, Ankley G, Blancato J, Breen M, Conolly R, Dix D, Houck K, Hubal E, Judson R, Rabinowitz J, Richard A, Setzer RW, Shah I, Villeneuve D, Weber E (2008) Computational toxicology–a state of the science mini review. Toxicol Sci 103(1):14–27
24. Benfenati E (2007) Predicting toxicity through computers: a changing world. Chem Cent J 1:32
25. Hong H, Xie Q, Ge W, Qian F, Fang H, Shi L, Su Z, Perkins R, Tong W (2008) Mold(2), molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. J Chem Inf Model 48(7):1337–1344
26. Devillers J, Doré JC, Guyot M, Poroikov V, Gloriozova T, Lagunin A, Filimonov D (2007) Prediction of biological activity profiles of cyanobacterial secondary metabolites. SAR QSAR Environ Res 18(7–8):629–643
27. Chen M, Bisgin H, Tong L, Hong H, Fang H, Borlak J, Tong W (2014) Toward predictive models for drug-induced liver injury in humans: are we there yet? Biomark Med 8(2):201–213
28. Hong H, Chen M, Ng HW, Tong W (2016) QSAR Models at the US FDA/NCTR. Methods Mol Biol 1425:431–459
29. Karplus M (2003) Molecular dynamics of biological macromolecules: a brief history and perspective. Biopolymers 68(3):350–358
30. Fermi E, Pasta P, Ulam S, Tsingou M. Studies of the nonlinear problems. United States: 1955 1955-05-01
31. Alder BJ, Wainwright TE (1957) Phase transition for a hard sphere system. J Chem Phys. 27:1208–1209
32. Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. General method. J Chem Phys. 31:459
33. Rahman A (1964) Correlations in the motion of atoms in liquid argon. Phys Rev 136(2A):A405–A411
34. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590
35. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103(2):227–249
36. Roccatano D, Barthel A, Zacharias M (2007) Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. Biopolymers 85(5–6):407–421
37. Sharma S, Ding F, Dokholyan NV (2007) Multiscale modeling of nucleosome dynamics. Biophys J 92(5):1457–1470
38. Stillinger FH, Rahman A (1974) Improved simulation of liquid water by molecular dynamics. J Chem Phys. 60:1545–1557
39. Clementi E, Corongiu G (1980) A theoretical study on the water structure for nucleic acids bases and base pairs in solution at T = 300 K. J Chem Phys 72(7):3979–3992
40. Parrinello M, Rahman A (1980) Crystal structure and pair potentials: a molecular-dynamics study. Phys Rev Lett 45(14):1196–1199
41. Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 72(4):2384–2393
42. Bruccoleri RE, Karplus M (1990) Conformational sampling using high-temperature molecular dynamics. Biopolymers 29(14):1847–1862
43. Hospital A, Goñi JR, Orozco M, Gelpí JL (2015) Molecular dynamics simulations: advances and applications. Adv Appl Bioinform Chem: AABC. 8:37–47
44. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718
45. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688

46. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802
47. Ponder JW, Richards FM (1987) An efficient newton-like method for molecular mechanics energy minimization of large molecules. J Comput Chem 8(7):1016–1024
48. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30(10):1545–1614
49. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19
50. Smith W, Forester TR. (1996) DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. J Mol Graph 14(3):136–141
51. Refson K (2000) Moldy: a portable molecular dynamics simulation program for serial and parallel computers. Comput Phys Commun 126(3):310–329
52. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: Proceedings of the 2006 ACM/IEEE conference on Supercomputing; Tampa, Florida. 1188544: ACM. p 84
53. MacKerell AD, Brooks CL, Nilsson L, Roux B, Won Y, Karplus M (1998) {CHARMM}: the energy function and its parameterization with an overview of the program. In: Schleyer, editor. 1: John Wiley & Sons: Chichester p 271–277
54. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117(19):5179–5197
55. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25(13):1656–1676
56. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236
57. Nelson MT, Humphrey W, Gursoy A, Dalke A, Kalé LV, Skeel RD, Schulten K (1996) NAMD: a parallel, object-oriented molecular dynamics program. Int J Supercomput Appl High Perform Computing 10(4):251–268
58. Lippert RA, Bowers KJ, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Shaw DE (2007) A common, avoidable source of error in molecular dynamics integrators. J Chem Phys 126(4):046101
59. Bowers KJ, Dror RO, Shaw DE (2006) The midpoint method for parallelization of particle simulations. J Chem Phys 124(18):184109
60. Meharenna YT, Poulos TL (2010) Using molecular dynamics to probe the structural basis for enhanced stability in thermal stable cytochromes P450. Biochemistry 49(31):6680–6686
61. Skopalik J, Anzenbacher P, Otyepka M (2008) Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. J Phys Chem B 112(27):8165–8173
62. Hendrychova T, Anzenbacherova E, Hudecek J, Skopalik J, Lange R, Hildebrandt P, Otyepka M, Anzenbacher P (1814) Flexibility of human cytochrome P450 enzymes: molecular dynamics and spectroscopy reveal important function-related variations. Biochim Biophys Acta 1:58–68
63. Lampe JN, Brandman R, Sivaramakrishnan S, de Montellano PR (2010) Two-dimensional NMR and all-atom molecular dynamics of cytochrome P450 CYP119 reveal hidden conformational substates. J Biol Chem 285(13):9594–9603
64. Park H, Lee S, Suh J (2005) Structural and dynamical basis of broad substrate specificity, catalytic mechanism, and inhibition of cytochrome P450 3A4. J Am Chem Soc 127(39):13634–13642

65. Fishelovitch D, Hazan C, Shaik S, Wolfson HJ, Nussinov R (2007) Structural dynamics of the cooperative binding of organic molecules in the human cytochrome P450 3A4. J Am Chem Soc 129(6):1602–1611
66. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE 3rd (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33(12):889–897
67. Kirchmair J, Williamson MJ, Tyzack JD, Tan L, Bond PJ, Bender A, Glen RC (2012) Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. J Chem Inf Model 52(3):617–648
68. Shu M, Hui Wen N, Michael O, Heng L, Hao Y, Weigong G, Weida T, Huixiao H (2016) Homology model and ligand binding interactions of the extracellular domain of the human α4β2 nicotinic acetylcholine receptor. J Biomed Sci Eng 09:41–100
69. Shen J, Zhang W, Fang H, Perkins R, Tong W, Hong H (2013) Homology modeling, molecular docking, and molecular dynamics simulations elucidated α-fetoprotein binding modes. BMC Bioinform 14(Suppl 14):S6–S6
70. Heitz F, Van Mau N (2002) Protein structural changes induced by their uptake at interfaces. Biochimica et Biophysica Acta (BBA)—Protein Struct Mol Enzym 1597(1):1–11
71. Hong H, Branham WS, Dial SL, Moland CL, Fang H, Shen J, Perkins R, Sheehan D, Tong W (2012) Rat α-fetoprotein binding affinities of a large set of structurally diverse chemicals elucidated the relationships between structures and binding affinities. Chem Res Toxicol 25(11):2553–2566
72. Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, Arkin IT, Jensen MØ, Xu H, Trbovic N, Friesner RA, Palmer AG, Shaw DE (2008) Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. J Phys Chem B 112(19):6155–6158
73. Lange OF, Lakomek N-A, Farès C, Schröder GF, Walter KFA, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320(5882):1471
74. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hübner CG, Kern D (2007) Intrinsic motions along an enzymatic reaction trajectory. Nature 450:838
75. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. Science 334(6055):517
76. Gibson JB, Goland AN, Milgram M, Vineyard GH (1960) Dynamics of radiation damage. physical. Review 120(4):1229–1253

Suguna Devi Sakkiah is Research Scientist in the Division of Bioinformatics and Biostatistics of FDA's National Center for Toxicological Research, Jefferson, Arkansas. She received her Ph.D. in computational chemistry from Gyeongsang National University, South Korea, and M.Sc. from Bharathiar University, Coimbatore, India. In 2010, she received a young scientist award during her Ph.D. She conducted postdoctoral researches on computer-aided drug design and structure–activity relationships studies using various modeling tools in Gwangju Institute of Science and technology, South Korea, University of California, Cedars-Sinai Medical Center, and FDA/National Center for Toxicological Research, USA. Currently, she is working on computational modeling of endocrine receptors. She developed multiple predictive models to predict chemicals binding with estrogen and androgen receptors to assist safety evaluation of FDA-regulated products. She published more than 50 articles in scientific journals.

Rebecca Kusko is SVP and Head of R&D at Immuneering Corporation. She completing her undergraduate degree at MIT in biological engineering and earning a Ph.D. in genetics and genomics from BU before joining Immuneering in 2014. At Immuneering, she has been accelerating pharma R&D using advanced algorithms. Additionally, she is Board Member of the MAQC Society, an initiative around best practices for big data analysis.

Weida Tong is Director of Division of Bioinformatics and Biostatistics at FDA's National Center for Toxicological Research (NCTR/FDA). He has served as Science Advisory Board Member for several large projects involving multiple institutes in Europe and the USA. He also holds several adjunct positions at universities in the USA and China. His division at FDA works to develop bioinformatic methodologies and standards to support FDA research and regulation and to advance regulatory science and personalized medicine. The most visible projects from his group are (1) leading the Microarray Quality Control (MAQC) consortium to develop standard analysis protocols and quality control metrics for emerging technologies to support regulatory science and precision medicine; (2) development of liver toxicity knowledge base (LTKB) for drug safety; (3) in silico drug repositioning for the enhanced treatment of rare diseases; and (4) development of the FDA bioinformatics system, ArrayTrackTM suite, to support FDA review and research on pharmacogenomics. In addition, his group also specializes in molecular modeling and QSARs with a specific interest in estrogen, androgen, and endocrine disruptor. He has published more than 230 papers and chapters.

Huixiao Hong is Chief of Bioinformatics Branch, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research (NCTR), US Food and Drug Administration (FDA), working on the scientific bases for regulatory applications of bioinformatics. Before joining the FDA, he was Manager of Bioinformatics Division of Z-Tech, an ICFI company. He was Research Scientist Position at Sumitomo Chemical Company in Japan and Visiting Scientist at National Cancer Institute at National Institutes of Health. He was also Associate Professor and Director of Laboratory of Computational Chemistry at Nanjing University in China. He is Member of steering committee of OpenTox, Member of Board Directors of US MidSouth Computational Biology and Bioinformatics Society, and in the leadership circle of US FDA modeling and simulation working group. He published more than 180 scientific papers and served as Editor-in-Chief, and Editorial Board Member for multiple peer-reviewed journals. He received his Ph.D. from Nanjing University in China in 1990 and conducted research in Leeds University in England in 1990–1992.

Part II
Applications in Regulatory Science

Chapter 11
Applicability Domain: Towards a More Formal Framework to Express the Applicability of a Model and the Confidence in Individual Predictions

Thierry Hanser, Chris Barber, Sébastien Guesné, Jean François Marchaland and Stéphane Werner

Abstract A common understanding of the concept of applicability domain (AD) is that it defines the scope in which a model can make a reliable prediction; in other words, it is the domain within which we can trust a prediction. However, in reality, the concept of confidence in a prediction is more complex and multi-faceted; the applicability of a model is only one aspect amongst others. In this chapter, we will look at these different perspectives and how existing AD methods contribute to them. We will also try to formalise a holistic approach in the context of decision-making.

Keywords Applicability domain · Decision domain · TARDIS · QSAR · Machine learning · Confidence modelling

Abbreviations

3D	Three Dimension
AD	Applicability Domain
DD	Decision Domain

T. Hanser (✉) · C. Barber · S. Guesné · J. F. Marchaland · S. Werner
Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS, UK
e-mail: Thierry.Hanser@lhasalimited.org

C. Barber
e-mail: Chris.Barber@lhasalimited.org

S. Guesné
e-mail: Sebastien.Guesne@lhasalimited.org

J. F. Marchaland
e-mail: Jean-Francois.Marchaland@lhasalimited.org

S. Werner
e-mail: Stephane.Werner@lhasalimited.org

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_11

kNN K-Nearest Neighbours
QSAR Quantitative Structure Activity Relationship
OECD Organization of Economic Co-operation and Development
PCA Principal Component Analysis

11.1 Introduction: Confidence in an Individual Prediction

Quantitative/Qualitative Structure Activity Relationship (QSAR) models can be used in different contexts ranging from early virtual screening to late safety assessment in the process of drug development. The ability to gauge the level of confidence in predictions provided by such models becomes more and more important as the drug candidates get closer to human exposure. In the early context of screening large chemical libraries in the quest for active structures, it is acceptable for a model to make a certain number of mistakes provided that the number of erroneous predictions is not too big. In such a case, the global accuracy of a model gives us a sufficient estimate of the confidence in the predictions and we expect a model with a good accuracy to lead to a useful selection across the whole compound library. The model's accuracy can be measured a priori using internal and external validation methods and provides an intrinsic sense of confidence in the predictions from a statistical standpoint; however, it does not tell us how much we can trust individual predictions. On the other hand, when it comes to human safety assessment, we are measuring the risk of adverse events induced by a specific molecule. In this case, the global accuracy of the model is not helpful, and we need a way to estimate the accuracy of an individual prediction. Whereas in the context of virtual screening, we use the accuracy of the model, in the context of human safety assessment, we need the accuracy of a prediction. Model accuracy and prediction accuracy are two very different concepts. To better illustrate the distinction, let us assume that we have built a good model providing an 83% accuracy on a challenging toxicity prediction task. If we use this model in the context of screening thousands of compounds, we can be relatively confident; indeed, it feels like going to the casino playing with strongly biased dice; we know that in the long run, we will achieve substantial gain (Fig. 11.1a). However, in the context of risk assessment, we focus on one individual prediction with potentially life-threatening consequences. In this case, the same 83% is a one in six chances of a lethal outcome (Fig. 11.1b). The two different contexts lead to two distinct ways to consider confidence in predictions.

When focusing on a specific compound, the question becomes "Can we trust this specific individual prediction?" which combines elements of assessing the legitimacy of the model, the reliability of the prediction for a specific compound and the level of uncertainty that can be tolerated in making a decision. All these aspects make perceiving the confidence in a prediction difficult to formalise. An attempt to formally describe the legitimate scope of a model has been introduced as the AD. The role of the AD is to define the boundaries within which a model can be used and provides sufficiently accurate predictions. A well-defined AD has become a key feature for

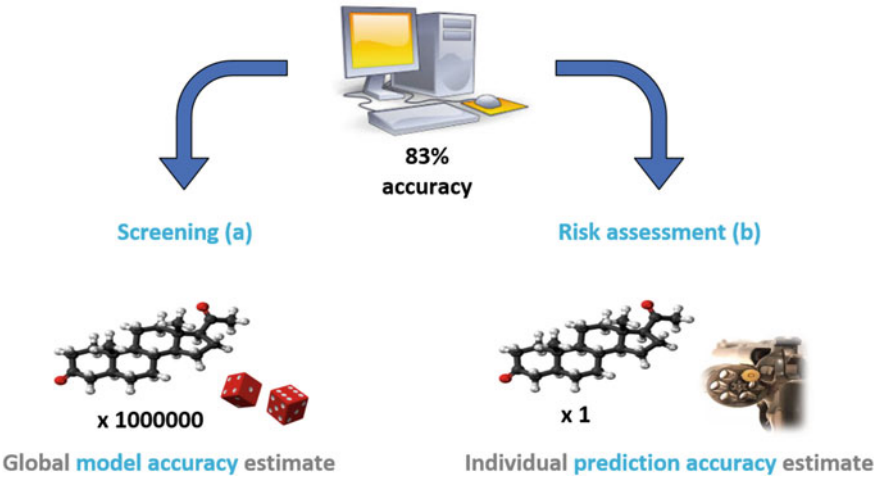

Fig. 11.1 In the context of virtual screening (a) a global accuracy of 83% provides an acceptable level of confidence across a large number of structures to select from. On the other hand, in the context of risk assessment (b), the same 83% corresponds to a one in six chances of observing a potentially lethal outcome. The confidence is therefore perceived as weaker in this latter case

in silico prediction systems, and the Organization of Economic Co-operation and Development (OECD) has included AD as part of the requirements for (Q)SAR models [1], the OECD defines AD as follows:

> Applicability Domain is the response and chemical structure space in which the model makes predictions with a given reliability.

Thanks to this formalisation effort, most modern *in silico* prediction systems feature a way to identify if a compound is part of their AD or not, and thus provide a confidence estimate at the individual prediction level. Unfortunately, we observe a strong heterogeneity in the wide spectrum of methods developed for this purpose. Existing methods are based on variable definitions of the AD and often rely on different ways to consider the problem at the source. In their well-structured review [2], M. Mathea et al. compare the main methods available and their contribution in formalising AD. Despite the important and valuable effort to tackle the AD problem [3–15], the lack of a standard definition and the variety of existing implementations, dramatically reduces the ability to assess a prediction or compare predictions from different systems.

Although the OECD definition is an important step forward to express the intended aim of the AD and formalise its scope, it remains vague about practical aspects and does not provide implementation guidelines. The OECD definition is based on several difficult concepts and formalising each of these concepts is a challenging task. For instance:

- When is it valid to use a model to make a prediction?

- Are there given structures for which the model should not be used?
- How do we define the boundaries of the chemical space?
- How do we define the reliability of a prediction?
- When is a given prediction reliable enough?
- Is the prediction conclusive, i.e. is the outcome likely?

One of the main sources of confusion is probably a natural desire to compile all these different questions into one: "Is the query compound inside or outside the applicability domain?", in other words, into a closed yes/no question. This compilation has several drawbacks; first, it is very difficult to combine information of a different nature and express the resulting concept in a single metric, secondly, the resulting AD implementation is less interpretable since it becomes very challenging, a posteriori, to disentangle the merged information. Finally, the assessment of the confidence in a prediction follows a chronology that is not captured in a single closed question. The importance of such a timeline will be discussed later. Another risk is to address only a subset of these questions which is unfortunately the case for many AD methodologies.

11.2 Decision Domain

A more holistic vision of AD becomes apparent when looking from the broader decision perspective; this approach was introduced by Hanser et al. [16]. Indeed, QSAR models are most useful if they provide sufficient support to the user to enable a confident decision to be made. If this is not the case, then the model offers little value.

From such a perspective, we can identify three well-defined concerns, each addressing a specific AD aspect along with a clear chronology that leads to a three-staged approach:

1. *Applicability*: Can the model be applied to make a prediction for my query?
2. *Reliability*: Is the resulting prediction reliable enough for the intended use case?
3. *Decidability*: Is the outcome of the prediction decisive (unequivocal)?

Together, these requirements define the scope for a confident decision based on a prediction (Fig. 11.2) called the decision domain (DD).

The decision domain can be defined as follows:

> The Decision Domain is the scope within which it is possible to make a confident decision based on a valid, reliable and decisive prediction.

By separating the three key concepts embedded in a prediction and addressing them individually, we can use the appropriate methodologies for assessing these aspects. This new perspective results in a more formal framework that supports the decision-making process. Existing AD methodologies can be mapped onto this framework, and we will use this structure to introduce these approaches at their corresponding level.

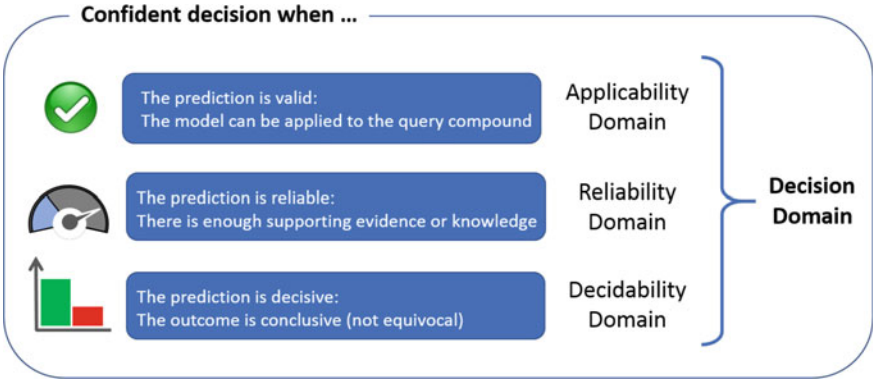

Fig. 11.2 A confident decision can be made based on a model prediction, when the prediction is valid (in the Applicability Domain), reliable (in the Reliability Domain) and decisive (in the Decidability Domain). If all three criteria are met, the prediction is said to be in the Decision Domain

11.2.1 Applicability

Prior to achieving any prediction, it is critical to check if it is valid to apply the model for the compound being predicted; in other words, we check that the model's specifications are compliant with the intended use case. These steps can be seen as a checklist with questions like:

- *Is the chemical class of the query compound supported by the model?* An example of such a specification could be "accept organic compounds but exclude polymers, proteins and inorganic molecules". There are many possible causes for the limitations of a model. For instance, for some classes of compounds, the model could be unable to compute some of the required descriptors or the model's internal chemical representation could be limited to a specific class of compound, e.g. organic molecules only due to the complex bonds in inorganic molecules, the size of the proteins structures or the repeated motifs in polymers (Fig. 11.3). Such specifications are statically attached to the model irrespective of the predicted compound or the training dataset.
- *Are the values of the descriptors of the predicted compound within the range of values the model has been trained or designed for?* Here the descriptor space is used to restrain the applicability domain of the model to ensure that the query structure does not exhibit unusual properties for which the model has not been trained. Different existing AD methods are based on this criteria amongst which the most popular are:
  - *Range Box*: This naïve approach defines the boundaries of the applicability domain as the range between the minimum and the maximum values, for each individual descriptor variable as observed in the training data. If any descriptor of the predicted compound has a value outside the interval [minimum-maximum]

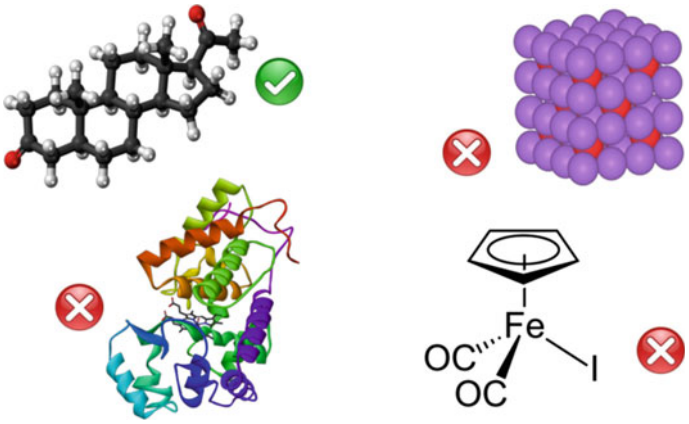

Fig. 11.3 A model may only accept some classes of query compounds. For instance, polymers, proteins and organometallic structures may be too complex or too large for the models' internal representation and should be excluded from the Applicability Domain

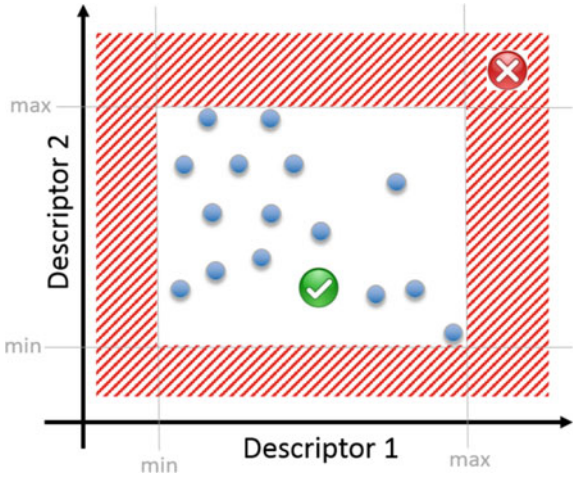

Fig. 11.4 Min-max descriptor ranges observed in the training data (blue) define the acceptable values of the query structure descriptors in the form of a hyper-box (here a rectangle in the case of two dimensions)

of the corresponding variable in the training data, then the model is not applicable, and this compound is labelled as "Outside the Applicability Domain" (Fig. 11.4). This approach bears an important drawback as it allows regions containing very few training examples inside the hyper-box defined by these ranges. As the number of descriptors (dimension of the descriptor space) use by the model increases, these regions become rapidly predominant due to the dimensionality curse [17].

– *Convex Hull*: This approach is very similar to the range box method with the difference that instead of a hyper-box defined in the descriptor space, the convex hull is the minimum surface that wraps the training data, defining a more precise

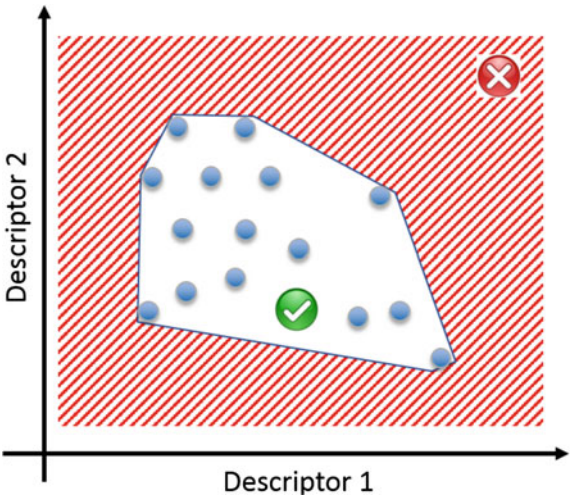

Fig. 11.5 Convex hull, is the minimum surface containing the descriptor range hyper-box. The convex hull surface defines a more complex and more precise AD than the range box approach; it allows the exclusion of "corners" of the descriptor space for which there are no training data points

AD volume. The convex hull can exclude regions that are in the range of an individual descriptor and yet distant from the extreme data points for this descriptor when looking from another descriptor's perspective; these regions can be seen as the "corners" of the hyper-box (Fig. 11.5). The convex hull method suffers from two main limitations; first finding the convex hull surface becomes computationally expensive as the number of dimensions increases, secondly, as for the hyper-box, for large numbers of descriptors and wide ranges of values, the volume contained in the convex hull becomes very large and the data points are sparsely distributed creating an effect of information "dilution" within the AD.

– *Dimension reduction*: When the number of descriptors becomes too large, it is possible to apply well-established dimension reduction techniques and define the AD within the resulting low-dimensional space (typically two or three dimensions) rather than in the direct original space. The main benefits of this approach are that the resulting AD scope focuses on the most important dimensions as identified by the reduction methodology. The lower dimension space also reduces the computational cost of AD methods like the convex hull and finally the density of information provided by the training data is higher owing to the reduced representation space. Principal Component Analysis (PCA) can be used as the dimension reduction methodology [18].

- *Has the model "seen" all the structural features present in the query compound?* Here "seen" means the features have been observed in the training data with sufficient representation. Such a criterion would not be captured by a simple descriptor range-based approach and needs to be addressed as well. Typical structural features can be based on atom-centred circular environment [19] or predefined patterns [20], alternatively pharmacophoric features can be used, which are often based on topological and 3D atom pairs/triplets [21–23].

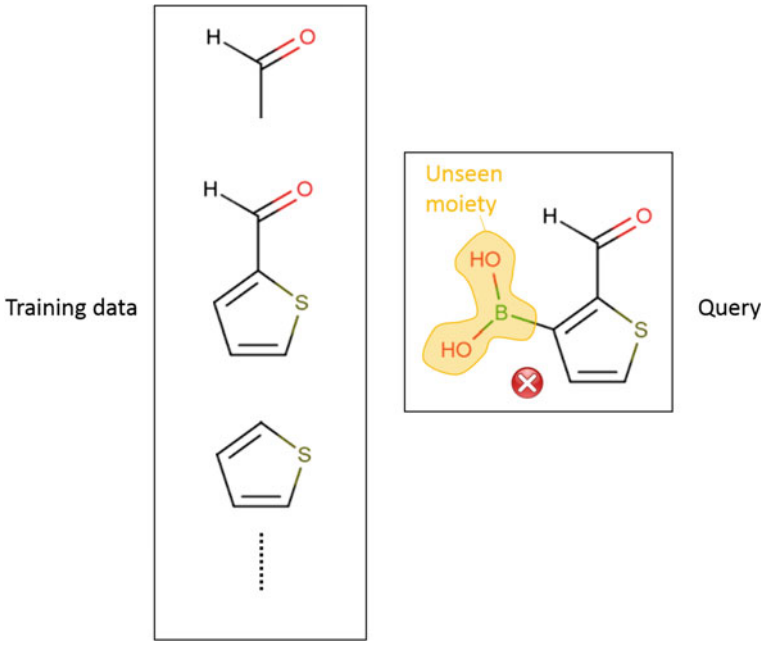

Fig. 11.6 Structural moieties present in the query compound and unseen in the model's training data or knowledge are blind spots and should invalidate the application of the model. For instance, the boronic acid group in this example is likely to cause mutagenicity. If unaware of this type of moiety, a mutagenicity *in silico* model should not be applied for the above query as the query is outside the applicability domain

The selection of such an approach should consider the relevance of structural features to the endpoint (ideally from a mechanistic perspective). Although often neglected, structural feature awareness is an important model applicability condition, especially in the context of risk assessment; a novel structural or pharmacophoric feature present in the query compound could induce toxicity. Previously unseen features constitute blind spots for the model and such models are no longer applicable (Fig. 11.6).

We have seen three representative cases in which it is not suitable to use a given model, however, there are many other possible criteria like mechanistic or metabolic prerequisites [24]. To be on the safe side, the AD of a model should consider all these criteria.

Using knowledge of the endpoint and of the model's methodology and its limitations, the model's designer is responsible for specifying the set of restrictions that define the scope in which the model can be applied, and it is the task of the prediction software to implement and to enforce these criteria.

Importantly, applicability is a model-intrinsic property, it is not dependent on the use case, and the model is either applicable for a query or it is not. In the latter case,

the query instance is labelled as "Outside the Applicability Domain". The AD of a model is a necessary but not sufficient condition for being confident in a prediction; it is also important to consider the reliability and the assertiveness of the prediction.

It is worth noting that at this stage we have not yet used the model to perform a prediction.

11.2.2 Reliability

Once established that the model is applicable, it becomes legitimate and meaningful to perform an actual prediction; we can call such a prediction a "valid" prediction. Whether the prediction is reliable or not will depend on the quantity, quality and relevance of the information available to the model. This supporting information can be in the form of training examples for statistical models, or a knowledge base in the case of expert systems. Typically, we would expect that a model based on training data containing compounds similar to the query compound, will produce a more reliable prediction than a model for which the query structure is an outlier. Indeed, it is fair to assume that the quantity and quality of information available to the model for a given query structure is provided by its nearest neighbours in the descriptor space. The number of close neighbours and the average distance of these neighbours capture the information density in the SAR region of the predicted structure. Since some data points have a degree of redundancy, it is useful to also take into account their dispersion; for instance, data points close to each other provide similar evidence and their combination holds less information than two better dispersed data points (more efficient domain coverage). Finally, the reliability of the data points themselves may impact the quality of the information; typically, experimental data obtained following good laboratory practice are likely to bear more information and less noise. All these elements contribute to the level of evidence supporting a reliable prediction (Fig. 11.7).

Different AD methods are based on this hypothesis including:

- *Distance to model*: These methods attempt to measure the distance in terms of dissimilarity between the query compound and the training dataset. There are many ways to define this distance; for instance, we could consider only the distance between the predicted structure and the closest structure in the training set (Fig. 11.8a). Alternatively, it is possible to consider the distance between the query structure and the centre of the training data using a virtual centroid point. Different similarity metrics (Tanimoto, Euclidean, Manathan, etc.) can be used to compute the actual value of the distance [25].
- *Information density*: The distance to the model is a coarse expression of the information available to the model, it does not consider internal data density variation since the training data is treated as a whole. A more fine-grained approach consists of measuring the average distance to the k closest compounds [26]. This approach can be extended into a continuous estimation of the information density using a

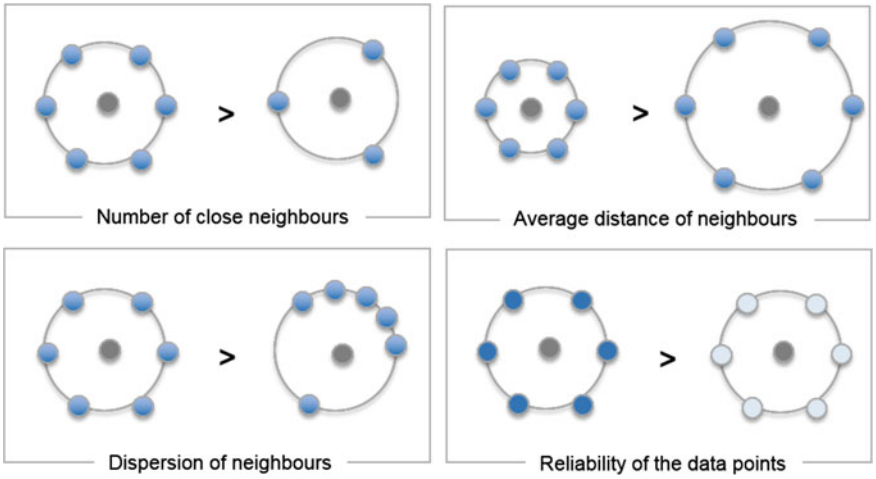

Fig. 11.7 Different aspects to take into account when considering the quality and quantity of the information available to the model to make a prediction. The left configurations of neighbours are assumed to induce a more reliable prediction

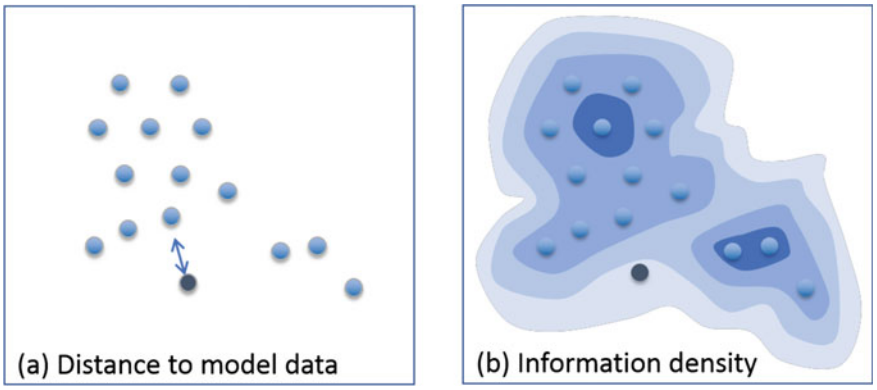

Fig. 11.8 Distance to model (**a**) and information density (**b**) can be used to evaluate the reliability of a prediction based on the amount of supporting evidence for a given query compound

kernel function based on the similarity between the query and the training data points [27]. The kernel density estimations allow to approximate a data density map in the descriptor space (Fig. 11.8b).

The role of the reliability is to inform the end user about the strength of the supporting information; the more supporting evidence available, the more reliable the prediction, and the more confidence we can have in this prediction. Reliability is usually expressed as a quantitative value and calibrated between 0 (there was no relevant information to support the prediction) and 1 (the predicted compound was known to the model). Values between 0 and 1 give an indication of the level of

reliability of the prediction and given the use case, the user can set a desired threshold below which the prediction is deemed unreliable. For instance, in the context of virtual screening, a low reliability level might be acceptable, whereas in the case of human safety assessment, the end user might accept only predictions with high reliability. Reliability is a prediction attribute as opposed to applicability which is a model property.

Reliability is not dependent on the result of the prediction. To better understand this decoupling, let us compare the model to a group of people asked a question in a specific domain and depositing their answer in a shared sealed envelope. If the people are picked randomly in a public area, before we even look in the envelope, we would consider the contents of the envelope to be less reliable than if we had chosen a group of experts in the relevant domain. We can see the outcome of the algorithm as a closed envelope, where the reliability captures how much we trust the contents of the envelope prior to opening it.

The actual desired level of reliability is use-case-dependent and can be set by the end user. The reliability metric must be calibrated and normalised. Methods based on information density can be used for this purpose. If a prediction's reliability falls below the required threshold, the prediction is said to be "Outside the Reliability Domain". Whether the result in the envelope points strongly or weakly towards a given class (classification model) or value (regression model) is not captured by the reliability; it is expressed through the likelihood assigned to these outcomes.

11.2.3 Decidability

After checking that we can use the model to make a valid prediction and that this prediction is reliable enough for the intended use case, we can finally consider the actual outcome of the model. Reusing the previous envelope analogy, we can now open this envelope to see the actual answers and check the consistency across them. Agreement amongst the supporting evidence is the main driver towards a clear conclusion of the model. If the supporting data for a given prediction converge, the model can build a more decisive prediction, whereas if the supporting information is self-contradicting, the prediction will be equivocal and therefore less decisive (Fig. 11.9).

This result is usually expressed by the model in the form of a probability distribution across the different classes or values, or in the form of discrete likelihood levels as in the case of expert systems. For instance:

- Naïve Bayes classifiers directly assign a posterior probability for each class.
- k-nearest neighbours (kNN) models can express a distribution of likelihood based on the distribution of the labels of the k-nearest neighbours and their distance to the query within the descriptor space.
- Similarly, a likelihood distribution can be derived from Random Forest predictions based on the relative vote count (at the individual tree level) for each possible class or value.

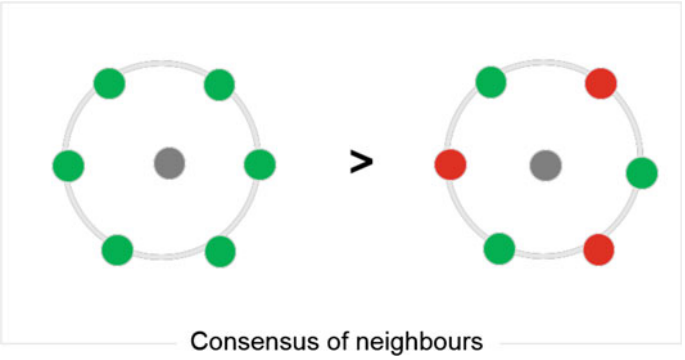

Fig. 11.9 When supporting evidence converge towards the same conclusion, the model can build a more decisive prediction, whereas if the supporting information is self-contradicting, the prediction will be equivocal and therefore less decisive. Note that, in both cases, the reliability of the prediction is the same (same number of data points, same distance and same dispersion), only the decidability is different

In the case of the Derek Nexus expert system, the assertiveness in the prediction is expressed by the reasoning engine, using one of the following likelihood levels: impossible, improbable, doubted, equivocal, plausible, probable or certain [28].

If a class (or a value) is significantly more likely than others, it suggests that the model is more decisive about the outcome; the user can have more trust in the prediction and in turn, take a more confident decision based on this prediction. On the other hand, if all the possible outcomes have similar likelihoods then the model is not decisive, the prediction is inconclusive and the user is less confident when making a decision. The decidability level can be observed by the gap between class likelihoods or, in the case of regression, by the standard deviation of the predicted value's likelihood distribution (Fig. 11.10). This principle can be seen from the distance to the decision boundary perspective; if the query compound lies far from the decision boundary of the model, then the prediction will be decisive, whereas, if the query is close to the decision boundary, the prediction is equivocal and therefore non-decisive.

The level of decidability can be expressed in the form of a real value ranging from 0 (all outcomes have equal likelihood) and 1 (the model is certain of a specific outcome). Like for the reliability, depending on the use case, the required level of decidability may vary. For instance, it is not desirable to take a risk assessment decision based on an inconclusive prediction and a high decidability threshold may therefore be set. On the other hand, in the context of virtual screening, it is possible to prioritize compounds using the relative values of decidability for given desired properties and a low level of decidability may be sufficient. The decidability level should be calibrated to actually reflect the accuracy expectation (correlation between decidability and observed accuracy); conformal predictors provide a mathematical framework to achieve this calibration and can be applied to any prediction methodologies [29, 30].

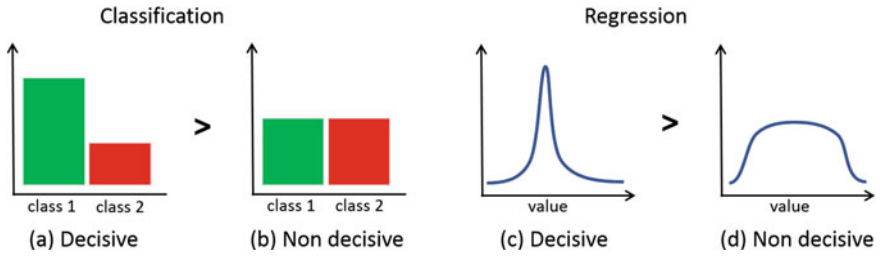

Fig. 11.10 Result of a prediction can be expressed in the form of a distribution of likelihood for each possible class in the case of classification models, or value in the case of regression estimators. If the likelihood for a given class or value stands out, then the prediction is more decisive and induces high confidence (**a**, **c**); if all outcomes have similar likelihoods, then the prediction is inconclusive and induces low confidence (**b**, **d**)

11.3 Framework

We saw that the confidence in a prediction is the result of the combination of three different aspects that can be handled separately in a cascade of assessments. The first step in this sequence is to ensure that the model is suitable for the intended task and can therefore be applied (applicability). If the model is applicable, the prediction will be valid, and a first level of confidence is met. In the second step, the quantity, quality and relevance of the supporting evidence is assessed in order to estimate how reliable the prediction is (Reliability). If the prediction is reliable enough for a given use case, it provides an additional level of confidence. Finally, if the prediction is both, valid and reliable, then the last level of confidence depends on how decisive the model is in terms of its results (Decidability). If the model expresses a decisive outcome, then the confidence in the valid, reliable and conclusive prediction is high (Fig. 11.11). Note that the chronology of these steps is important. If a model cannot be applied, it is not legitimate to do a prediction and is therefore meaningless to estimate the reliability; similarly, if the prediction is valid but not reliable, its result cannot be trusted regardless of how decisive the outcome is, since this assertiveness is itself not reliable.

11.4 Tardis

The confidence in a model's outcome as defined by the decision domain is capturing the objective and intrinsic components of a prediction. When used in the context of decision-making and especially in the context of risk assessment, users leverage their expertise to validate the QSAR predictions; the latter should therefore be interpretable and provide information regarding the evidence used by the model to build a conclusion. For instance, a kNN model may expose the k-nearest neighbours used to construct the prediction. Understanding the model's rational and accessing

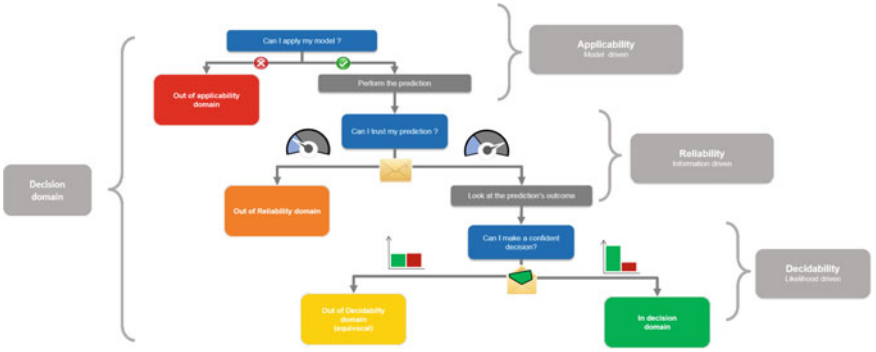

Fig. 11.11 Estimating the confidence in an individual QSAR prediction can be achieved stepwise with a gradual consolidation of the confidence level. The legitimate application of the model, the reliability of the prediction and finally the assertiveness of the outcome are assessed in a cascading process following a strict chronology

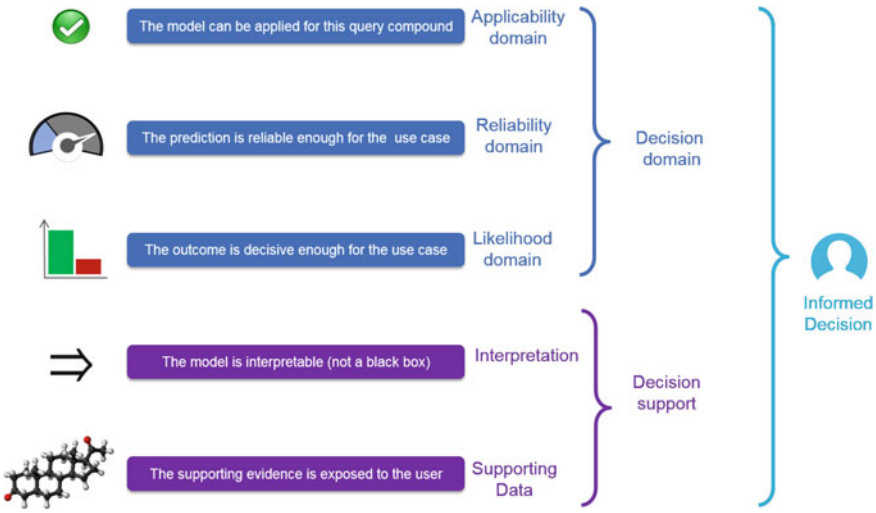

Fig. 11.12 Combining the decision domain framework along with information supporting the decision (interpretation of the prediction and access to the underlying evidence) provides a comprehensive decision framework and helps the end user take a well-educated decision

the underlying evidence help to extend the assessment of the prediction and grow the confidence in subsequent decisions. This transparency of the prediction together with the decision domain form a complete framework that helps the end user make a well-informed decision (Fig. 11.12).

The different elements of this framework are captured by the TARDIS principle (Transparency, Applicability, Reliability, Decidability, Interpretability, Support).

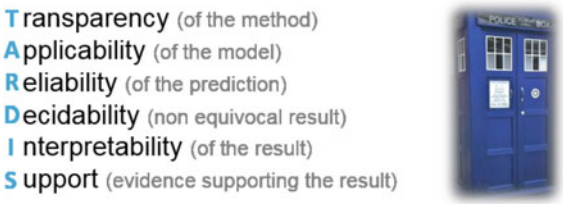

Fig. 11.13 TARDIS principle captures the criteria for comprehensive support to help the end user make a confident decision. *Note* TARDIS also refers to a Blue Police Box in a British Sci-Fi TV show called Dr. Who [31]

When fulfilling all these criteria, a prediction provides an ideal decision-making support system for the end user (Fig. 11.13).

11.5 Conclusion

Applicability Domain and confidence in predictions are difficult concepts to frame and formalise. Different aspects are involved such as the legitimacy of using a model for a given task, the reliability of the resulting prediction and how decisive the outcome is. All these criteria are important and contribute to the assessment of the confidence in the prediction. By separating these concerns and applying the appropriate validation methodologies, it is possible to define a more formal framework that defines the scope in which it is possible to make a confident decision based on a valid, reliable and decisive prediction. Today, not many AD methodologies separate these concepts nor propose a holistic approach; furthermore, current AD methodologies lack standardisation and the end user may be confused when working with different models. It would be beneficial for the QSAR community to converge towards a formal and comprehensive framework, with a normalised way of expressing the confidence in a prediction across models and applications.

In the context of decision-making, the confidence in a prediction should also be completed, when possible, with the interpretation of the outcome and a presentation of the supporting evidence. Such comprehensive support allows the user to combine their expertise with the result and the understanding of the *in silico* prediction to make a well-informed decision.

References

1. OECD (2007) Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. In: OECD series on testing and assessment, No. 69. OECD Publishing, Paris. https://doi.org/10.1787/9789264085442-en. Accessed 10 Sept 2018
2. Mathea M et al (2016) Chemoinformatic classification methods and their applicability domain. Mol Inf 35(5):160–180
3. Eriksson L et al (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 111(10):1361–1375
4. Carrió P et al (2014) Applicability domain analysis (ADAN): a robust method for assessing the reliability of drug property predictions. J Chem Inf Model 54(5):1500–1511
5. Netzeva TI et al (2005) Current status of methods for defining the applicability domain of (quantitative) structure—activity relationships. Altern Lab Anim 32(2):155–173
6. Dragos H et al (2009) Predicting the predictability: a unified approach to the applicability domain problem of QSAR models. J Chem Inf Model 49(7):1762–1776
7. Sahigara F et al (2012) Comparison of different approaches to define the applicability domain of QSAR models. Molecules 17(5):4791–4810
8. Ochi S et al (2017) Structure modification toward applicability domain of a QSAR/QSPR model considering activity/property. Mol Inf 36(12):1700076
9. Sheridan RP (2012) Three useful dimensions for domain applicability in QSAR models using random forest. J Chem Inf Model 52(3):814–823
10. Sahigara F et al (2013) Defining a novel k-nearest neighbours approach to assess the applicability domain of a QSAR model for reliable predictions. J Cheminform 5(1):27
11. Sheridan RP et al (2004) Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. J Chem Inf Comput Sci 44(6):1912–1928
12. Toplak M et al (2014) Assessment of machine learning reliability methods for quantifying the applicability domain of QSAR regression models. J Chem Inf Model 54(2):431–441
13. Sheridan RP (2015) The relative importance of domain applicability metrics for estimating prediction errors in QSAR varies with training set diversity. J Chem Inf Model 55(6):1098–1107
14. Roy K et al (2015) On a simple approach for determining applicability domain of QSAR models. Chemom Intell Lab Syst 145:22–29
15. Gadaleta D (2016) Applicability domain for QSAR models: where theory meets reality. Intern J Quant Struct Prop Relat 1(1):45–63
16. Hanser T (2016) Applicability domain: towards a more formal definition. SAR QSAR Environ Res 27(11):893–909
17. Wikipedia https://en.wikipedia.org/wiki/Curse_of_dimensionality. Accessed 10 Sept 2018
18. Nikolova-Jeliazkova N, Jaworska J (2005) An approach to determining applicability domains for QSAR group contribution models: an analysis of SRC KOWWIN. Altern Lab Anim 33(5):461–470
19. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50(5):742–754
20. MACCS structural keys (2011) Accelrys, San Diego, CA
21. Carhart RE et al (1985) Atom pairs as molecular features in structure-activity studies: definition and applications. J Chem Inf Comput Sci 25(2):64–73
22. Gobbi A, Poppinger D (1998) Genetic optimization of combinatorial libraries. Biotech Bioeng 61(1):47–54
23. Nilakantan R et al (1987) Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. J Chem Inf Comput Sci 27(2):82–85
24. Dimitrov S et al (2005) A stepwise approach for defining the applicability domain of SAR and QSAR models. J Chem Inf Model 45(4):839–849
25. Willett P et al (1998) Chemical similarity searching. J Chem Inf Comput Sci 38(6):983–996

26. Aniceto N et al (2016) A novel applicability domain technique for mapping predictive reliability across the chemical space of a QSAR: reliability-density neighbourhood. J Cheminformatics 8:69
27. Mussa HY et al (2015) The Parzen Window method: in terms of two vectors and one matrix. Pattern Recogn Lett 63:30–35
28. Philip N et al (2013) Assessing confidence in predictions made by knowledge-based systems. Toxicol Res 4(2):70–79
29. Norinder U et al (2016) Conformal prediction to define applicability domain—a case study on predicting ER and AR binding. SAR QSAR Environ Res 27(4):303–316
30. Forreryd A et al (2018) Predicting skin sensitizers with confidence—Using conformal prediction to determine applicability domain of GARD. Toxicol In Vitro 48:179–187
31. Wikipedia. https://en.wikipedia.org/wiki/TARDIS

Thierry Hanser is leading the Molecular Informatics Group at Lhasa Limited. Thierry obtained a Ph.D. in Chemo-informatics (Université de Strasbourg, France, 1992) and joined Professor Elias James Corey's group at Harvard University (1995) for a postdoctoral position followed by a second postdoctoral period with Professor Peter Johnson at the University of Leeds (1996). In 2004, Thierry started his own company, Ixelis, where he developed innovative solutions for chemical information management. Thierry joined Lhasa Limited in 2006 where he initially led the design and development of Lhasa Limited's Cheminformatics platform. His current research activity includes bridging cheminformatics and artificial intelligence/machine learning in order to design new knowledge discovery and modelling methodologies in predictive pharmacology/toxicology.

Chris Barber is the Chief Executive Officer at Lhasa Limited. He graduated with a BSc in Chemistry (Southampton, United Kingdom, 1988) and a Ph.D. in Organic Chemistry (Southampton, United Kingdom, 1991). He joined Lhasa Limited as the Director of Science in October 2011 and was appointed to the position of CEO in June 2017. Whilst Director of Science, he provided the strategic leadership and direction for Lhasa's scientific output. Chris joined Lhasa Limited from Pfizer, where he worked as a medicinal chemist for 20 years, eventually leading teams across a range of endpoints from early lead discovery through to delivering candidates for development.

Sébastien Guesné is a Senior Scientist at Lhasa Limited. He graduated with a Ph.D. in Organic and Synthetic Chemistry from the University of Exeter, United Kingdom, in 2005. Sébastien applied his laboratory and chemistry skills over eight years on medicinal chemistry projects in research assistant positions. He joined Lhasa Limited in 2013 as a Scientist where his early role was to derive knowledge from data as SAR for toxicological endpoints such as skin sensitisation, mutagenicity and hepatotoxicity. In 2016, he took over as project leader the long-running collaborative project between Lhasa and the National Institute of Health Sciences (NIHS) of Japan. In April 2018, Sébastien transitioned to a Research Chemo-informatician role in the Molecular Informatics Group. In this role, he is responsible for the design and implementation of cheminformatics' solutions, such as the use of the conformal prediction algorithm in a more formal framework of applicability domain, intended to support the development of new and existing knowledge discovery and QSAR modelling methodologies.

Jean François Marchaland is a Research Software Developer in the Molecular Informatics Group at Lhasa Limited. He graduated with a Ph.D. in Chemo-informatics from the Université de Strasbourg, France, in 1996. He joined Lhasa Limited in August 2015. His role involves the development of a new dataset registry which will be used in several of the research projects. His route to Lhasa Limited has been varied, beginning in academia working for Professor Peter Johnson at the University of Leeds in several cheminformatics postdoctoral positions before moving to

software development. Prior to joining Lhasa, Jean-Francois worked on developing software for sports databases and building systems for mobile device management and billing analysis.

Stéphane Werner is a Research Chemo-informatician in the Molecular Informatics Group at Lhasa Limited. Stéphane obtained a Ph.D. in Chemistry in 2009 from the Université de Strasbourg, France. Stéphane joined Lhasa Limited in 2007 and initially contributed to the design and development of Lhasa Limited's Cheminformatics platform. His current research activity includes the development of machine learning and QSAR internal frameworks to improve the prediction of toxicological endpoints.

Chapter 12
Application of Computational Methods for the Safety Assessment of Food Ingredients

Patra Volarath, Yu (Janet) Zang and Shruti V. Kabadi

Abstract At the Office of Food Additive Safety (OFAS) in the Center for Food Safety and Applied Nutrition at the United States Food and Drug Administration, scientists review toxicological data submitted by industry or published in scientific journals as a part of premarket safety assessments of food ingredients. OFAS also reviews relevant safety data during postmarket assessments of food ingredients as new toxicological data or exposure information become available. OFAS is committed to maintaining a high standard of science-based safety reviews and to staying abreast of novel computational approaches used by industry that could add value to improve safety assessments of food ingredients. In this chapter, we discuss some computational approaches, including quantitative structure–activity relationships, toxicokinetic modeling and simulation, and bioinformatics, as well as OFAS's in-house food ingredient knowledgebase. We describe the scientific utility of these computational approaches for improving the efficiency of the review process and reducing uncertainties in decisions about the safe use of food ingredients and highlight some challenges with their use for food ingredient safety assessments.

Keywords QSAR · SAR · TK/PBTK · Database · Bioinformatics · Food ingredients · Safety · Allergenicity · IVIVE

P. Volarath (✉) · Y. (Janet) Zang · S. V. Kabadi
FDA/CFSAN/OFAS, 5001 Campus Drive, HFS 275, College Park, MD 20740, USA
e-mail: Patra.Volarath@fda.hhs.gov

Y. (Janet) Zang
e-mail: Yu.Zang@fda.hhs.gov

S. V. Kabadi
e-mail: Shruti.Kabadi@fda.hhs.gov

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_12

Abbreviations

3D	Three dimensional
5:3 acid	5:3 fluorotelomer carboxylic acid
AA	Amino acid
AOL	AllergenOnline database
AUC	Area Under the Curve
BBDR	Biologically based dose response
CASRN	Chemical Abstract Service Registration Numbers
CERES	Chemical Evaluation and Risk Estimation System
CosIng	Cosmetic Ingredient
CR	Compound Registration
CRS-IDs	CERES compound identifiers
CXFSAN	Center for Food Safety and Applied Nutrition
DSSTox	Distributed Structure-Searchable Toxicity
E-memos	Electronic memoranda
FAO	Food and Agriculture Organization
FARM	Food Application Regulatory Management System
FARRP	Food Allergy Research and Resource Program
FCN	Food contact notification
FCS	Food contact substance
FOIA	Freedom of Information Act
FTOH	Fluorotelomer alcohol
GSIDs	General Substance Identifiers
GRAS	Generally recognized as safe
HPT	Hypothalamic–pituitary–thyroid
IgE	Immunoglobulin E
IUIS	International Union of Immunological Societies
IVIVE	In vitro to in vivo extrapolation
Km	Michaelis–Menten constant
NCBI	National Center for Biotechnology Information
OFAS	Office of Food Additive Safety
PAFA	Priority-Based Assessment of Food Additive
PBTK	Physiologically based toxicokinetic
PDF	Portable Document Format
PLETHEM	Population Lifecourse Exposure-to-Health-Effects Models
PNC	Pre-notification consultation
QSAR	Quantitative Structure Activity Relationship
RCA	Research Collaboration Agreements
SAR	Structure–Activity Relationship
SD	Structure data
SMILES	Simplified molecular-input line-entry system
STARI	Scientific Terminology and Regulatory Information

12.1 Introduction

The Office of Food Additive Safety (OFAS) in the Center for Food Safety and Applied Nutrition (CFSAN) of the United States Food and Drug Administration (US FDA) regulates safe use of food ingredients. The term "food ingredients" includes food additives and color additives used in food, substances classified as food contact substances (FCS; also known as indirect food additives), and substances classified as generally recognized as safe (GRAS) as well as substances derived from bioengineering for their intended use in food [1]. OFAS review scientists assess chemical identity, exposure information [2, 3], and toxicological data [1, 4] included in industry submissions to support the safety of food ingredients under the intended conditions of their use. OFAS review scientists also conduct independent literature searches to find any additional information relevant to the reviews and analyses. Each evaluation is performed to determine with reasonable certainty that the food ingredient is not harmful when used as intended. In some instances, a postmarket safety evaluation may be conducted for a certain food ingredient due to a change in exposure to the consumer as a result of its intended use or due to the availability of new scientific data that raises questions regarding its safety under the conditions of use. Under these circumstances, OFAS review scientists would perform an updated safety assessment based on contemporary methodology and guidelines.

As computational science advances and new data become available, integrating predictive toxicology methods into safety assessments has become an agency-wide priority. An example of this is the FDA's "Predictive Toxicology Roadmap" [5]. This roadmap identifies specific toxicological areas that could benefit from improved predictivity, such as modeling, to support decision making whenever there are data gaps. This effort shows FDA's commitment to developing and adopting new technologies, including computational toxicology. At OFAS, the review scientists are exploring the scientific utility of computational methods to assist with food ingredient safety evaluations, particularly when scientific questions may not be directly answered by experimental study data. In this chapter, we provide an overview of the underlying concepts and applications of some of these methods for evaluating the safety of food ingredients. These include:

(a) quantitative structure–activity relationship (QSAR) modeling to support the safety evaluation of those substances whose exposure is less than 150 μg/person/day and have limited and insufficient toxicity data,
(b) toxicokinetic (TK) modeling and simulation to gain insight into the internal exposure and modes of action of substances with intermediate to high toxicological potential, and
(c) bioinformatics approaches to evaluate potential allergenic effects from food proteins. Furthermore, we describe OFAS's in-house food ingredient knowledgebase, Chemical Evaluation and Risk Estimation System (CERES) that contains administrative, chemical, and safety data on a diverse set of food ingredients. Lastly, we discuss current challenges for utilizing these computational methods to their full potential for evaluating safety of food ingredients.

12.2 Computational Approaches Currently Used in Food Ingredient Safety Assessments

In this section, we describe basic principles and applications of some computational methods that are currently utilized for safety assessment of food ingredients, such as *in silico* models and databases. While validation "Standard Operating Procedures" and similar aspects of "Good Practices" are crucial for scientific accountability in safety assessments, these aspects are not covered in depth here. It is expected that practitioners will adhere to quality control and assurance procedures in performing computational modeling [6].

12.2.1 QSAR Modeling

QSAR modeling represents a subclass of *in silico* models that predicts potential toxicity of chemicals using molecular descriptors. Molecular descriptors are computer-calculated chemical and physicochemical features that describe the chemical structures and properties, respectively. Examples of descriptors commonly used in QSAR modeling are summarized in Table 12.1 [7]. Different descriptors have advantages and disadvantages, and the selection of descriptors should be based on the type of models that need to be built [8, 9]. Once appropriate descriptors are selected, an algorithm (such as linear regression or multiple linear regression) can be applied to generate a QSAR model. In general, chemicals whose activities fit the same QSAR model are assumed to have similar biological/biochemical functions and are, therefore, expected to behave through the same mechanism and exert similar toxicity. In OFAS, QSAR models have been used to support premarket safety assessments and provide guidance to industry and other stakeholders during prenotification consultations (PNCs) for future food contact substance notifications (FCNs) [10]. For example, an FCN must include all data and other information that form the basis of the determination that the FCS is safe under the intended conditions of use. Data must include primary toxicological [4] and chemical [2] information. Typically, with respect to the chemical data this includes information about residual starting material, catalysts, adjuvants, production aids, byproducts and breakdown products of the FCS. FDA encourages petitioners and notifiers to contact the agency before making a submission to discuss various issues related to a submission. Prior to submitting an FCN, industry can submit a PNC to discuss regulatory and scientific aspects of their intended submission, including the data needed to support the FCN. In response to a PNC, OFAS review scientists routinely use QSAR models to identify potential safety questions related to genetic toxicity and carcinogenicity for substances with no or limited safety data. In some cases, certain toxicological endpoints, such as developmental and reproductive toxicity, may be evaluated using QSAR models. Thus, upon review of a PNC, the FDA can recommend the types of toxicity data needed to come to a safety conclusion based on the results of the QSAR analyses. Review scientists'

Table 12.1 Examples of commonly used descriptors in QSAR [7]

Descriptors	Examples
Physicochemical descriptors	Partition coefficient (logP), acid dissociation constant, Hammett constant, and Taft steric constant
Topological descriptors	Wiener index, Zagreb group indices, Balaban *J* index, and molecular connectivity index
Structural descriptors	Chiral centers, molecular weight, rotatable bonds, and hydrogen bond donor/acceptor
Thermodynamics descriptors	ALogP, Fh2o, Foct, and Hf
Quantum chemical descriptors	Mulliken atomic charges and quantum topological molecular similarity indices
Molecular shape analysis descriptors	Difference volume (DIFFV), common overlap steric volume (COSV), and root mean square to shape reference (ShapeRMS)

recommendations to industry may also indicate the need for additional toxicity data beyond those typically recommended for compounds with exposure within a given toxicological testing tier associated with the exposure of a substance based on its intended use [2, 3, 10].

OFAS review scientists are currently exploring QSAR modeling for food ingredient safety assessments. Licenses to commercial software packages are obtained through Research Collaboration Agreements (RCA), formerly called Cooperative Research and Development Agreements [11]. Under these agreements, OFAS's food ingredient chemical and toxicity data are shared with the participating vendors. The records shared with the collaborators are the same as can be obtained through a request under the Freedom of Information Act (FOIA). The shared data are used to improve the vendors' QSAR models that, in turn, get incorporated into their software packages which become available to OFAS review scientists. Software packages under the RCAs include Leadscope Enterprise and Model Applier [12], DEREK Nexus [13], and Vitic Nexus [14], MultiCASE's CaseUltra [15], Prous Scientific [16], ACD/Labs 2012 (Window version) [17], and MN/AM ChemTunes [18]. Publicly available tools and databases, such as ChemIDplus [19], ToxNet [20], and Toxtree [21], are also used.

12.2.2 TK Modeling and Simulation

Toxicokinetics (TK) is a discipline that evaluates the disposition of a substance at toxicologic doses, its relationship with occurrence, and a time course of toxic effects [22]. Over the years, several mathematical approaches or models have been developed for examining the TK of substances and predicting their biological effects. These approaches are referred to as methods of TK modeling and simulation as they

involve developing a model to apply certain TK conditions (i.e., TK modeling) and computing the results of the applied conditions on the mathematical model (i.e., simulation). TK modeling and simulation is based on mass balance and kinetics of reactions involved in absorption, distribution, metabolism, and elimination of a substance (or its metabolites). TK modeling and simulation are used to estimate parameters that define the disposition of a substance (and its metabolites) referred to as markers of internal exposure. Based on the underlying concepts of mathematical modeling, TK modeling and simulation can be classified into two types of models: classical and physiological [23].

12.2.2.1 Classical TK Models

Classical TK models assume that the body is a system of one or more compartments, though the compartments may not exactly correspond with anatomical structures. Classical TK models are of two types:

Compartment TK models: Compartment TK models, also known as data-based models, consist of a central compartment and may have multiple peripheral compartments. A one-compartment TK model (Fig. 12.1a) assumes that the changes in concentration over time reflect proportional changes in the tissue (or blood) over the same amount of time and that the elimination follows a first-order process. In contrast, a two-compartment TK model (Fig. 12.1b) assumes that upon administration into the central compartment, the substance distributes between two compartments; however, it does not achieve instantaneous distribution or equilibration between the two compartments [24].

Compartment TK modeling has been commonly used for examining the TK of many food ingredients for years. For example, several research articles [25–29] were published in the 1970s and 1980s that examined the TK of styrene, whose polymers are approved for food contact uses. Some studies have reported increased incidences in lung tumors following chronic inhalation [30] or oral [31] exposure to styrene in rodents. Although the human relevance of the reported carcinogenic findings in rodents has been a subject of debate [32], analysis of TK data across different exposure routes generated by compartment modeling [25–29] provides important information on TK profiles of styrene after oral versus non-oral (e.g., inhalation) exposure. More recently, due to advancements in the field of TK modeling and simulation that provide opportunities of incorporating physiological parameters or biological response for predicting effects (described further under "Physiological TK models"), investigators have been slowly moving away from compartment modeling. However, compartment modeling is still often used for hazard identification, examination of TK profiles of substances (and their metabolites), and identification of data gaps necessary for planning and designing future studies needed for further evaluating the toxicological potential of a substance.

Noncompartment TK models: Noncompartment TK models assume that the estimation of TK parameters does not depend on the number of compartments. Noncompartment TK modeling is used to estimate markers of internal exposure, such as

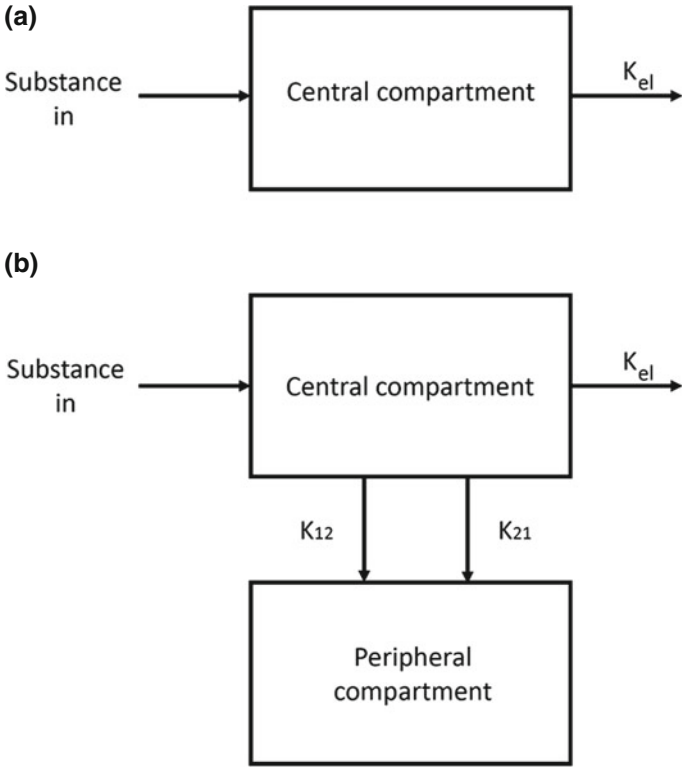

Fig. 12.1 Representation of a one-compartment model and a two-compartment model. **a** A One-Compartment Model: This model represents administration of the test substance into the central compartment, and k_{el} (h^{-1}) is the elimination rate constant. **b** A Two-Compartment Model: This model represents a central compartment (compartment 1) and a peripheral compartment (compartment 2), with rate constants, k_{12} and k_{21} representing the rate of transfer between the two compartments, C1 and C2 representing concentrations in the two compartments at time t, and k_{el} as the elimination rate constant. In simple compartment models, the rate of elimination as well as transfer between compartments is assumed to be following first-order rate of kinetics

the area under the curve (AUC) [33]. Using noncompartment TK modeling, AUC can be estimated by the trapezoidal rule which involves calculation of the area of the reported concentration versus time plot by numerically integrating areas of small sections (or trapezoids) of the curve [33].

Kabadi et al. (2018) recently performed a preliminary TK assessment of 6:2 fluorotelomer alcohol (6:2 FTOH) using noncompartment TK modeling following on a thorough review of reported data from published rat and human TK studies on the test substance [34]. FTOHs are components of high-molecular-weight polymers which are subjects of several effective FCNs for use as grease-proofing agents in food contact paper and paperboard. The results of noncompartment TK modeling of 6:2 FTOH (and its metabolites) indicated that one of the metabolites of 6:2 FTOH, 5:3

fluorotelomer carboxylic acid (5:3 A), has the potential to accumulate and contribute to the potential biopersistence of 6:2 FTOH [35]. This analysis identified 5:3 A as an important biomarker for biomonitoring studies of 6:2 FTOH [35]. Furthermore, the TK parameters estimated from the noncompartment modeling helped identify factors to consider in future toxicological studies to determine more conclusively whether, and to what extent, 6:2 FTOH biopersists in humans [35]. Like compartment modeling, noncompartment modeling not only provides valuable information on the TK profile of a substance, but also is useful in identifying data gaps necessary for planning and designing toxicity studies for conclusively determining toxicological potential.

12.2.2.2 Physiological TK Models

Physiological TK models incorporate known or hypothesized biological processes into the analysis of TK of a substance (and its metabolites). Unlike classical TK models, in physiological TK models the rate constants are not defined by data, but by physiological and anatomical components and biochemistry of a substance that influence the TK of the substance [23, 36]. These models are referred to as physiologically based TK (PBTK) models. PBTK modeling includes mathematical representation of important physicochemical and biological factors that influence the TK of substances [36–39]. The whole body is divided into tissue compartments identified as anatomical structures that are involved in the TK of a substance and defined with appropriate physiological characteristics, such as blood flow rate (ml/min or l/h), cardiac output (l/h), alveolar ventilation rate (l/h), and partition coefficients between total concentration in the tissue and freely diffusible concentration in the blood or interstitial fluid. A basic structure of a PBTK model is provided in Fig. 12.2. This model represents inhalation and oral exposure to a substance. The tissue compartments included in the PBTK model (Fig. 12.2) are lungs, fat, liver, rapidly perfused and slowly perfused tissues and the gastrointestinal tract (GIT). The rapidly perfused tissue compartment, as the name suggests, consists of tissues that are heavily perfused with blood, such as heart, kidneys, brain, etc. In contrast, the slowly perfused tissue compartment represents tissues that are less perfused with blood, such as muscle, skin, and bone.

PBTK modeling is commonly used for prediction of tissue dosimetry and internal exposure of substances and their metabolites. In addition, PBTK models can be used for extrapolation of a TK response from high dose to low dose, between different exposure routes, and across species, if appropriate physiological and TK data required to validate a model are available [36–39]. For example, over the years several PBTK models have been published on styrene [40–42]. A PBTK model for predicting the kinetic behavior of inhaled styrene in humans using data from rats was developed by Ramsey and Andersen (1984) which predicted that styrene metabolism was saturated at inhaled concentration of 200 ppm and higher in rats, mice, and humans [40]. Another PBTK model for estimating the body burden of a key metabolite of styrene, styrene-7,8-oxide (STO), was developed by Csanady et al. (1994) that described the

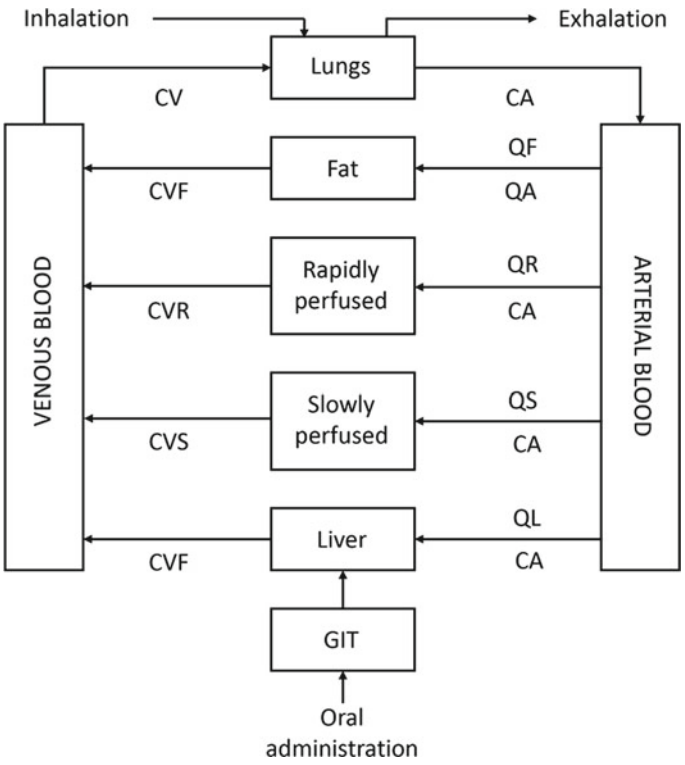

Fig. 12.2 Representation of a physiologically based pharmacokinetic (PBPK) model. The PBPK model above represents inhalation and oral exposure to a substance and consists of the lungs, fat, liver, rapidly perfused and slowly perfused tissue compartments, and the gastrointestinal tract (GIT). Certain physiological parameters are represented to describe the model: QC = cardiac output, CA = concentration in arterial blood, QF = blood flow through the fat, QR = blood flow through rapidly perfused tissues, QS = blood flow through slowly perfused tissues, QL = blood flow through the liver, CVL = concentration in the venous blood from the liver, CVS = concentration in the venous blood from the slowly perfused tissues, CVR = concentration in the venous blood from the rapidly perfused tissues, CVF = concentration in the venous blood from the fat, CV = concentration in the venous blood to the lungs

distribution and metabolism of styrene and STO in rats, mice, and humans following inhalation, intravenous, oral, and intraperitoneal administration of STO. The model represented oxidation of styrene into STO, the intracellular first-pass hydrolysis of STO catalyzed by epoxide hydrolase, and conjugation of STO with glutathione, described by an ordered sequential ping-pong mechanism between glutathione, STO, and glutathione *S*-transferase [41]. A variety of software packages are available, making PBTK modeling one of the most commonly used TK approaches in different sectors. Some commonly used software packages for PBTK modeling are Berkeley Madonna (developed at the University of California at Berkeley by Robert Macey

and George Oster), Simcyp (provided by Simcyp Limited based in Sheffield, UK), and GastroPlus (provided by Simulations Plus based in Lancaster, CA).

More recently, biological response has been incorporated into PBTK modeling, which is referred to as biologically based dose response (BBDR) modeling. BBDR modeling assumes that a toxic response is a function of the concentration of the substance in the target tissue and describes a biological response or any mechanism related to toxicity. A BBDR model was developed by McLanahan et al. (2008) for dietary iodide and the hypothalamic–pituitary–thyroid (HPT) axis in adult rats. The BDDR-HPT model [43] consisted of submodels for dietary iodide, thyroid-stimulating hormone (TSH), and thyroid hormones T_3 and T_4. The key biological processes described in the model included the influence of T_4 on TSH production, stimulation of thyroidal T_4 and T_3 production by TSH, TSH upregulation of the thyroid sodium/iodide symporter, and recycling of iodide from metabolism of thyroid hormones. This model predicted effects on the HPT axis caused by insufficient dietary iodide intake and successfully simulated perturbations in serum T_4 when compared with experimental results. This BDDR-HPT axis model [43] provides a strong basis for use in conjunction with PBTK models for thyroid-active chemicals to evaluate and predict dose-dependent HPT alterations based on modes of action [44, 45].

12.2.3 Bioinformatic Approaches for Analysis of Potential Allergenicity of Proteins

The allergenicity potential is an essential aspect of assessing the safety of food ingredients that have protein components. Since most food allergies are mediated through immunoglobulin E (IgE) which reacts with specific linear or conformational epitopes, the degree of amino acid (AA) sequence similarities between the query protein and known allergens could be used to predict the likelihood of the query protein to induce a cross-reactivity through IgE binding [46]. OFAS refers to the guideline of Codex Alimentarius (Codex) [47, 48] that describes the principles of AA sequence-based bioinformatic analysis as one of the first steps in allergenicity risk assessment. Bioinformatic tools for AA sequence alignment and similarity comparisons are commonly used in OFAS as one of the criteria to predict the potential allergenicity risk of food ingredients with protein components, including direct additives such as phycocyanins in the color additive spirulina extract, microbially derived food processing enzymes, and food proteins from genetically engineered plants.

12.2.3.1 Allergenic Protein Database

As the first step in their allergenicity risk assessments, the Codex recommends that the AA sequence of a query protein be compared against all scientifically known allergens. Thus, the predictivity of the bioinformatic methods relies on the collec-

tive information provided by the allergen databases used, which can vary by their content, organization, and accessibility [49]. Currently, the most frequently used database is the Food Allergy Research and Resource Program (FARRP) allergen database (also known as the AllergenOnline, database, AOL), maintained at the University of Nebraska-Lincoln (http://www.allergenonline.org). Candidate proteins are collected into the AOL database from the National Center for Biotechnology Information (NCBI) protein database, the World Health Organization/International Union of Immunological Societies (WHO/IUIS) allergen nomenclature database, and peer-reviewed publications on allergenicity. Each candidate protein is evaluated by an expert panel in FARRP using a set of transparent predetermined criteria to determine whether it should be classified as an allergen and then added into the AOL database [50]. The current version (Version 18B) of the AOL database includes 2089 AA sequences from 831 taxonomic-protein groups that are accepted with evidence of allergic serum IgE-binding and/or biological activity. In addition to the AOL database, there are several other online allergen databases available for bioinformatic usage, such as the UniProt (Previously referred to SwissProt) allergen database, the Structural Database of Allergenic Proteins (SDAP), Allermatch, and Allfam (a database of allergen families) [51, 52]. However, the peer-review process for each of these databases has not been sufficiently defined and most of these databases are not regularly updated, which limits their utility in bioinformatic analyses.

12.2.3.2 Criteria for Bioinformatic Analysis

Bioinformatic tools available online at AllergenOnline.org allow users to submit the AA sequence of the query protein for overall or local sequence alignments, searching for "matches" of identical short-sequence segments against the known and putative allergenic proteins in the AOL database. Regarding the criteria of possible IgE cross-reactivity, Codex recommends: (1) a threshold of 35% identity in a segment (sliding window) of 80 or more AAs, corresponding to the typical size of a protein domain containing IgE epitopes [53] and (2) identical sequence matches of six or eight contiguous AAs of the query protein with an allergenic protein in the allergen database. The FAO/WHO [54] recommends searching for a segment of six AAs, considering six AAs as the minimal length of IgE recognizable sequence. However, using only six AAs could return many false positives. Therefore, the AOL database uses the 8-AA identity search criteria. Although Codex recommends full-length sequence alignment using FASTA or BLASTP algorithms, a threshold degree of homology that suggests allergenicity between the query protein and a known allergen has not been clearly defined in the Codex guidelines. However, published analyses have determined that at least 50–70% homology in proteins would be required to show cross-reactivity [55].

While the scientific significance and sensitivity of the above-mentioned Codex criteria have been challenged [53, 56–58], they are widely accepted and used for allergenicity risk assessment by stakeholders. A negative sequence homology result indicates that a protein is unlikely to be cross-reactive to a known allergen, whereas

a positive result suggests that further examinations, such as serum screening studies, may be needed. It should be noted that the allergenicity assessment uses a weight-of-evidence framework that not only utilizes the AA sequence homology analysis, but also considers many other factors such as the source and stability of the new protein, and serum cross-activity test (if necessary) [58].

12.2.4 OFAS's Food Ingredient Knowledgebase CERES

In addition to the *in silico* methods described above, the development of databases for storage and organization of data to support the *in silico* activities and results from the safety assessment of food ingredients are also important. At OFAS, a food ingredient knowledgebase called CERES is available in-house to the review scientists to enable efficient utilization of available chemical and toxicological data while performing safety assessments of food ingredients. The CERES knowledgebase is comprised of data from several sources. One of these sources is the chemical and toxicological data submitted to OFAS in industry submissions, such as food additive petitions and FCNs, that provide information to support the authorization of new or expanded uses of food additives. These industry submissions typically contain chemical, toxicological, environmental, microbiological, and other data supporting the safety assessment of a new food additive. From these submissions, the Chemical Abstract Service Registration Numbers (CASRN), chemical names, submission type, submission number, chemicals' use types/functions, etc., are abstracted into CERES, allowing the connection of the chemical and toxicological data in CERES to the original dossiers submitted to FDA. The toxicity data included in the submissions are also carefully reviewed and abstracted into CERES by an internally trained data harvesting team. The harvested toxicity data includes in vitro assays (e.g., bacterial mutagenesis, in vitro chromosome aberration, in vitro micronucleus, in vitro mammalian mutagenesis, and cytotoxicity) and in vivo assays (e.g., acute, sub-chronic, chronic, reproductive, developmental, reproductive-developmental, multi-general reproductive, carcinogenicity, and in vivo chromosome aberration). The most important information stored in the CERES knowledgebase is the institutional knowledge generated during the safety evaluations of those submissions, which includes the reviewers' evaluations of and conclusions on the toxicity data submitted regarding the food substances and its impurities. Lastly, the knowledgebase contains the subsumed data from the Priority-Based Assessment of Food Additive (PAFA) database. PAFA is the FDA's legacy database that contained chemical and detailed toxicological (oral, acute, and genetic toxicity) data on approximately 1700 direct food additives regulated in the USA since the 1970s [59]. Other information such as the annual usage in food, estimated daily U.S. dietary intake and the Joint Committee on Food Additives Allowable Daily Intake Values, and the FDA Redbook structure categories of the chemicals are also contained within the database. A full list of the PAFA data fields can be found in Benz et al. 1991 [59]. The PAFA system became inactive in 2010,

and the legacy records were transferred into CERES in 2012 to ensure their access to OFAS review scientists.

The information contained in CERES is organized in a chemical-centric structure, where all data are connected to the chemicals through unique CERES compound identifiers (CRS-IDs). CERES users can interact with the system (Fig. 12.3a) by entering a chemical's name, identifiers, CASRN, or simplified molecular-input line-entry system (SMILES) code. Multiple chemical IDs, names, and SMILES can be batch processed in CERES. Chemical structure (and substructure) sketching is also allowed. The users can search CERES using either chemical data alone or coupled chemical-toxicity data. Once the search criteria are specified, searches can be

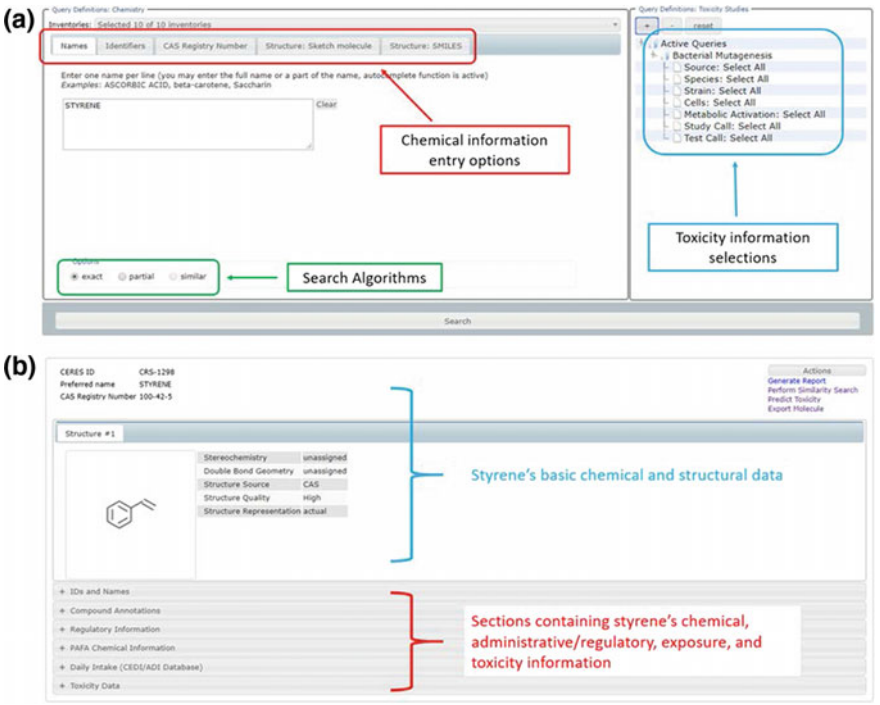

Fig. 12.3 Query page and chemical information display in CERES. **a** This view shows different options to query the system: chemical names, identifiers, CASRNs, structure drawing, and SMILES (red). The chemical search can also be coupled with toxicity data search (blue). A search algorithm (green) can also be specified accordingly (default is set at "exact" search). **b** This view displays styrene's basic chemical and structural data (blue) on top, followed by expandable sections describing the compound's chemical, regulatory/administrative, exposure, and toxicity data in details (red). The "IDs and Names" section contains a list of identifiers and chemical names/synonyms of styrene and their sources. The "Compound Annotations" section contains information on the chemical's molecular formula, material/composition type, and use types/functions. The "Regulatory Information" and "PAFA Chemical Information" sections contain the regulatory submission and PAFA data, respectively. The "Daily Intake (CEDI/ADI Database)" section contains the internally collected exposure data, cumulative estimated daily intake (CEDI) and acceptable daily intake (ADI) values. Finally, the "Toxicity Data" section contains a list of the toxicity data

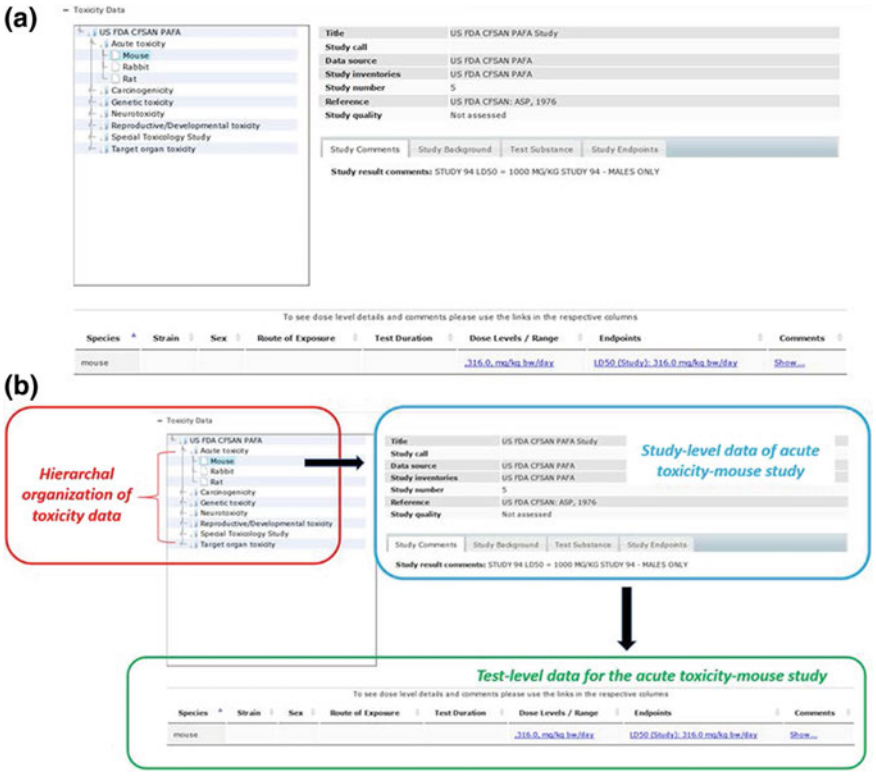

Fig. 12.4 Styrene's toxicity data organization. **a** The information display in this view is divided into three sections: an overall hierarchal organization of all toxicity data for styrene (**b**, red box), study data of a selected study (**b**, blue box), and the list of test data that support the study data (**b**, green box)

performed using exact, partial, or similarity search algorithms. The result page displays a list of the chemicals that matched the search criteria. A compound can be selected from the result page to view its chemical detail, regulatory/administrative, exposure, (e.g., cumulative estimated daily intake values), and toxicity information (Fig. 12.3b). The toxicity data are displayed in a hierarchal order (Fig. 12.4). The data are displayed at two levels: the study level and test level. The study level includes study title, overall study call, data source, study number, reference, study quality, general comments on the study, and background information, whereas the test-level information includes species, strain, sex, route of exposure, test duration, dose levels, and endpoints.

In addition to serving as the OFAS food ingredient knowledgebase, CERES is designed to support structure–activity relationship (SAR) modeling. It is, therefore, important for the CERES chemicals to cover as much chemical space as possible and consist of chemicals and chemical structures from external sources (e.g., Tox21, Cosmetic Ingredient Database, US EPA's ToxCast). For these chemicals, their original identifiers (e.g., the Distributed Structure-Searchable Toxicity General Substance

Identifiers (DSSTox GSIDs), Cosmetic Ingredient (CosIng) reference numbers) are tracked to allow data cross-referencing between CERES and the source databases. External chemicals and their structures are linked to CERES chemicals through the CRS-IDs. This setup allows cheminformatics analyses to be performed across all chemicals in CERES.

Cheminformatics capabilities are built into CERES to perform different informatics tasks using workflows. Taking the "Chemical Structure Similarity" workflow as an example, it calculates the structure similarity scores (based on the RDKit and MACCS key fingerprints) of up to five chemical structures and the results are displayed as a similarity matrix as shown in Fig. 12.5a. Another example of a workflow is data

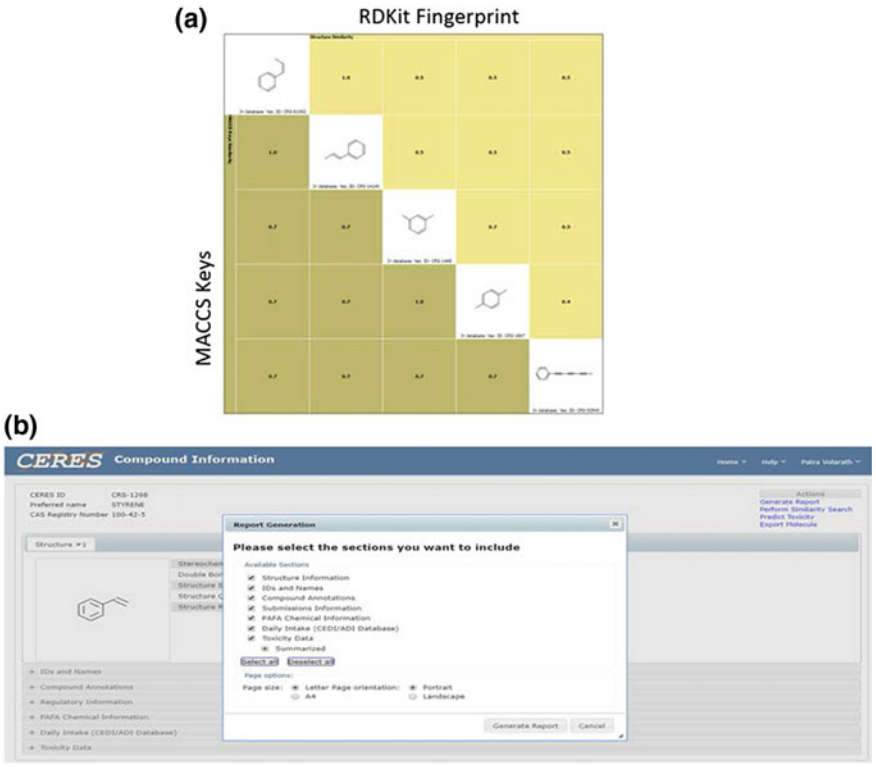

Fig. 12.5 Similarity Score Matrix and Chemical's Information Organization in CERES. **a** The similarity score matrix of five chemicals in CERES (structures are shown diagonally) is shown. The scores are calculated using the Tanimoto coefficient equation [85]. The scores in the upper right region of the structures are calculated using the RDK fingerprint, and the scores in the lower left region are calculated using the MACCS keys. The scores are ranged from 0 to 1. The closer the score is to 1, the more similar the compounds are to one another. **b** In this example, the styrene's information in CERES (structure, IDs and Names, Compound Annotation, Regulatory Information, PAFA Chemical Information, Daily Intake, and Toxicity Data) is shown and organized under different tabs. The report generation function allows the information under each of these tabs to be selected and exported into a PDF

export. This workflow exports the chemical/toxicity/regulatory data of the selected CERES chemical in a portable document format (PDF) document (Fig. 12.5b). The structure data (SD) can also be exported as an SD file that can be processed by external cheminformatics tools. Additional informatics workflows such as read-across and weight-of-evidence calculations are under consideration for development.

12.3 Current Challenges and Future Directions

Although the computational approaches described above exhibit great potential for improving the efficiency of the safety assessment process and expanding our understanding of potential toxicity of food ingredients, these methods have limitations that need to be recognized and carefully examined.

12.3.1 QSAR Modeling

QSAR modeling has many challenges [60]. One challenge related to QSAR model development for food ingredient safety evaluation is the limitation of current software to handle inorganic and metal-containing compounds, salts, and mixtures. An experiment was conducted on CFSAN's public food ingredient inventories (Table 12.2) [61–66]. The chemical records from these inventories were screened for inorganic, organometallic, metal complexes, metals, and polymers. From the 19,265 total records, 10,884 records had definable structures that were further reviewed by OFAS scientists. The structures of these 10,884 records were then programmatically processed to remove small inorganic fragments for those structures that contained multiple chemical components, neutralize charged molecules, generate 3D structures, and flag duplicates. This analysis removed 8609 structures and yielded 2275 final food ingredients whose structures could be used for QSAR analysis. This workflow is typically performed to systematically remove chemicals that cannot be efficiently processed by cheminformatics software. As shown in this example, out of the starting 19,265 food ingredient records, only 2275 (~12%) had structures that could be used in the analysis. This example indicates the need to explore and identify issues within the excluded chemical records and structures. At the same time, it illustrates current challenges in the softwares' abilities to handle polymers and inorganic/metal-containing compounds, which also need to be further explored. In the future, as more reliable data on polymers and inorganic/metal-containing compounds become available, it is expected that the structure handling capabilities in the software will improve. The software enhancements, together with the knowledge gained from examining the excluded chemical records, will enable more of the food ingredients that currently fall under these categories to be processed and used to improve QSAR models for the safety assessment of food ingredients.

Table 12.2 Publicly available U.S. FDA CFSAN food ingredient inventories (with references)

Food ingredient inventories	Abbreviation	# RECs	#Structures
U.S. Substances Added to Food Inventory (formerly called Everything Added to Food in the USA, or EAFUS) [62]	–	3968	2443
Food Contact Substances [63]	FCS	1155	391
Flavor and Extract Manufacturer's Association [61]	FEMA	2758	1742
Generally recognized as safe [64]	GRAS	572	40
Indirect Food Additives [65]	INDIRECT	3237	1790
Priority-based Assessment of Food Additives [61]	PAFA	7202	4341
Select Committee on GRAS Substance [66]	SCOGS	373	137
Total		19,265	10,884

Despite the limitations, QSAR models are useful and can potentially be integrated with other modeling approaches to enhance their prediction capabilities. For example, QSAR can be used to predict a substance's partition coefficient and metabolic parameters, such as Vmax which represents the maximum velocity of a metabolic reaction, and Km which is the Michaelis–Menten constant that represents the concentration of a substance when the reaction velocity is half of the maximum velocity for the reaction [67–69]. The predicted TK parameters can further be used for building PBTK models to simulate a substance's TK behavior [67–69]. This approach can be particularly useful for substances whose TK or toxicity data are insufficient or not available. Some commercial PBTK modeling software packages, such as GastroPlus (Simulations Plus), have incorporated this integrated QSAR-PBTK function in their platforms to predict such TK parameters based on chemical structures. However, the applications of this approach in the safety assessment of food ingredients are still limited due to insufficient chemical information on such substances and their metabolites [68].

12.3.2 TK Modeling and Simulation

Currently, there are insufficient TK datasets available on food ingredients, contaminants, or similar environmental chemicals required to validate and apply these models. This represents a challenge for developing well-validated TK models for the safety assessment of food ingredients, particularly in sensitive populations that include infants and children, pregnant and lactating women, the elderly, and people with compromised health status. One possibility to overcome this challenge is to extrapolate data from non-oral exposure studies using PBTK modeling for performing a safety assessment; however, the TK profile of substances may vary with the exposure routes. In such cases, these assessments need to be performed

on a case-by-case basis. Furthermore, the utilization of TK modeling and simulation requires a level of proficiency that is achieved by receiving appropriate training and access to resources that are necessary for applying these approaches for safety evaluation of food ingredients. Although there are vast resources available for developing TK/PBTK models, translating or transferring a model from one platform to another may be challenging due to unique applications of the models and lack of sufficient familiarity with all available platforms [70]. Several attempts have been made to develop open-source packages to support TK modeling and simulation for high-throughput TK, such as the "R-package" [71] and "Population Lifecourse Exposure-To-Health-Effects Model" (PLETHEM) [72]. A consistent data exchange and information sharing process for TK model development and quality control for safety evaluation can provide opportunities for exporting the TK models from one platform to another, thereby improving the accessibility of the developed TK models [70].

In recent years, additional *in silico* approaches based on TK have been developed by integration with other computational methods to facilitate advancement in the field of food ingredient safety assessment. One example is the integration of QSAR with PBTK which has been described in Sect. 12.3.1 of this chapter. Another example of such an integration is the use of in vitro to in vivo extrapolation (IVIVE) for TK modeling that allows utilization of data from in vitro systems for examining individual processes that could be integrated to determine effects in an intact organism [73]. If it is assumed that the toxic response of a substance is a function of its concentration in the target tissue, in vitro systems can be used to estimate parameters that can be further extrapolated to in vivo systems. Some commonly used in vitro systems for examining pharmacokinetic processes include isolated perfused liver, tissue slices, hepatocytes, subcellular fractions (microsomes and cytosol), and recombinantly expressed enzymes [73]. Over the years, several methods of performing IVIVE based on the use of different scaling factors have been proposed [73–75]. One of the most common applications of IVIVE for safety assessment is its integration with PBTK modeling [76–80]. IVIVE can be used to estimate physiological and pharmacokinetic parameters, such as metabolic rate parameters (e.g., Vmax and Km values for Michaelis–Menten kinetics) that can further be included in PBTK model development for safety assessment [76–78].

12.3.3 Bioinformatic Approaches for Allergenicity Assessment

A limitation of bioinformatic methods is that they are unable to predict de novo food sensitization as they rely on existing known allergens, IgE epitopes, or even sequence motifs leading to non-IgE mediated food allergy (such as gluten sensitivity). In addition, the AA sequence alignment only helps to identify linear epitopes but is not very useful in identifying conformational epitopes. Although efforts have been made

to improve the predictivity of allergenicity from 3D structures [52], more research is needed to identify and update the allergenic structural motifs and algorithms need to be developed to predict structural similarity.

As mentioned previously, there is no specific guidance to determine the likelihood of cross-reactivity based on full-length sequence homology. It has been proposed in some recent publications [81, 82] to use the E-value in FASTA or PLASTP alignment, which represents the probability that the alignment might occur by chance, to determine the degree of sequence homology. The developer of FASTA software stated that an E-value of less than 10^{-6} indicated a high certainty of homology [83]. However, because the E-value depends on the size of the database, there is currently no clear regulatory guideline for using a threshold E-value in allergenicity assessment. Nevertheless, some scientists believe that using the E-value from full-length sequence homology has a stronger scientific basis compared to the 80 AA sliding window [53, 84]. Therefore, additional studies are needed to validate and standardize the application of E-value in assessing protein allergenicity.

12.3.4 CERES Knowledgebase Expansion

A challenge in expanding the CERES knowledgebase is integrating data with those from other sources that have different data structures (interoperability). Currently, the chemical information in CERES, the OFAS's submission repository system called Food Application Regulatory management (FARM) system, and the CFSAN's Scientific Terminology and Regulatory Information (STARI) system is being integrated to create a centralized chemical information system in CFSAN, as well as to enable efficient data abstractions from FARM and STARI into the CERES's knowledgebase. This data synchronization task is difficult as the three systems have different data architectures. A temporary data model has been created to mitigate these challenges and enforce data standardizations.

Another challenge in the knowledgebase expansion is data quality. At present, the chemical, toxicity, and administrative data are required to go through external quality control processes before being parsed into CERES. This mechanism can introduce errors and misinterpretations as the data are manually transferred from one process to another. Several data entry tools are being developed to address these issues. These tools are intended to capture data directly from OFAS scientists within their standard workflow process, thus minimizing data transfer errors and enforcing data standardizations. An example of such a tool is electronic memoranda (e-memos). E-memos are Web forms that use the format of the official OFAS chemistry and toxicology memoranda that are written by the review scientists to summarize the technical reviews of food ingredients. A typical chemistry review memorandum contains descriptions of the chemical identity (chemical names and CASRN), intended use of the substance, method of manufacture, and dietary intake values for the substance, along with the chemistry reviewer's comments/recommendations on the proposed use based on a review of the chemistry information. A typical toxicology memorandum primarily

includes a summary of the toxicological data submitted to support the estimated exposure level of the additive under its intended condition of use and the toxicologist reviewer's conclusions/recommendations on the proposed use of the substance based on a review of the toxicology information. The e-memos allow the chemical, study-level toxicity data, and administrative data to be automatically captured into CERES. Other tools that are under development are toxicity data entry (TDE) and compound registration (CR) tools. The TDE tool is designed to allow the data harvest team to directly enter the test-level toxicity data from the submissions that support the study-level data collected from the e-memo. Similarly, the CR tools will allow OFAS chemical registrars to directly enter and correct the food ingredients' chemical information in the knowledgebase.

12.4 Conclusions

OFAS review scientists evaluate available information about food ingredients included in industry premarket submissions as well as data available in the public domain or agency files to determine if there is reasonable certainty that the substances are not harmful under their intended conditions of use. With recent advancements in toxicological testing methods and computational science, OFAS review scientists are expanding the use of *in silico* methods (SAR/QSAR, TK/PBTK, and bioinformatics) as additional tools for evaluating the safety of food substances. In addition, in-house tools (CERES and others) have been developed not only to organize and store the institutional data, but also to provide modern computational capabilities that may allow OFAS scientists to fill data gaps using existing information. Although the *in silico* methods described in this chapter have been well-utilized in the field of clinical drug development, they are not as routinely used for assessing safety of food ingredients. Considering that OFAS review scientists evaluate the safety of diverse classes of food ingredients that cover a different chemical space than drugs, it is important to understand the limitations of utilizing these computational methods for food ingredient safety assessment. Nevertheless, efforts have been made to overcome some of the challenges related to different computational approaches used for their safety assessment. Lastly, it is important to emphasize that there is a need for effective communication and collaboration among scientists from all sectors: government, industry, and academia, who are interested in the development and application of computational methods for supporting an efficient food ingredient safety assessment process.

Acknowledgements The authors would like to acknowledge the following peer reviewers from the FDA for their intellectual contribution to the chapter: Dr. Michael Adams (OFAS, CFSAN), Dr. Kirk Arvidson (OFAS, CFSAN), Dr. Jason Aungst (OFAS, CFSAN), Dr. Omari Bandele (OFAS, CFSAN), Dr. Supratim Choudhuri (OFAS, CFSAN), Dr. Mary Ditto (OFAS, CFSAN), Dr. Jeffrey Fisher (OR, NCTR), Dr. Suzanne Fitzpatrick (OCD, CFSAN), Celeste Johnston (OFAS, CFSAN), Dr. Antonia Mattia (OFAS, CFSAN), Dr. Geoffrey Patton (OFAS, CFSAN), Dr. Catherine Whiteside (OFAS, CFSAN), and Andrew Zajac (OFAS, CFSAN).

Disclaimer The views and data interpretations expressed in this chapter represent that of the authors and not necessarily of the US FDA. The use of trade names, company names, product names, etc., is for example/illustrative purposes only and should not be construed as endorsements by the FDA or the authors.

References

1. CFSAN-FDA. Redbook: guidance for industry and other stakeholders. Toxicological principles for the safety assessment of food ingredients, College Park, MD: CFSAN, US FDA, DHHS, 1993 Revised July 2007. Report No. https://www.fda.gov/downloads/food/guidanceregulation/ucm222779.pdf
2. FDA (2007) Guidance for industry: preparation of premarket submissions for food contact substances (chemistry recommendations), College Park, MD. https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm081818.htm
3. FDA (2009) Guidance for industry: recommendations for submission of chemical and technological data for direct food additive petitions, College Park, MD. https://www.fda.gov/Food/GuidanceRegulation/ucm124917.htm
4. FDA (2002) Guidance for industry: preparation of food contact notifications for food contact substances (toxicology recommendations). DFCN/OFAS/CFSAN, College Park, MD. https://www.fda.gov/food/guidanceregulation/guidancedocumentsregulatoryinformation/ingredientsadditivesgraspackaging/ucm081825.htm
5. FDA (2017) FDA's predictive toxicology roadmap. https://www.fda.gov/downloads/scienceresearch/specialtopics/regulatoryscience/ucm587831.pdf
6. Workgroup EM, Marshall SF, Burghaus R, Cosson V, Cheung SY, Chenel M, DellaPasqua O, Frey N, Hamren B, Harnisch L, Ivanow F, Kerbusch T, Lippert J, Milligan PA, Rohou S, Staab A, Steimer JL, Tornoe C, Visser SA (2016) Good practices in model-informed drug discovery and development: practice, application, and documentation. CPT Pharmacometrics Syst Pharmacol 5(3):93–122
7. Roy K, Kar S, Das RN (2015) Chemical information and descriptors. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Academic Press, Tokyo
8. Danishuddin, Khan AU (2016) Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discov Today 21(8):1291–1302
9. Parthasarathi R, Dhawan A (2018) In silico approaches for predictive toxicology. In: Dhawan AK, Kwon S (eds) In vitro toxicology. Academic Press, pp 91–109
10. Arvidson KB, Chanderbhan R, Muldoon-Jacobs K, Mayer J, Ogungbesan A (2010) Regulatory use of computational toxicology tools and databases at the United States Food and Drug Administration's Office of Food Additive Safety. Expert Opin Drug Metab Toxicol 6(7):793–796
11. FDA (2018) FDA Cooperative Research and Development Agreements (CRADAs). https://www.fda.gov/ScienceResearch/CollaborativeOpportunities/CooperativeResearchandDevelopmentAgreementsCRADAs/ucm122820.htm (cited 2018 06/06)
12. Leadscope (2018) Leadscope: enterprise and model applier. http://www.leadscope.com/ (cited 2018 06/06)
13. Lhasa Limited (2018) Lhasa Limited: Derek Nexus. https://www.lhasalimited.org/products/derek-nexus.htm (cited 2018 06/06)
14. Lhasa Limited (2018) Lhasa Limited: Vitic Nexus. https://www.lhasalimited.org/products/vitic-nexus.htm (cited 2018 06/06)
15. MultiCASE (2017) MultiCASE. http://www.multicase.com/ (04/07/2018)
16. Valencia A, Prous J, Mora O, Sadrieh N, Valerio LG Jr (2013) A novel QSAR model of Salmonella mutagenicity and its application in the safety assessment of drug impurities. Toxicol Appl Pharmacol 273(3):427–434

17. ACD/Labs (2018) ACD/labs. https://www.acdlabs.com/support/version/ (07/08/2018)
18. MNAM (2018) ChemTunes: https://www.mn-am.com/products/chemtunes (cited 04/07/2018)
19. NLM (2018) ChemIDplus. https://chem.nlm.nih.gov/chemidplus/ (cited 2018 06/06)
20. NLM (2018) TOXNET. https://toxnet.nlm.nih.gov/ (cited 2018 06/06)
21. Ideaconsult Ltd (2018) Toxtree. http://toxtree.sourceforge.net/ (cited 2018 06/06)
22. Welling PG (1995) Differences between pharmacokinetics and toxicokinetics. Toxicol Pathol 23(2):143–147
23. Shen DD (2013) Toxicokinetics. In: Klaassen CD (ed) Cassarett and Doull's toxicology: the basic science of poisons (8th edn). McGraw-Hill Professional Publishing
24. Dhillon SG, Gill K (2006) Basic pharmacokinetics. In: Dhillon S (ed) Clinical pharmacokinetics (1st edn). Pharmaceutical Press
25. Withey JR (1978) The toxicology of styrene monomer and its pharmacokinetics and distribution in the rat. Scand J Work Environ Health 4(Suppl 2):31–40
26. Withey JR, Collins PG (1977) Pharmacokinetics and distribution of styrene monomer in rats after intravenous administration. J Toxicol Environ Health 3(5–6):1011–1020
27. Withey JR, Collins PG (1978) Styrene monomer in foods a limited Canadian Survey. Bull Environ Contam Toxicol 19(1):86–94
28. Withey JR, Collins PG (1979) The distribution and pharmacokinetics of styrene monomer in rats by the pulmonary route. J Environ Pathol Toxicol 2(6):1329–1342
29. Withey JR, Karpinski K (1985) Fetal distribution of styrene in rats after vapor phase exposures. Biol Res Pregnancy Perinatol 6(2):59–64
30. Cruzan G, Cushman JR, Andrews LS, Granville GC, Johnson KA, Bevan C, Hardy CJ, Coombs DW, Mullins PA, Brown WR (2001) Chronic toxicity/oncogenicity study of styrene in CD-1 mice by inhalation exposure for 104 weeks. J Appl Toxicol 21(3):185–198
31. NCI. Bioassay of styrene for possible carcinogenicity. NCI, 1979 NCI-CG-TR-185
32. Cruzan G, Bus JS, Andersen ME, Carlson GP, Banton MI, Sarang SS, Waites R (2018) Based on an analysis of mode of action, styrene-induced mouse lung tumors are not a human cancer concern. Regul Toxicol Pharmacol 95:17–28
33. Rosenbaum SE (2016) Introduction to noncompartmental analysis. In: Rosenbaum SE (ed) Basic pharmacokinetics and pharmacodynamics: an integrated textbook and computer simulations (2nd edn). Wiley, Hoboken
34. Russell MH, Himmelstein MW, Buck RC (2015) Inhalation and oral toxicokinetics of 6:2 FTOH and its metabolites in mammals. Chemosphere 120:328–335
35. Kabadi SV, Fisher J, Aungst J, Rice P (2018) Internal exposure-based pharmacokinetic evaluation of potential for biopersistence of 6:2 fluorotelomer alcohol (FTOH) and its metabolites. Food Chem Toxicol 112:375–382
36. Yang RS, Dennison JE, Andersen ME, Ou YC, Liao KH, Reisfeld B (2004) Physiologically based pharmacokinetic and pharmacodynamic modeling. In: Holland EC (ed) Mouse models of human cancer. Wiley, pp 391–405
37. Andersen ME, Krishnan K (1994) Physiologically based pharmacokinetics and cancer risk assessment. Environ Health Perspect 102(Suppl 1):103–108
38. Clewell HJ 3rd, Andersen ME (1994) Physiologically-based pharmacokinetic modeling and bioactivation of xenobiotics. Toxicol Ind Health 10(1–2):1–24
39. Krishnan K, Clewell HJ 3rd, Andersen ME (1994) Physiologically based pharmacokinetic analyses of simple mixtures. Environ Health Perspect 102(Suppl 9):151–155
40. Ramsey JC, Andersen ME (1984) A physiologically based description of the inhalation pharmacokinetics of styrene in rats and humans. Toxicol Appl Pharmacol 73(1):159–175
41. Csanady GA, Mendrala AL, Nolan RJ, Filser JG (1994) A physiologic pharmacokinetic model for styrene and styrene-7,8-oxide in mouse, rat and man. Arch Toxicol 68(3):143–157
42. Sarangapani R, Teeguarden JG, Cruzan G, Clewell HJ, Andersen ME (2002) Physiologically based pharmacokinetic modeling of styrene and styrene oxide respiratory-tract dosimetry in rodents and humans. Inhal Toxicol 14(8):789–834
43. McLanahan ED, Andersen ME, Fisher JW (2008) A biologically based dose-response model for dietary iodide and the hypothalamic-pituitary-thyroid axis in the adult rat: evaluation of iodide deficiency. Toxicol Sci 102(2):241–253

44. McLanahan ED, White P, Flowers L, Schlosser PM (2014) The use of PBPK models to inform human health risk assessment: case study on perchlorate and radioiodide human lifestage models. Risk Anal 34(2):356–366
45. Schlosser PM (2016) Revision of the affinity constant for perchlorate binding to the sodium-iodide symporter based on in vitro and human in vivo data. J Appl Toxicol 36(12):1531–1535
46. Aalberse RC, Stadler BM (2006) In silico predictability of allergenicity: from amino acid sequence via 3-D structure to allergenicity. Mol Nutr Food Res 50(7):625–627
47. Commission CA. Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants. Annex on the Assessment of Possible Allergenicity. 2003 CAC/GL 45-2003
48. Commission CA (2009) Foods derived from modern biotechnology. Italy, Rome
49. Gendel SM (2009) Allergen databases and allergen semantics. Regul Toxicol Pharmacol 54(3 Suppl):S7–S10
50. Goodman RE, Ebisawa M, Ferreira F, Sampson HA, van Ree R, Vieths S, Baumert JL, Bohle B, Lalithambika S, Wise J, Taylor SL (2016) AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. Mol Nutr Food Res 60(5):1183–1198
51. Choudhuri S (2004) Bioinformatics for beginners: genes, genomes, molecular evolution, databases and analytical tools, London
52. Remington B, Broekman HCH, Blom WM, Capt A, Crevel RWR, Dimitrov I, Faeste CK, Fernandez-Canton R, Giavi S, Houben GF, Glenn KC, Madsen CB, Kruizinga AK, Constable A (2018) Approaches to assess IgE mediated allergy risks (sensitization and cross-reactivity) from new or modified dietary proteins. Food Chem Toxicol 112:97–107
53. Herman RA, Song P, Thirumalaiswamysekhar A (2009) Value of eight-amino-acid matches in predicting the allergenicity status of proteins: an empirical bioinformatic investigation. Clin Mol Allergy 7:9
54. FAO/WHO (2001) Evaluation of allergenicity of genetically modified foods
55. Aalberse RC (2000) Structural biology of allergens. J Allergy Clin Immunol 106(2):228–238
56. Herman RA, Song P, Kumpatla S (2015) Percent amino-acid identity thresholds are not necessarily conservative for predicting allergenic cross-reactivity. Food Chem Toxicol 81:141–142
57. Cressman RF, Ladics G (2009) Further evaluation of the utility of "sliding window" FASTA in predicting cross-reactivity with allergenic proteins. Regul Toxicol Pharmacol 54(3 Suppl):S20–S25
58. Ladics GS (2008) Current codex guidelines for assessment of potential protein allergenicity. Food Chem Toxicol 46(Suppl 10):S20–S23
59. Benz RD, Irausquin H (1991) Priority-based assessment of food additives database of the U.S. Food and Drug Administration Center for Food Safety and Applied Nutrition. Environ Health Perspect 96:85–89
60. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010
61. Arvidson K, Rathman J, Volarath P, Mostrag A, Tarkhov A, Bienfait B, Vitcheva V, Yang C (eds) Evaluation of the chemical inventories in the US FDA's Office of Food Additive Safety for human health endpoints using a toxicity prediction system. In: 55th Society of toxicology annual meeting 2016, New Orleans, LA
62. FDA (2017) U.S. FDA Substances Added to Food Inventory (formerly EAFUS). https://www.accessdata.fda.gov/scripts/fdcc/?set=FoodSubstances
63. FDA (2017) U.S. FDA inventory of effective food contact substance (FCS) notifications. https://www.fda.gov/food/ingredientspackaginglabeling/packagingfcs/notifications/default.htm
64. FDA (2018) U.S. FDA inventory of GRAS notices. https://www.accessdata.fda.gov/scripts/fdcc/?set=GRASNotices
65. FDA (2018) U.S. FDA List of Indirect Additives used in Food Contact substances. https://www.fda.gov/food/ingredientspackaginglabeling/packagingfcs/indirectadditives/default.htm

66. FDA (2018) U.S. FDA GRAS Substances (SCOGS) Database. https://www.fda.gov/food/ingredientspackaginglabeling/gras/scogs/default.htm
67. Price K, Krishnan K (2011) An integrated QSAR-PBPK modelling approach for predicting the inhalation toxicokinetics of mixtures of volatile organic chemicals in the rat. SAR QSAR Environ Res 22(1–2):107–128
68. Peyret T, Krishnan K (2011) QSARs for PBPK modelling of environmental contaminants. SAR QSAR Environ Res 22(1–2):129–169
69. Peyret T, Poulin P, Krishnan K (2010) A unified algorithm for predicting partition coefficients for PBPK modeling of drugs and environmental chemicals. Toxicol Appl Pharmacol 249(3):197–207
70. Tan YM, Worley RR, Leonard JA, Fisher JW (2018) Challenges associated with applying physiologically based pharmacokinetic modeling for public health decision-making. Toxicol Sci 162(2):341–348
71. Pearce R, Woodrow Setzer R, Strope C, Sipes N, Wambaugh J (2017) HTTK: R package for high-throughput toxicokinetics. J Stat Softw 79(4):1–26
72. ScitoVation. PLETHEM 2017 (cited 2017). Available from http://www.scitovation.com/plethem.html
73. Bale AS, Kenyon E, Flynn TJ, Lipscomb JC, Mendrick DL, Hartung T, Patton GW (2014) Correlating in vitro data to in vivo findings for risk assessment. ALTEX 31(1):79–90
74. Barter ZE, Bayliss MK, Beaune PH, Boobis AR, Carlile DJ, Edwards RJ, Houston JB, Lake BG, Lipscomb JC, Pelkonen OR, Tucker GT, Rostami-Hodjegan A (2007) Scaling factors for the extrapolation of in vivo metabolic drug clearance from in vitro data: reaching a consensus on values of human microsomal protein and hepatocellularity per gram of liver. Curr Drug Metab 8(1):33–45
75. Lipscomb JC, Poet TS (2008) In vitro measurements of metabolism for application in pharmacokinetic modeling. Pharmacol Ther 118(1):82–103
76. Sweeney LM, Himmelstein MW, Gargas ML (2001) Development of a preliminary physiologically based toxicokinetic (PBTK) model for 1,3-butadiene risk assessment. Chem Biol Interact 135–136:303–322
77. Johanson G, Filser JG (1996) PBPK model for butadiene metabolism to epoxides: quantitative species differences in metabolism. Toxicology 113(1–3):40–47
78. Kohn MC, Melnick RL (2000) The privileged access model of 1,3-butadiene disposition. Environ Health Perspect 108(Suppl 5):911–917
79. Martin SA, McLanahan ED, Bushnell PJ, Hunter ES 3rd, El-Masri H (2015) Species extrapolation of life-stage physiologically-based pharmacokinetic (PBPK) models to investigate the developmental toxicology of ethanol using in vitro to in vivo (IVIVE) methods. Toxicol Sci 143(2):512–535
80. Campbell J, Van Landingham C, Crowell S, Gentry R, Kaden D, Fiebelkorn S, Loccisano A, Clewell H (2015) A preliminary regional PBPK model of lung metabolism for improving species dependent descriptions of 1,3-butadiene and its metabolites. Chem Biol Interact 238:102–110
81. Mirsky HP, Cressman RF Jr, Ladics GS (2013) Comparative assessment of multiple criteria for the in silico prediction of cross-reactivity of proteins to known allergens. Regul Toxicol Pharmacol 67(2):232–239
82. Song P, Herman RA, Kumpatla S (2014) Evaluation of global sequence comparison and one-to-one FASTA local alignment in regulatory allergenicity assessment of transgenic proteins in food crops. Food Chem Toxicol 71:142–148
83. Pearson WR (2016) Finding protein and nucleotide similarities with FASTA. Curr Protoc Bioinform 53(1):3–9
84. Silvanovich A, Bannon G, McClain S (2009) The use of E-scores to determine the quality of protein alignments. Regul Toxicol Pharmacol 54(3 Suppl):S26–S31
85. Willett PB, Barnard JM, Downs GM (1998) Chemical similarity searching. J Chem Inf Comput Sci 38(6):983–996

Patra Volarath is Lead Cheminformatics Scientist in the Office of Food Additive Safety (OFAS) at the Center for Food Safety and Applied Nutrition (CFSAN) at the United States Food and Drug Administration. She received her Ph.D. in Biophysical Chemistry (2008) and M.S. in Molecular Genetics and Biochemistry (2002) from Georgia State University. She joined the National Center for Computational Toxicology at the U.S. Environmental Protection Agency as a Postdoctoral Fellow in 2010, before joining OFAS in 2013 to work on the FDA's Chemical Evaluation and Risk Estimation System (CERES) project.

Yu (Janet) Zang is Board-Certified Toxicologist at United States Food and Drug Administration (FDA) Center for Food Safety and Applied Nutrition (CFSAN). As a toxicology reviewer, she conducts pre-market safety review of FDA-regulated food substances, including direct and indirect food additives, color additives, and other food ingredients. She received her B.S and M.S. degrees in Environmental Biology from Nanjing University, China. She obtained her Ph.D. in Pharmacology and Toxicology from University of Louisville School of Medicine. Before she joined FDA in 2009, she completed her toxicology postdoctoral training at Johns Hopkins Bloomberg School of Public Health. She has served as the toxicological and risk assessment resource person on many national and international food safety working groups convened by NIH, NTP, OECD, FAO, and WHO. She is selected to the current (2016–2020) Joint FAO/WHO Expert Committee on Food Additives (JECFA) Roster of Toxicological and Epidemiological Experts and has authored a number of JECFA monographs. She has also served in several leadership committees within the Society of Toxicology (SOT), including the Carcinogenesis Specialty Section and Food Safety Specialty Section.

Shruti V. Kabadi is Pharmacologist at the United States Food and Drug Administration (US FDA). She works as a toxicology reviewer in the Office of Food Additive Safety (OFAS) at the Center for Food Safety and Applied Nutrition (CFSAN) where she performs primary toxicological reviews of food contact notifications (FCN) and prenotification consultations (PNC). She received her Ph.D. in Pharmacology from Massachusetts College of Pharmacy and Health Sciences (MCPHS) in Boston in 2009. She completed her postdoctoral training in neuropharmacology (traumatic brain injury) at Georgetown University (Washington, DC) and University of Maryland (Baltimore, MD). She worked as a research faculty member at the School of Medicine of the University of Maryland prior to joining the FDA in 2014. Over the past four years, she has specialized in the use of mechanistically informed modeling approaches for safety assessments of food ingredients. She has utilized classic as well as physiologic pharmacokinetic modeling for predicting toxicity or evaluating the safety of food substances. Her work has been published as peer-reviewed original research articles, review articles, and chapters. She has received awards for her work as a reviewer at FDA and has participated in and chaired sessions at international conferences. She teaches pharmacokinetics classes to undergraduate and graduate students in some universities outside of her work responsibilities at the US FDA.

Chapter 13
Predicting the Risks of Drug-Induced Liver Injury in Humans Utilizing Computational Modeling

Minjun Chen, Jieqiang Zhu, Kristin Ashby, Leihong Wu, Zhichao Liu, Ping Gong, Chaoyang (Joe) Zhang, Jürgen Borlak, Huixiao Hong and Weida Tong

Abstract Drug-induced liver injury (DILI) is a significant challenge to clinicians, drug developers, as well as regulators. There is an unmet need to reliably predict risk for DILI. Developing a risk management plan to improve the prediction of a

M. Chen (✉) · J. Zhu · K. Ashby · L. Wu · Z. Liu · H. Hong · W. Tong
Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA
e-mail: Minjun.Chen@fda.hhs.gov

J. Zhu
e-mail: Jieqiang.Zhu@fda.hhs.gov

K. Ashby
e-mail: Kristin.McEuen@fda.hhs.gov

L. Wu
e-mail: Leihong.Wu@fda.hhs.gov

Z. Liu
e-mail: Zhichao.Liu@fda.hhs.gov

H. Hong
e-mail: Huixiao.Hong@fda.hhs.gov

W. Tong
e-mail: Weida.Tong@fda.hhs.gov

P. Gong
Environmental Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS 39180, USA
e-mail: Ping.Gong@usace.army.mil

C. (Joe) Zhang
School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS 39406, USA
e-mail: Chaoyang.Zhang@usm.edu

J. Borlak
Hannover Medical School, Center of Pharmacology and Toxicology, Hannover, Germany
e-mail: Borlak.Juergen@mh-hannover.de

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_13

drug's hepatotoxic potential is a long-term effort of the research community. Robust predictive models or biomarkers are essential for assessing the risk for DILI in humans, while an improved DILI annotation is vital and largely affects the accuracy and utility of the developed predictive models. In this chapter, we will focus on the DILI research efforts at the National Center for Toxicological Research of the US Food and Drug Administration. We will first introduce our drug label-based approach to annotate the DILI risk associated with individual drugs and then upon these annotations we developed a series of predictive models that could be used to assess the potential of DILI risk, including the "rule-of-two" model, DILI score model, and conventional and modified Quantitative structure–activity relationship (QSAR) models.

Keywords Modeling · Risk management · QSAR · Rule-of-two · DILI score

Abbreviations

DF	Decision Forest
DILI	Drug-Induced Liver Injury
DILIN	Drug-Induced Liver Injury Network
EMA	European Medicines Agency
FDA	Food and Drug Administration
LTKB	Liver Toxicity Knowledge Base
MOA	Mode of Action
QSAR	Quantitative Structure–Activity Relationship
RM	Reactive Metabolites

13.1 Introduction

Drug-induced liver injury (DILI) poses a significant challenge to the medical and pharmaceutical communities as well as regulatory agencies. Many drugs have failed during clinical trials, and over 50 drugs were withdrawn from the worldwide market due to the concern of DILI risk [1]. Because of its significant impact on public health, a series of guidances were published by regulatory agencies to request that the pharmaceutical industry better assesses DILI risk during drug development, including the US Food and Drug Administration (FDA)'s guidance "Drug-Induced Liver Injury: Premarketing Clinical Evaluation" and the European Medicines Agency (EMA)'s "Non-clinical guidance on drug-induced hepatotoxicity" [2].

One significant challenge encountered by drug developers and regulators stems from the lack of sensitive screening methodologies to identify DILI signals at the early stage of drug development, especially before the first-in-human testing [3]. While animal studies remain the "gold standard" of testing strategies in preventing

potentially toxic drug candidates from entering clinical trials [4–6], it is not perfect and sometimes fails to detect hepatotoxic drug candidates; a retrospective analysis revealed that such tests failed in about 45% of DILI cases found in clinical trials [7]. In one notorious example, five subjects in a phase 2 clinical trial experienced fatal hepatotoxicity induced by fialuridine, while this investigational nucleoside analogue showed no liver damage in animal studies [8]. There is unmet need to more reliably predict risk for DILI in humans and to overcome current limitations.

Many worldwide efforts have been launched to better understand and address DILI issues. In the USA, the drug-induced liver injury network (DILIN) was funded by National Institute of Healthy since the year of 1995 and is still today actively collecting and analyzing cases of severe liver injury caused by prescription drugs, over-the-counter drugs, and alternative medicines, such as herbal products and supplements. Similar government supported drug-induced liver injury network efforts were recently established in Europe funded by European cooperation in Science and Technology (http://www.cost.eu/COST_Actions/ca/CA17112). The US FDA has a long-term effort to improve drug safety by better assessing pre-marketing and post-marketing data for identifying signs of toxicity. At the National Center for Toxicological Research, we have developed the Liver Toxicity Knowledge Base (LTKB) which contains diverse liver-related data such as drug properties, DILI mechanisms, and drug metabolism. that can be utilized to develop new models for assessing the risks for DILI in humans [1, 5, 9–21]. In this chapter, we will introduce our continuing efforts toward the development of computational models for the prediction of DILI risks in humans. First, we will present the drug label-based approach to annotate the risk for DILI associated with individual drugs, and then based on these annotations, we developed a panel of predictive models that could be used to assess drug candidates for their potential to cause DILI risk before human testing or during clinical trials.

13.2 Annotation of DILI Risk for Marketed Drugs

Annotation of DILI risk for drugs is challenging. Drugs could cause significantly different scales of DILI risk even when their chemical structures are similar. For example, alpidem and zolpidem both are anxiolytic drugs derived from the imidazopyridine family used as sleeping medication. These two drugs have similar chemical structures but distinct hepatotoxicity (Fig. 13.1): Alpidem was withdrawn due to hepatotoxicity while zolpidem is still widely used in clinical practice with rare hepatotoxicity observed. Drugs withdrawn from market due to hepatotoxicity and those without hepatotoxicity observed represent two extremes within the spectrum of the risk for humans. Most drugs are located within the middle of spectrum depending on the associated DILI risk.

The DILI annotation discussed here refers to the classification of risks of DILI exposure to the human population associated with the drug treatment for various diseases. An improved annotation of DILI is vital and largely affects the accuracy and

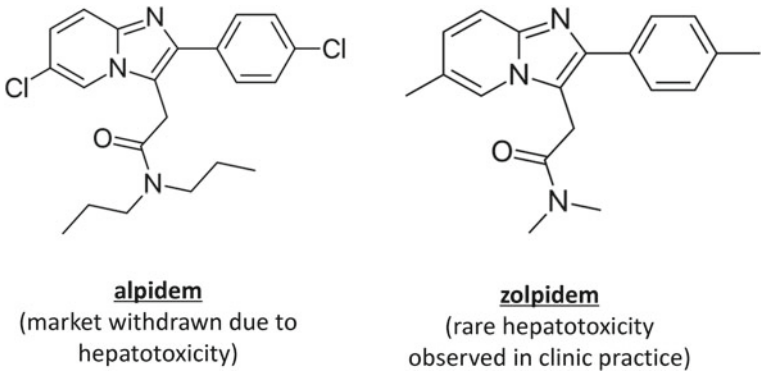

Fig. 13.1 Distinct hepatotoxicity observed between alpidem and zolpidem even though their chemical structures are similar

utility of a predictive model [22]. At least three attributes including severity, causality, and incidence need to be considered when assessing a drug's potential to cause DILI [1]. However, annotating a drug's DILI risk is not trivial in clinical practice [23] due to several hurdles to be considered, i.e., (1) the uncommon occurrence of DILI, (2) the various complicated clinical DILI manifestations, (3) the deficiency of accurate biomarkers for DILI diagnosis, (4) the complications in causality adjudication, (5) and the severe under-reporting of DILI cases.

There is not a single resource which could provide all the information required for an accurate DILI annotation [1]. The research community has put great efforts to address this challenging issue as summarized in some reviews [10]. Overall, the approaches to annotate DILI risk are either based on case reports or on monograph. Case reports can be collected by on-going DILI research projects such as US DILI network and Spain DILI registry, reported in literature [24–26], or retrieved from the FDA's adverse event reporting system [27–29]. Monographs are written by experts based on collection of evidence from a variety of sources, such as the FDA drug labeling [1], the Physicians' Desk Reference [30], and the US pharmacopeia. The information in the monograph documents was authoritative but not updated as frequently as the case reports [31–33]. Given the lack of a "gold standard" that defines DILI risk, certain drugs could have diverse annotations due to the different definitions and data sources for annotations [34]. A comparison among different annotations was reported [35–37]. Overall, the agreements among annotations are acceptable, and normally a higher concordance among hepatotoxicity drugs was present as compared to the non-hepatotoxicity drugs [10, 15, 38].

We selected FDA-approved drug labeling as the main supporting evidence to annotate drugs for their DILI risk for humans. Drug labeling is an authoritative document summarizing drug safety information based on the comprehensive evaluation of data from preclinical studies, clinical testing, post-marketing surveillance, and publications in literature. The information within drug labels summarizes the consensus and serious thoughts from experts at that time with the consideration of all three criteria

(i.e., severity, causality, and incidence) mentioned above [1]. We developed a schema to gather the information from FDA-approved drug labeling to annotate DILI risk and created a benchmark dataset which contained 287 drugs that were categorized into three levels of DILI severity: most-DILI-concern, less-DILI-concern, and no-DILI-concern [33]. Specifically, the 137 drugs categorized as most-DILI-concern are those that were suspended, withdrawn, or issued a black box warning due to hepatotoxicity or had gotten warnings and precautions with moderate or severe DILI concern. Eighty-five drugs categorized as less-DILI-concern had been issued warnings and precaution with mild DILI concern or only recorded hepatotoxicity in the Adverse Reactions section of drug labels. Sixty-five drugs listed as no-DILI-concern are those with no DILI concern mentioned in their drug labels.

The safety data contained in drug labeling are not perfect. A major concern of drug labels was weakness in causality assessment [1], i.e., the definite causal relationship is not mandatorily required for drug labeling, and the regulators were authorized by law to issue a warning when a clinically significant hazard is identified for a drug with reasonable evidence of causality (http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=201.57). Additionally, any modification or updating of the safety information in drug labeling is a stringent and lengthy process that likely causes a time lag from the most updated clinical findings [39]. Meanwhile, case reports could have better timing and be more sensitive to any potential alert signals caused by drugs. Therefore, by incorporating the case report information derived from up-to-date literature and on-going DILI research projects such as the US DILIN project, the drug labeling-based annotation of DILI risk could be further improved.

Upon these considerations, we further refined the labeling-based annotation schema by weighing evidence of case reports together with the information from FDA-approved drug labeling to improve the accuracy of DILI annotation. More specifically, the refined annotation schema was built upon a collection of well-vetted cases (verified via thorough case evaluation by DILI experts) and adjudicated cases (verified using the standardized clinical causality assessment system, i.e., Roussel Uclaf Causality Assessment Method [40]). With this collected causality information, the DILI risk of individual drugs was re-evaluated by complementing drug labeling with available evidence of verified causality. This new schema classified drugs into four categories as detailed as below:

- Withdrawn drugs and those with a black box warning for severe liver injury were classified as verified most-DILI-concern (^{v}Most-DILI-concern) drugs because they are consistently classified as high DILI risk among several published datasets.
- For those drugs which had been warned with severe or moderate DILI occurrence in their labels (i.e., isoniazid) [1], the verification process of causality is needed for the assessment in the new schema: The causality verified drugs will be classified as the ^{v}Most-DILI-concern, otherwise will be reassigned as "Ambiguous DILI-concern."

- Similarly, the less-DILI-concern drugs could be reassigned as verified less-DILI-concern (^{v}Less-DILI-concern) or "Ambiguous DILI-concern," which will depend on whether evidence of causality is available.
- The verified no-DILI-concern drug (^{v}No-DILI-concern) can be confirmed only if the drug was not verified as a cause of DILI in literature and no DILI mentioned in its drug label.

The refined schema was applied to 1036 marketed drugs approved by the FDA before 2010, namely DILIrank dataset, including 192 ^{v}Most-DILI-concern drugs, 278 ^{v}Less-DILI-concern drugs, 312 ^{v}No-DILI-concern drugs, and 254 Ambiguous DILI drug. Notably, given that the existing knowledge will advance over time, the schema we applied in the DILIrank will continuously be updated along with the newly reported DILI cases.

13.3 Predictive Models Developed at NCTR

Developing a risk management plan to improve prediction of a drug's hepatotoxic potential is a long-term effort of the research community [41], and predictive models or biomarkers are essential for assessing the risk for DILI in humans at early stages of drug development, even before the first test in humans. The developments of predictive models for DILI are nicely summarized in several seminal reviews [42]. Here, we briefly introduced some continuing efforts at the FDA's National Center for Toxicological Research for the developing models to predict the risk for DILI in humans, such as the "rule-of-two" model, DILI score model, and conventional and modified Quantitative structure–activity relationship (QSAR) models.

13.3.1 The "Rule-of-Two" Model [11]

Many drugs withdrawn from the market or issued a black box warning due to hepatotoxicity were prescribed at a daily dose of 100 mg or greater [43, 44] while drugs given at a lower daily dose of <10 mg experienced less severe events, suggesting a potential relationship between hepatotoxicity risk and daily dose [31, 45]. Consequently, some experts recommended avoiding the development of drugs requiring a high daily dose to reduce the potential adverse events [42, 46, 47]. Meanwhile, many drugs given at high daily doses are found with little or no risk of DILI, therefore, suggesting that daily dose alone is not a reliable approach to guide drug development, regulatory application, and clinical practice.

Besides daily dose, lipophilicity is an important physicochemical property [48] and is frequently modulated to improve bioavailability and pharmacological activity. Lipophilicity could affect hepatocyte uptake and drug ADMET (i.e., absorption, distribution, metabolism, elimination) behaviors [49], and many lines of evidence

also implicate lipophilicity to be linked to drug toxicity. Nonetheless, it was unclear whether the combination of daily dose and lipophilicity related to risk for DILI in humans.

To better examine the combined effects of daily dose and lipophilicity, a data repository of 164 drugs labeled for their liver liabilities derived from the LTKB-benchmark dataset were used, including $N = 116$ most-DILI-concern drugs and $N = 48$ no-DILI-concern drugs. Lipophilicity was measured by the octanol-water partition coefficient (i.e., logP) which was calculated from the atomic-based prediction of AlogP using Pipeline Pilot (version 8.0, Accelrys Inc, San Diego, CA), and it was categorized into three groups: <1, 1–3, and ≥3 as recommended by literature (13). Daily doses were majorly retrieved from the WHO's ATC database (http://www.whocc.no/atc_ddd_index) and were divided into the groups of <100 mg, 10–100 mg, and ≥100 mg per day as suggested by literature [43, 46].

When the 164 drugs of the dataset were put into the scatter plot of daily doses and logP, the upper right quadrant at a high daily dose and a high logP was majorly distributed with most-DILI-concern drugs. Few no-DILI-concern drugs appeared in this region. The relative risk for DILI associated with various doses and logP constellations was further assessed. Specifically, the subgroup of daily doses ≥100 mg and logP ≥ 3 was associated with a significantly higher proportion of hepatotoxic drugs as compared to the rest of subgroups altogether (96% vs. 41%, odds ratio: 14.05, $P < 0.001$). The analysis demonstrated that a statistically significant association between logP and risk for DILI was observed for the drugs given at daily doses of ≥ 100 mg, while no statistically significant relationship between logP and hepatotoxicity was obtained for the drugs given at daily doses of less than 100 mg.

Similar findings were observed from another independent dataset of 179 oral drugs that 85% of the "rule-of-two" positives are associated with hepatotoxicity as compared with 59% in the "rule-of-two" negatives (odds ratio: 3.89, $P < 0.01$). These evidences together suggest that a drug given at a daily dose of ≥100 mg and with a high logP ≥ 3, namely as the "rule-of-two," is associated with a significant high risk for DILI in humans.

The "rule-of-two" is a simple but effective model to predict the risk for DILI in humans and has been independently evaluated by the drug safety scientists. In a study by Paul Leeson from UK [50], the "rule-of-two" was applied to predict the drugs that failed in drug development due to hepatotoxicity in humans, and 13 of 22 (59%) failed drug candidates were found as "rule-of-two" positives (see Table 13.1). This practice demonstrated that the "rule-of-two" model can be applied to assess drug candidates with similar or even better performance than that among marketed drugs, even though the chemical spaces of drugs candidates in development has significantly shifted from those marketed drugs approved decades ago. Furthermore, another study from a Pfizer team found that the "rule-of-two" model performs better than the three mechanistic endpoints they selected (i.e., cytotoxicity, mitochondrial impairment, and BSEP inhibition) by single, dual combination or triple combinations when evaluated by a total of 125 drugs [51]. Moreover, the "rule-of-two" model was also applied to the direct-acting antiviral for the treatment of chronic hepatitis C and successfully identified the DILI potential associated with Vieraki Pak [52].

Table 13.1 "Rule-of-two model" for prediction of drugs that failed in clinical development due to hepatotoxicity in humans

Compound	Max Daily Dose, mg	cLogP	Rule-of-two prediction
Darbufelone	10	3.7	
Fialuridine	19	0.0	No DILI
Pralnacasan	1200	2.2	
Zamifenacin	40	6.0	
TAK-875	50	4.7	
LY-2409021	90	7.4	
MK-0893	120	7.8	Most DILI
Fiduxosin	120	4.9	Most DILI
CP-457920	120	2.2	
CP-085958	200	4.6	Most DILI
Falnidamol	200	3.8	Most DILI
Pafuramidine	200	4.8	Most DILI
Sitaxentan	300	3.4	Most DILI
ADX-10059	200	4.1	Most DILI
CP-368296	300	2.4	
Telcagepant	560	4.0	Most DILI
CP-724714	500	4.6	Most DILI
CP-422935	500	6.8	Most DILI
Tasosartan	600	2.5	
Solithromycin	800	3.7	Most DILI
CP-456773	1200	3.4	Most DILI
Aplaviroc	1600	3.9	Most DILI
Predicted Most-DILI-concern			13/22 (59%)

Data were collected from Leeson, 2018 [50]

13.3.2 DILI Score Model [12]

The "rule-of-two" model provides added value for predicting DILI risk in humans but could not foresee degree of severity [53, 54]. Additionally, besides dose and lipophilicity, some other mechanistic factors could contribute to the predictive models, facilitating the development of quantitative metrics [55].

Covalent binding of reactive metabolites (RM) is an important toxicity mechanistic factor that could cause direct cellular toxicity or modulate immune reactions [56]. Numerous drugs were reported to generate RM, although their causative relationship for human DILI is still controversial and inconclusive [57]. However, some reports suggest that protein adducts caused by RM seen with drugs are not necessarily asso-

ciated with liver injury [58–60]. Furthermore, a large-scale retrospective analysis demonstrated that the level of covalent binding has no correlation with incidence of liver toxicity observed in vivo in preclinical studies [57]. Even though, considering the possible toxic implications, industry still strongly recommend to minimize the potential of RM formation for drug [61–63] with a target threshold of <50 pmol of RM bound to 1 mg protein [64].

We applied logistic regression analysis to investigate the association between daily dose, logP, RM formation, and DILI risk by using $N = 192$ FDA-approved drugs. The multivariate regression analysis suggested that daily dose, logP, and RM formation all contributed independently to predicting DILI risk, and their contributions were ranked by the order of RM > daily dose/Cmax > logP per the regression coefficients. Consequently, we developed a DILI score model [12] derived from daily dose, logP, and RM: 0.608 * $\log_e$(daily dose/mg) + 0.227 * logP + 2.833 * (RM formation); here, RM was assigned as 1 or 0 based on whether a drug could produce reactive metabolites. As an example, alpidem given at a daily dose of 150 mg/day has a logP of 5.6 and produces RM which resulted in a DILI score of 0.608 * $\log_e(150)$ + 0.216 * 5.6 + 2.833 * 1 = 7.15. Meanwhile, zolpidem (a drug with the same mode of action, similar chemical structure, and preclinical safety profile but with distinct liver toxicity) has a logP of 1.20 and is given at a daily dose of 10 mg, which resulted in a DILI score of 4.51.

The developed DILI score model was evaluated by three independently published datasets assessing its capability to predict the severity of DILI risk in humans. The first dataset was derived from the LTKB-BD with a total of $N = 354$ drug annotated with DILI potential, including 124 most-DILI-concern drugs, 162 less-DILI-concern drugs, and 68 with no-DILI-concern. The second dataset with $N = 227$ drugs retrieved from Greene et al. [24] had $N = 130$ human hepatotoxicity drugs, $N = 44$ drugs with weak evidence, and $N = 53$ drugs with no evidence. The third dataset comes from Suzuki et al. [26] and considered the severity of human hepatotoxicity, of which a total of 182 drugs were obtained consisting of $N = 35$ withdrawn drugs, $N = 61$ with reported acute liver failure cases, and $N = 86$ general DILI drugs. Overall, an increased DILI score significantly correlates with the severity of liver injury. In the first dataset, the DILI risk score decreased in the order of most-DILI-concern > less-DILI-concern > no-DILI-concern [1], and each of the subsequent comparisons was statistically significant ($P < 0.001$). In Greene et al. [24] dataset, DILI score also correctly predicted drugs with evidence for overt human hepatotoxicity having significantly higher DILI scores than those with weak evidence ($P < 0.001$) and not unexpectedly followed those without any evidence for developing DILI ($P < 0.001$). For the data from Suzuki et al. [26], the algorithm also correctly predicted severe DILI cases ($P < 0.001$).

Furthermore, the DILI score model was applied to $N = 165$ clinical cases collected from NIH LiverTox database (https://livertox.nih.gov/), and it was demonstrated that the DILI score correlated with the severity of clinical outcome. The DILI score model was also applied to successfully distinguish some drug pairs such as minocycline/doxycycline, trovafloxacin/moxifloxacin, and benzbro-

marone/amiodarone, which are defined by their molecular structure (tanimoto similarity > 0.5) and similar mode of action but discordant toxicity [65].

13.3.3 Conventional QSAR [13]

QSAR models have been extensively applied to predict drug-induced liver injury due to their ability to produce rapid results without requiring physical drug substance [22, 24, 66, 67]. So far, most of the QSAR-DILI models' report limited predictive performance, with accuracies of approximately 60% or less, especially when the models are challenged by external validation sets. We implemented an improved strategy to develop the QSAR model for predicting DILI in humans using a robust annotation of DILI risk relying on FDA-approved drug labeling and applying an extensive modeling validation strategy to ensure the model performance was sustainable and better than by chance.

Our conventional QSAR was developed by using a decision forest (DF) algorithm to correlate the chemical structures with their DILI risk in humans based on a set of drugs as the training set. The DF algorithm is a supervised machine learning technique utilizing a modified decision tree model by employing a consensus technique to combine multiple heterogeneous decision trees to achieve a more accurate predictive model. The DF algorithm is developed by our laboratory, and the software is publicly available @ https://www.fda.gov/ScienceResearch/BioinformaticsTools/DecisionForest/default.htm. Meanwhile, the chemical structures of drugs were codified into a digital format (i.e., chemical descriptors) as the input for the machine learning algorithm DF. Here, we utilized the Mold2 molecular descriptors to transform the 2-dimensional chemical structures into 777 chemical descriptors. Mold2 is also developed by NCTR and freely available at https://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2/default.htm.

The training set to develop the QSAR model included 197 drugs (NCTR training set), which were annotated by FDA-approved drug labeling as discussed previously. The drug label-based DILI annotation proved to be robust and consistent as compared to other annotations [37], which is critical for the development of an improved QSAR model. The developed models were evaluated by internal and external validations. Internal validation employed a 2000 run of 10-fold cross-validation based on the NCTR training set. External validation of the QSAR models was applied to 3 different datasets with a total of 438 unique drugs: NCTR validation dataset with $N = 190$ drugs, Greene et al. dataset with $N = 328$ drugs, and Xu et al. dataset with $N = 241$ drugs. The validation results in Table 13.2 show that when using the NCTR annotated training or validation set, the predictive performance of the QSAR model had an accuracy of 69.7% for internal cross-validation and 68.9% for external validation. Meanwhile, the external validation assessed by Greene and Xu et al. datasets was at accuracies of 61.6 and 63.1%, respectively. The performances evaluated by different datasets are largely consistent, the occasional variations might reflect the quality of annotation, and the diverse drugs included in the datasets.

Besides the QSAR model for predicting two classes of DILI risk, we also developed another model to assess the three classes of DILI risk (i.e., most-DILI, less-DILI, and no-DILI) [68]. The model was developed by using decision forest (DF) and Mold2 structural descriptors together with DILIrank dataset with >1000 drugs evaluated for their likelihood of causing DILI in humans, of which >700 drugs were classified into three categories used for the model development. Similarly, with two classes of QSAR model, the three-class models were evaluated via cross-validations, bootstrapping validations, and permutation tests for assessing the potential chance correlation. Moreover, prediction confidence analysis was also conducted to provide an additional interpretation of prediction results. These results indicated that the 3-class model showed higher accuracy in differentiating most-DILI drugs from no-DILI drugs than the 2-class DILI model with a potential to categorize DILI risk into a higher resolution.

13.3.4 Modified QSAR Models

Besides developing conventional QSAR models based on chemical structure information only, we also tried to incorporate other drug information, especially those related to DILI-relevant biological functions, to improve model performance. For instance, understanding the mode of action (MOA) of a drug is critical in safety assessment. Therefore, it is promising to improve the predictive model by considering MOA of drugs on DILI. To achieve that, we have developed an algorithm named

Table 13.2 Conventional QSAR performance evaluated by cross-validation and independent validation

	Cross-validation ($N = 2000$ runs)	Independent validation		
	NCTR training set[a]	NCTR validation set	Greene dataset	Xu dataset
Drugs	197 ($P/N =$ 81/116)	190 ($P/N =$ 95/95)	328 ($P/N =$ 214/114)	241 ($P/N =$ 132/109)
Accuracy (%)	69.7 ± 2.9	68.9	61.6	63.1
Sensitivity (%)	57.8 ± 6.2	66.3	58.4	60.6
Specificity (%)	77.9 ± 3.0	71.6	67.5	66.1
PPV (%)	64.6 ± 4.3	70.0	77.2	68.4
NPV (%)	72.6 ± 2.5	68.0	46.4	58.1

Cross-validated results come from the mean values of 2000 runs from 10-fold cross-validations. Independent validation results are predicted results based on the three validation sets, i.e., NCTR validation set, Greene et al. dataset, and Xu et al. dataset

[a]mean ± relative standard deviation

MOA-DILI [69], integrating the MOA and structural information to enhance DILI prediction.

Different from a conventional QSAR model, the modified model will utilize MOA information to categorize drugs, i.e., drugs would be categorized into active or inactive group for each specific MOA. The underlying hypothesis is that MOA-specific drugs would share similar DILI mechanisms and thus would be predicted by the same QSAR models. In other words, we will develop one model to distinguish DILI drugs from all MOA active drugs and another model to separate DILI drugs from all MOA inactive drugs. Finally, these two QSAR models, for active and inactive drugs, respectively, were merged into one assay-specific QSAR model (Fig. 13.2a).

A total of 17 toxicity-relevant MOA assays was curated from the Tox21 dataset [70], including estrogen receptor (ER), androgen receptor (AR), mitochondrial toxicity, p53, PPAR gamma, etc. Therefore, 17 specific MOA-QSAR models were developed, and a consensus approach was applied to determine the DILI risk associated with drugs. Some feature selection strategies (i.e., sequential forward selection) were used to determine DILI-relevant MOAs (assays) for the final model.

The proposed MOA-DILI model was tested on 333 drugs with both clinical DILI annotation and Tox21 assay data available. Mold2 software [71] was used to generate chemical descriptors for the development of QSAR models. Hold-out and cross-validation were used to evaluate the model performance. For the hold-out approach, the 333 drugs were randomly split into 2/3 (222 drugs) and 1/3 (111 drugs). The former (2/3) were used to develop a model while the latter (1/3) were used to evaluate

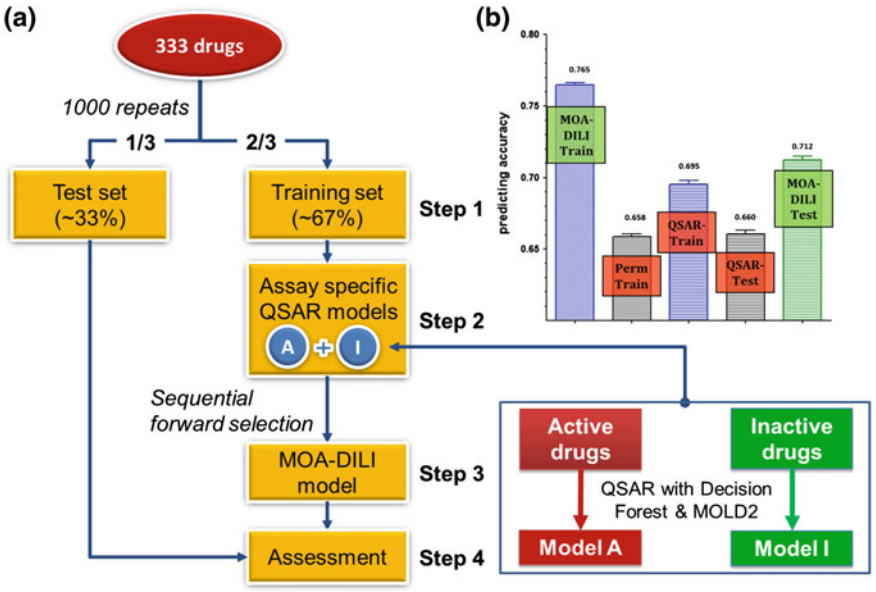

Fig. 13.2 **a** Workflow for MOA-DILI modeling and **b** modeling performance of the MOA-DILI model

the model. The hold-out process was repeated 1000 times to generate training/test sets pairs. Cross-validation was applied inside the training set to evaluate model performance. Label permutation testing with the DILI severity annotations randomly shuffled was applied to check whether the model could generate results better than random.

The optimized MOA-DILI model employed four assays, i.e., ARE-bla (antioxidant response element), ER-luc-bg1-4e2-antagonist (ERalpha, BG1 cell line), gh3-tre-antagonist (thyroid receptor), and PPARG-bla-agonist (peroxisome proliferator-activated receptor gamma). Furthermore, a prediction accuracy of 0.757 in 5-fold cross-validation and 0.695 in hold-out testing was observed for the optimized MOA-DILI model, which is significantly higher than the results obtained from the permutation test (Fig. 13.2b). Moreover, this optimized model has a significantly higher predictive performance than the conventional QSAR model only (Table 13.3), demonstrating the improved predictive power for hepatotoxicity by integrating MOA data of drugs.

Another modified QSAR model was also developed, namely DILI prediction systems [72] which aims to translate the post-marketing surveillance information back to the preclinical stage for improving DILI prediction performance. In DILI prediction systems model, it is hypothesized that there exists a set of hepato-related side effects with discriminative power to distinguish between drugs with or without the risk for DILI. Then, *in silico* models could be developed for those hepato-related side effects based on drug's chemical structure with machine learning algorithms. Based on SIDER datasets [73], 13 different hepato-related side effects were identified and corresponding models were developed by using naïve Bayesian classifier in a single cohesive prediction system. The DILI prediction systems yielded 60–70% accuracies when evaluated using drugs from different DILI annotations. Furthermore, it was found that when a drug was predicted as positive by at least three side effects, the positive predictive value could be boosted to 91%.

13.4 Conclusion

Reliably predicting the risk for DILI in humans is still an unmet need in the research community [34]. Accurate annotation of DILI risk is vital for the development of

Table 13.3 Overall Performance of AOPs-DILI model in training and test set

Model types	5-fold cross-validation	Hold-out test
MOA-DILI model	0.757 (0.022)	0.695 (0.043)
Conventional QSAR model	0.658 (0.031)	0.663 (0.04)
Label permutated model	0.582 (0.042)	0.500 (0.063)

robust predictive models for prediction of DILI risk in humans; however, appropriate annotation is not a trivial task. We utilized the FDA-approved drug labels to annotate a given drug's risk for DILI in humans, which was demonstrated to be robust and consistent across different types of drugs. The schema was further improved by weighing evidence of case reports and was applied to 1036 FDA-approved drugs to classified into three verified DILI groups (i.e., ^{v}Most-, ^{v}Less-, and ^{v}No-DILI-concern) with an additional group of drugs with DILI concern but without verified causality (ambiguous annotation).

Besides the improved DILI annotations, we could develop better models by utilizing the relevant contributing factors and advanced modeling technologies. We have developed a series of computational predictive models that use in silico or physicochemical methods, including the "rule-of-two" model, DILI score model, conventional QSAR model for the prediction of two classes and multiple classes of DILI, and modified QSAR model including MOA-DILI model and DILI prediction systems model. Some models such as "rule-of-two" were independently validated and successfully identified drugs with significant hepatotoxicity. In the future, some emerging technologies (e.g., high-throughput screening or high-content assay, induced Pluripotent Stem Cells (iPSCs), engineered human liver cocultures, and 3D cell culture) [74–78] could be incorporated into predictive models for a better identification of DILI risk liability at the early stage of drug development. In addition to the drug properties we discussed above, host factors and their interactions with drug properties [79, 80] should be considered and this information should be incorporated into current drug-based models to improve prediction of DILI.

Disclaimer This article reflects the views of the authors and should not be construed to represent FDA's views or policies.

References

1. Chen M et al (2011) FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today 16(15–16):697–703
2. Avigan MI, Muñoz MA (2018) Perspectives on the regulatory and clinical science of drug-induced liver injury (DILI). In: Drug-induced liver toxicity. Springer, Berlin, pp 367–393
3. Noureddin N, Kaplowitz N (2018) Overview of mechanisms of drug-induced liver injury (DILI) and key challenges in DILI research. In: Drug-induced liver toxicity. Springer, Berlin, pp 3–18
4. Fielden MR, Kolaja KL (2008) The role of early in vivo toxicity testing in drug discovery toxicology. Expert Opin Drug Saf 7(2):107–110
5. Zhang M, Chen M, Tong W (2012) Is toxicogenomics a more reliable and sensitive biomarker than conventional indicators from rats to predict drug-induced liver injury in humans? Chem Res Toxicol 25(1):122–129
6. Khan I, Hausner E (2018) Regulatory toxicological studies: identifying drug-induced liver injury using nonclinical studies. In: Drug-induced liver toxicity. Springer, Berlin, pp 395–410
7. Olson H et al (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol 32(1):56–67

8. Lewis W et al (1996) Fialuridine and its metabolites inhibit DNA polymerase gamma at sites of multiple adjacent analog incorporation, decrease mtDNA abundance, and cause mitochondrial structural defects in cultured hepatoblasts. Proc Natl Acad Sci 93(8):3592–3597
9. Chen M et al (2013) Liver Toxicity Knowledge Base (LTKB)—a systems approach to a complex end point. Clin Pharmacol Ther 95(5):409–412
10. Chen M et al (2014) Toward predictive models for drug-induced liver injury in humans: are we there yet? Biomark Med 8(2):201–213
11. Chen M, Borlak J, Tong W (2013) High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. Hepatology 58(1):388–396
12. Chen M, Borlak J, Tong W (2016) A model to predict severity of drug-induced liver injury in humans. Hepatology 64(3):931–940
13. Chen M et al (2013) Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs. Toxicol Sci 136(1):242–249
14. Chen M et al (2015) Drug-induced liver injury: interactions between drug properties and host factors. J Hepatol 63(2):503–514
15. Chen M et al (2016) DILIrank—the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. Drug Discov Today 21(4):648–653
16. Chen M et al (2014) An improved testing strategy to predict risk for drug-induced liver injury in humans using high-content screening assays and the 'Rule-of-two' model. Arch Toxicol 88(7):1439–1449
17. Liu Z et al (2011) Translating clinical findings into knowledge in drug safety evaluation–drug induced liver injury prediction system (DILIps). PLoS Comput Biol 7(12):e1002310
18. Mishra P, Chen M (2017) Application of "Rule-of-two" model to direct-acting antivirals for treatment of chronic hepatitis C: can it predict potential for hepatotoxicity? Gastroenterology 152(6):1270–1274
19. Wang Y et al (2013) A unifying ontology to integrate histological and clinical observations for drug-induced liver injury. Am J Pathol 182(4):1180–1187
20. Yu K et al (2014) High daily dose and being a substrate of cytochrome P450 enzymes are two important predictors of drug-induced liver injury. Drug Metab Dispos 42(4):744–750
21. Yu K et al (2014) Mining hidden knowledge for drug safety assessment: topic modeling of LiverTox as a case study. BMC Bioinform 15(Suppl 17):S6
22. Przybylak KR, Cronin MT (2012) In silico models for drug-induced liver injury—current status. Expert Opin Drug Metab Toxicol 8(2):201–217
23. Abboud G, Kaplowitz N (2007) Drug-induced liver injury. Drug Saf 30(4):277–294
24. Greene N et al (2010) Developing structure-activity relationships for the prediction of hepatotoxicity. Chem Res Toxicol 23(7):1215–1222
25. Gustafsson F et al (2013) A correlation between the in vitro drug toxicity of drugs to cell lines which express human P450s and their propensity to cause liver injury in humans. Toxicol Sci 137(1):189–211
26. Suzuki A et al (2010) Drugs associated with hepatotoxicity and their reporting frequency of liver adverse events in VigiBase: unified list based on international collaborative work. Drug Saf 33(6):503–522
27. Ursem CJ et al (2009) Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans. Part A: use of FDA post-market reports to create a database of hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54(1):1–22
28. Rodgers AD et al (2010) Modeling liver-related adverse effects of drugs using k nearest neighbor quantitative structure-activity relationship method. Chem Res Toxicol 23(4):724–732
29. Zhu X, Kruhlak NL (2014) Construction and analysis of a human hepatotoxicity database suitable for QSAR modeling using post-market safety data. Toxicology 321:62–72
30. Sakatis MZ et al (2012) Preclinical strategy to reduce clinical hepatotoxicity using in vitro bioactivation data for >200 compounds. Chem Res Toxicol 25(10):2067–2082
31. Lammert C et al (2008) Relationship between daily dose of oral medications and idiosyncratic drug-induced liver injury: search for signals. Hepatology 47(6):2003–2009

32. Weng Z et al (2015) A comprehensive study of the association between drug hepatotoxicity and daily dose, liver metabolism, and lipophilicity using 975 oral medications. Oncotarget 6(19):17031–17038
33. Xu JJ et al (2008) Cellular imaging predictions of clinical drug-induced liver injury. Toxicol Sci 105(1):97–105
34. Atienzar FA, Nicolas J-M (2018) Prediction of human liver toxicity using in vitro assays: limitations and opportunities. In: Drug-induced liver toxicity. Springer, Berlin, pp 125–150
35. Thakkar S et al (2018) The Liver Toxicity Knowledge Base (LKTB) and drug-induced liver injury (DILI) classification for assessment of human liver injury. Expert Rev Gastroenterol Hepatol 12(1):31–38
36. Chen M et al (2014) Toward predictive models for drug-induced liver injury in humans: are we there yet? Biomark Med 8(2):201–213
37. Chen M et al (2016) DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. Drug Discov Today 21(4):648–653
38. Thakkar S et al (2018) Drug-induced liver injury (DILI) classification and its application on human DILI risk prediction. In: Drug-induced liver toxicity. Springer, Berlin, pp 45–59
39. Regev A (2014) Drug-induced liver injury and drug development: industry perspective. Semin Liver Dis 34(02):227–239
40. García-Cortés M et al (2011) Causality assessment methods in drug induced liver injury: strengths and weaknesses. J Hepatol 55(3):683–691
41. Light DS, Aleo MD, Kenna JG (2018) Interpretation, integration, and implementation of in vitro assay data: the predictive toxicity challenge. In: Drug-induced liver toxicity. Springer, Berlin, pp 345–364
42. Ballet F (2010) Back to basics for idiosyncratic drug-induced liver injury: dose and metabolism make the poison. Gastroenterol Clin Biol 34(6–7):348–350
43. Walgren JL, Mitchell MD, Thompson DC (2005) Role of metabolism in drug-induced idiosyncratic hepatotoxicity. Crit Rev Toxicol 35(4):325–361
44. Uetrecht JP (1999) New concepts in immunology relevant to idiosyncratic drug reactions: the "danger hypothesis" and innate immune system. Chem Res Toxicol 12(5):387–395
45. Senior JR (2008) What is idiosyncratic hepatotoxicity? What is it not? Hepatology 47(6):1813–1815
46. Uetrecht J (2001) Prediction of a new drug's potential to cause idiosyncratic reactions. Curr Opin Drug Discov Dev 4(1):55–59
47. Lammert C et al (2010) Oral medications with significant hepatic metabolism at higher risk for hepatic adverse events. Hepatology 51(2):615–620
48. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. Nat Rev Drug Discov 6(11):881–890
49. Waring MJ (2010) Lipophilicity in drug discovery. Expert Opin Drug Discov 5(3):235–248
50. Leeson PD (2018) Impact of physicochemical properties on dose and hepatotoxicity of oral drugs. Chem Res Toxicol 31(6):494–505
51. Shah F et al (2015) Setting clinical exposure levels of concern for drug-induced liver injury (DILI) using mechanistic in vitro assays. Toxicol Sci 147(2):500–514
52. Mishra P, Chen M (2017) Direct-acting antivirals for chronic hepatitis C: can drug properties signal potential for liver injury? Gastroenterology 152(6):1270–1274
53. Kaplowitz N (2013) Avoiding idiosyncratic DILI: two is better than one. Hepatology 58(1):15–17
54. Lewis JH (2014) Drug-induced liver injury, dosage, and drug disposition: is idiosyncrasy really unpredictable? Clin Gastroenterol Hepatol 4(1):4–8
55. Aleo MD et al (2014) Human drug-induced liver injury severity is highly associated to dual inhibition of liver mitochondrial function and bile salt export pump. Hepatology 60(3):1015–1022
56. Stephens C, Andrade RJ, Lucena MI (2014) Mechanisms of drug-induced liver injury. Curr Opin Allergy Clin Immunol 14(4):286–292
57. Park BK et al (2011) Managing the challenge of chemically reactive metabolites in drug development. Nat Rev Drug Discov 10(4):292–306

58. Stepan AF et al (2011) Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. Chem Res Toxicol 24(9):1345–1410
59. Smith DA, Obach RS (2006) Metabolites and safety: what are the concerns, and how should we address them? Chem Res Toxicol 19(12):1570–1579
60. Uetrecht J (2006) Evaluation of which reactive metabolite, if any, is responsible for a specific idiosyncratic reaction*. Drug Metab Rev 38(4):745–753
61. Srivastava A et al (2010) Role of reactive metabolites in drug-induced hepatotoxicity. In: Adverse drug reactions. Springer, Berlin, pp 165–194
62. Park B et al (2011) Drug bioactivation and protein adduct formation in the pathogenesis of drug-induced toxicity. Chem Biol Interact 192(1):30–36
63. Kalgutkar AS et al (2005) A comprehensive listing of bioactivation pathways of organic functional groups. Curr Drug Metab 6(3):161–225
64. Evans DC et al (2004) Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. Chem Res Toxicol 17(1):3–16
65. Chen M et al (2012) A decade of toxicogenomic research and its contribution to toxicological science. Toxicol Sci 130(2):217–228
66. Hong H et al (2018) Quantitative structure–activity relationship models for predicting risk of drug-induced liver injury in humans. In: Drug-induced liver toxicity. Springer, Berlin, pp 77–100
67. Fisk L, Greene N, Naven R (2018) Physicochemical properties and structural alerts. In: Drug-induced liver toxicity. Springer, pp 61–76
68. Hong H et al (2017) Development of decision forest models for prediction of drug-induced liver injury in humans using a large set of FDA-approved drugs. Sci Rep 7(1):17311
69. Wu L et al (2017) Integrating drug's mode of action into quantitative structure-activity relationships for improved prediction of drug-induced liver injury. J Chem Inf Model 57(4):1000–1006
70. Tice RR et al (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121(7):756
71. Hong H et al (2008) Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. J Chem Inf Model 48(7):1337–1344
72. Liu Z et al (2011) Translating clinical findings into knowledge in drug safety evaluation—drug induced liver injury prediction system (DILIps). PLoS Comput Biol 7(12):e1002310
73. Kuhn M et al (2010) A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 6:343
74. Wong MW et al (2018) Status and use of Induced Pluripotent Stem Cells (iPSCs) in toxicity testing. In: Drug-induced liver toxicity. Springer, Berlin, pp 199–212
75. Monckton CP, Khetani SR (2018) Engineered human liver cocultures for investigating drug-induced liver injury. In: Drug-induced liver toxicity. Springer, Berlin, pp 213–248
76. Otieno MA, Gan J, Proctor W (2018) Status and future of 3D cell culture in toxicity testing. In: Drug-induced liver toxicity. Springer, Berlin, pp 249–261
77. Persson M (2018) High content screening for prediction of human drug-induced liver injury. In: Drug-induced liver toxicity. Springer, Berlin, pp 331–343
78. Porceddu M et al (2018) In vitro assessment of mitochondrial toxicity to predict drug-induced liver injury. In: Drug-induced liver toxicity. Springer, Berlin, pp 283–300
79. Chen M et al (2015) Drug-induced liver injury: Interactions between drug properties and host factors. J Hepatol 63(2):503–514
80. Stephens C, Lucena MI, Andrade RJ (2018) Host risk modifiers in idiosyncratic drug-induced liver injury (DILI) and its interplay with drug properties. In: Drug-induced liver toxicity. Springer, Berlin, pp 477–496

Minjun Chen is a principal investigator working at the Division of Bioinformatics and Biostatics of the US FDA's National Center for Toxicological Research and serve as the adjunct faculty

and mentor for the bioinformatics program joint by Univ. of Arkansas at Little Rock (UALR) and Univ. of Arkansas for Medical Sciences (UAMS). He received the FDA award for outstanding junior investigator (2012) and the NCTR scientific achievement award (2014) and the FDA scientific achievement award (2017). Currently, he co-chairs the FDA Liver Toxicity Working Group since 2014 and is the book editor titled "Drug-induced Liver Toxicity" published by the Springer in 2018. His primary research interests encompass drug-induced liver injury, drug safety, bioinformatics, and personalized medicine. He has authored or co-authored more than 90 original publications and book chapters.

Jieqiang Zhu is a ORISE fellow at NCTR. He earned his Ph.D. in Medicine Chemistry from Zhejiang University in China. His current interests focus on (1) development of prediction models for drug-induced liver injury and (2) genome-based prediction model for breast cancer risk.

Kristin Ashby recently completed her MS in Bioinformatics from the University of Arkansas at Little Rock and University of Arkansas for Medical Sciences. Her research has focused on integrating tools that predict drug-induced liver injury and evaluating predictive model parameters.

Leihong Wu is a staff fellow of the Division of Bioinformatics and Biostatistics at the US FDA's National Center for Toxicological Research (NCTR), where he currently focuses the research on developing in silico predictive models for drug- and dietary supplement-associated hepatotoxicity. He developed MOA-DILI approach, which utilizes chemical structure information as well as high content assay data to construct models to evaluate FDA-relevant drugs for liver toxicities. He also helps work on the development of Liver Toxicity Knowledge Base (LTKB) to integrate multiple resources, including (1) annotation data, (2) genomics data, (3) assay data, (4) chemical information, etc. He received his Ph.D. degree in Pharmacology from Zhejiang University, China in 2013 and was a postdoctoral fellow at NCTR from 2014 to 2017. He is author/co-author of 28 publications including 12 first author publications.

Zhichao Liu got his Ph.D. from Nankai University in China. He is currently a principal investigator at NCTR, FDA. Dr. Liu's background is across Chemistry, Biology, and Computer Science. In the past fifteen years, he led and took part in many cutting-edge projects from both industry and academia. Specifically, he developed innovative approaches for in silico drug repositioning framework to seek for novel opportunities for rare disease treatment development. The goals aim at providing the standard in silico pipeline for drug repositioning and early drug safety detection by retrieving, integrating, and organizing the information from chemical, biological, and clinical spaces, which help industry to seek for the optimal route to accelerate the drug development efficacy from advanced regulatory sciences perspective. His interests also widely cover (1) development of flexible and integrative risk prediction systems for drug safety evaluation and (2) development and application of standard pipeline for genomic data (e.g., microarray/NGS) analysis for drug safety and efficacy questions. Dr. Liu has published more than 60 papers in the prestigious journals.

Ping Gong is a principal investigator with multidisciplinary expertise in environmental genomics, bioinformatics, bioengineering, mechanistic and predictive toxicology, and molecular modeling. He earned a Bachelor's degree in Environmental Biology and Ecology from Peking University and a Ph.D. in Environmental Toxicology from the Institute of Applied Ecology, Chinese Academy of Sciences. He completed his postdoctoral studies in the Technology University of Berlin (TU Berlin), Swedish University of Agricultural Sciences (SLU), and Biotechnology Research Institute of National Research Council of Canada. As a research biologist, he has led a number of multi-year projects that have greatly improved the fundamental and mechanistic understanding underlying the observed toxicities of military unique and other environmental contaminants. His current research interests focus on mode of action-guided predictive toxicology, genetic variations-conferred herbicide resistance, epigenetics-driven transgenerational inheritance of phe-

notypic traits, and synthetic biology-based bioengineering and biocontrol of invasive species. He has advised more than 18 graduate students and published over 75 peer-reviewed research articles and book chapters. He also served on the Board of Directors of MidSouth Computational Biology and Bioinformatics Society (MCBIOS) as well as the editorial boards of Frontiers in Genetics and Environmental Toxicology and Chemistry.

Chaoyang (Joe) Zhang is a professor in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi (USM). His research interests include data mining, machines learning, big data analytics and their applications to interdisciplinary areas of bioinformatics, health care, and toxicity analysis. His research has been supported by the National Science Foundation (NSF), Department of Defense (DOD), National Academy of Sciences (NAS), and National Institute of Health (NIH). Dr. Zhang received his MS in computer science and Ph.D. in computational analysis and modeling from Louisiana Tech University. Prior to joining USM in 2003, he was a research assistant professor in the Department of Computer Science at the University of Vermont.

Jürgen Borlak was born in Neu-Ulm, Germany in 1958. After studies at Universities in Germany and abroad, he obtained his Doctorate in Pharmacology and Toxicology at the University of Reading, GB. Following residencies in the UK and France (Strasbourg), he was habilitated in pharmacology and toxicology and received the venia legendi ("Privatdozent") at Hannover Medical School in the year 2000. Two years later, he was appointed as a full professor of Pharmacology and Toxicology at Hannover Medical School. From 2002 onwards, he has been the Director of the Institute of Pharmaco- and Toxicogenomics at Hannover Medical School. This new field of genomic science applies a wide range of methods in genetics, molecular biology, molecular toxicology, and functional genomics for a better understanding of disease-causing mechanisms and drug-induced toxicities. An array of enabling technologies is applied for an identification of "druggable" targets and for a better understanding of inter-individual differences in drug response, therefore, allowing individualized drug treatment regimens and disease prevention strategies. Dr. Borlak is author of >260 original publications and 25 book chapters and editor of the Handbook of Toxicogenomics. He is reviewer and member of the editorial board for various scientific journals. Among others, he is an appointed expert of the World Health Organization (WHO), the European Medicines Agency (EMA), and the German regulatory agency (BfArM). He is also an international reviewer for many European, US, and Asian Research Organizations.

Huixiao Hong is the Chief of Bioinformatics Branch, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research (NCTR), US Food and Drug Administration (FDA), working on the scientific bases for regulatory applications of bioinformatics. Before joining the FDA, he was the Manager of Bioinformatics Division of Z-Tech, an ICFI company. He was a research scientist position at Sumitomo Chemical Company in Japan and a visiting scientist at National Cancer Institute at National Institutes of Health. He was also an Associate Professor and the Director of Laboratory of Computational Chemistry at Nanjing University in China. Dr. Hong is the member of steering committee of OpenTox, the member of board directors of US MidSouth Computational Biology and Bioinformatics Society, and in the leadership circle of US FDA modeling and simulation working group. He published more than 180 scientific papers and served as editor-in-chief and editorial board member for multiple peer-reviewed journals. He received his Ph.D. from Nanjing University in China in 1990 and conducted research in Leeds University in England in 1990–1992.

Weida Tong is Director of Division of Bioinformatics and Biostatistics at FDA's National Center for Toxicological Research (NCTR/FDA). He has served as a science advisory board member for several large projects involving multiple institutes in Europe and the USA. He also holds several adjunct positions at universities in the USA and China. His division at FDA works to develop bioinformatic methodologies and standards to support FDA research and regulation and to

advance regulatory science and personalized medicine. The most visible projects from his group are (1) leading the Microarray Quality Control (MAQC) consortium to develop standard analysis protocols and quality control metrics for emerging technologies to support regulatory science and precision medicine; (2) development of liver toxicity knowledge base (LTKB) for drug safety; (3) in silico drug repositioning for the enhanced treatment of rare diseases; and (4) development of the FDA bioinformatics system, ArrayTrackTM suite, to support FDA review and research on pharmacogenomics. In addition, his group also specializes in molecular modeling and QSARs with specific interest in estrogen, androgen, and endocrine disruptor. Dr. Tong has published more than 230 papers and book chapters.

Chapter 14
Predictive Modeling of Tox21 Data

Ruili Huang

Abstract As an alternative to traditional animal toxicology studies, the toxicology for the twenty-first century (Tox21) program initiated a large-scale, systematic screening of chemicals against target-specific, mechanism-oriented in vitro assays aiming to predict chemical toxicity based on these in vitro assay data. The Tox21 library of ~10,000 environmental chemicals and drugs, representing a wide range of structural diversity, has been tested in triplicate against a battery of cell-based assays in a quantitative high-throughput screening (qHTS) format generating over 85 million data points that have been made publicly available. This chapter describes efforts to build in vivo toxicity prediction models based on in vitro activity profiles of compounds. Limitations of the current data and strategies to select an optimal set of assays for improved model performance are discussed. To encourage public participation in developing new methods and models for toxicity prediction, a "crowd-sourcing" challenge was organized based on the Tox21 assay data with successful outcomes.

Keywords Computational modeling · Human toxicity · Animal toxicity · Adverse drug effect · In vitro assay · High-throughput screening

Abbreviations

ADE	Adverse Drug Effect
ACToR	Aggregated Computational Toxicology Online Resource
AUC-ROC	Area Under the Receiver Operating Characteristic curve
ASNN	Associative Neural Networks
BLA	Beta-lactamase

R. Huang (✉)
National Center for Advancing Translational Sciences, NIH, 9800 Medical Center Drive, Rockville, MD 20850, USA
e-mail: huangru@mail.nih.gov

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_14

CEBS	Chemical Effects in Biological Systems
CYP	Cytochrome P450
DMSO	Dimethylsulfoxide
DTA	Drug Target Annotation
EPA	Environmental Protection Agency
FN	False Negative
FP	False Positive
FDA	Food and Drug Administration
GPCR	G-Protein-Coupled Receptor
HTS	High-Throughput Screening
NCATS	National Center for Advancing Translational Sciences
NCCT	National Center for Computational Toxicology
NIEHS	National Institute of Environmental Health Sciences
NTP	National Toxicology Program
NR	Nuclear Receptor
QC	Quality Control
qHTS	Quantitative High-Throughput Screening
QSAR	Quantitative Structure–Activity Relationship
ROC	Receiver operating characteristic
SOM	Self-Organizing Map
SR	Stress Response
Tox21	Toxicology for the twenty-first Century
TN	True Negative
TP	True Positive
WFS	Weighted Feature Significance

14.1 Introduction

Animal-based in vivo models have been traditionally used to assess the toxicological effects of chemicals, with results extrapolated to foreshadow potentially harmful events in humans. More than 80,000 chemicals are currently registered for use in the United States, for 95% of which no data on human exposure and/or hazard is available [1]. In addition, about 2000 new chemicals are being introduced into our environment every year that may pose hazards for human health [1]. Traditional toxicity testing methods rely heavily on low throughput, expensive, and time-consuming animal studies, making it impossible to evaluate the in vivo toxicity of the fast growing number of chemicals in a cost-efficient and timely manner. The reliability of extrapolating test results derived from animals to health effects in humans poses another challenge due to species differences. High-throughput screening (HTS) techniques, now routinely used in conjunction with computational methods and information technology to probe how chemicals interact with biological systems, offer a new alternative to traditional toxicity testing. Through HTS assays, patterns of cellular

response induced by certain chemicals or chemical classes are established, which might be predictive of adverse health outcomes in humans.

Aiming to identify in vitro chemical signatures that could act as predictive surrogates for in vivo toxicity, the US Tox21 (toxicology in the twenty-first century) program was established in 2008 with an emphasis on developing new methodologies to evaluate the potential risk of thousands of environmental chemicals on human health [2–5]. The Tox21 program is a collaboration involving the National Toxicology Program (NTP) at the National Institute of Environmental Health Sciences (NIEHS), the National Center for Computational Toxicology (NCCT) at the US Environmental Protection Agency (EPA), the intramural research division of the National Center for Advancing Translational Sciences (NCATS), and the US Food and Drug Administration (FDA). The goal of the Tox21 program is threefold:

(1) analyze the patterns of compound-induced biological responses in order to identify toxicity pathways and compound mechanisms of toxicity;
(2) prioritize compounds for further extensive toxicological evaluation; and
(3) develop predictive models for biological response in human beings.

This inter-federal agency partnership established a collection of ~10,000 environmental chemicals and drugs (Tox21 10K library) to profile for potential effects on human health [6, 7]. The Tox21 10K library is screened against a large panel of cell-based assays in a quantitative high-throughput screening (qHTS) format as 15-pt titration series in triplicate [8]. qHTS generates a concentration–response curve for each compound, which greatly reduces the frequency of false positives and negatives. During phase II (2011–2017), the production phase of the program, the Tox21 10K library had been screened against a panel of more than 50 assays with an initial focus on the nuclear receptor (NR) [9–13] and stress response (SR) pathways (Table 14.1) [14–16] in qHTS format producing over 85 million data points to date [7, 17–19]. The results form a rich set of compound in vitro activity profiles that can serve as the basis for mechanism of compound toxicity hypotheses generation and predictive modeling. However, as with any new technology, the reliability and the relevance of the approach need to be evaluated and validated. Here, we describe a few examples using this dataset to build in silico models to predict in vivo including human toxicity. The performances of in vitro assay data-based models are compared with those of chemical structure information, animal toxicity data, and literature annotations on compound target and mode of action. Strategies are proposed to optimize the biological space coverage of in vitro assays in order to establish comprehensive compound activity profiles for improved toxicity prediction.

Table 14.1 Tox21 assay 10K library screens

Assay	Assay name	PubChem AID	Reproducibility score
Stress response pathway assays			
ATAD5	ATAD5	651632, 720516, 651634	99.51
DT40 Rad54/Ku70	DNA repair Rad54/Ku70 deficient isogenic chicken DT40 cell viability	70743015	90.61
DT40 Rev3	DNA repair Rev3 deficient isogenic chicken DT40 cell viability	743014	83.02
DT40 WT	Wild type chicken DT40 cell viability	743015	79.58
H2AX	H2AX	1224845, 1224847	93.71
P53	P53	651631, 720552, 651633	103.02
AP-1-BLA	Activator protein-1	1159526, 1159525, 1159528	91.65
ARE	Antioxidant responsive element	743202, 743219, 743203	81.85
HRE-BLA agonist	Hypoxia-inducible factor-1	1224846, 1224844	96.96
HSE-BLA	Heat shock factor response element	743210, 743228, 743209	95.46
NFκB-BLA agonist	Nuclear factor kappa B	1159509, 1159515, 1159518	100.31
ER stress	Endoplasmic reticulum stress	1159516, 1159517, 1159519	92.97
Mitochondria toxicity	Mitochondrial toxicity	720635, 720637, 720634	87.20
Nuclear receptor assays			
Aromatase	Aromatase	743083, 743139, 743084	93.66
AhR	Aryl hydrocarbon receptor	743085, 743122, 743086	84.05
AR-BLA agonist	Androgen receptor, ligand binding domain, agonist mode	743036, 743053	89.55
AR-BLA antagonist	Androgen receptor, ligand binding domain, antagonist mode	743035, 743063, 743033	96.44

(continued)

Table 14.1 (continued)

Assay	Assay name	PubChem AID	Reproducibility score
AR-MDA agonist	Androgen receptor, full length, agonist mode	743040, 1259381, 1259379, 1259387	99.74
AR-MDA antagonist	Androgen receptor, full length, antagonist mode	743042, 743054, 743041, 1259243, 1259242, 1259247	92.28
CAR agonist	Constitutive androstane receptor, agonist mode	1224839, 1224836	99.93
CAR antagonist	Constitutive androstane receptor, antagonist mode	1224838, 1224837	89.24
ER-BG1 agonist	Estrogen receptor alpha, full length, agonist mode	743079, 1259383, 1259386, 1259391	91.46
ER-BG1 antagonist	Estrogen receptor alpha, full length, antagonist mode	743080, 743091, 743081, 1259244, 1259241, 1259248	95.25
ER-BLA agonist	Estrogen receptor alpha, ligand binding domain, agonist mode	743075, 743077	95.25
ER-BLA antagonist	Estrogen receptor alpha, ligand binding domain, antagonist mode	743069, 743078, 743074	84.90
ER-beta-BLA agonist	Estrogen receptor beta, agonist mode	1259377, 1259380, 1259394	91.76
ER-beta-BLA antagonist	Estrogen receptor beta, antagonist mode	1259378, 1259382, 1259396	90.96
ERR	Estrogen-related receptor	1224834, 1224849, 1224848, 1259403, 1259404	95.66
ERR-PGC	Estrogen-related receptor with the pleiotropic PPAR-gamma coactivator (PGC)	1224842, 1224841, 1224840, 1259401, 1259402	93.35
FXR-BLA agonist	Farnesoid X receptor, agonist mode	743220, 743239, 743218	95.09
FXR-BLA antagonist	Farnesoid X receptor, antagonist mode	743217, 743240, 743221	88.48

(continued)

Table 14.1 (continued)

Assay	Assay name	PubChem AID	Reproducibility score
GR-BLA agonist	Glucocorticoid receptor, agonist mode	720691, 720719	94.89
GR-BLA antagonist	Glucocorticoid receptor, antagonist mode	720692, 720725, 720693	77.40
PPAR-delta-BLA agonist	Peroxisome proliferator-activated receptor delta, agonist mode	743212, 743227, 743211	91.91
PPAR-delta-BLA antagonist	Peroxisome proliferator-activated receptor delta, antagonist mode	743215, 743226, 743213	90.37
PPAR-gamma agonist	peroxisome proliferator-activated receptor gamma, agonist mode	743094, 743140	93.81
PPAR-gamma antagonist	Peroxisome proliferator-activated receptor gamma, antagonist mode	743191, 743199, 743194	85.15
RAR agonist	Retinoic acid receptor, agonist mode	1159553	89.15
RAR antagonist	Retinoic acid receptor, antagonist mode	1159552	71.50
ROR antagonist	Retinoid-related orphan receptor gamma, antagonist mode	1159521	72.95
RXR-BLA agonist	Retinoid X receptor, agonist mode	1159527, 1159529, 1159531	79.20
TR-beta agonist	Thyroid hormone receptor, agonist mode	743066	87.84
TR-beta antagonist	Thyroid hormone receptor, antagonist mode	743065, 743067, 743064	87.82
VDR-BLA agonist	Vitamin D receptor, agonist mode	743222, 743241, 743224	91.79

(continued)

Table 14.1 (continued)

Assay	Assay name	PubChem AID	Reproducibility score
VDR-BLA antagonist	Vitamin D receptor, antagonist mode	743223, 743242, 743225	88.97
Other assays			
HDAC	Histone deacetylase	1259364, 1259365, 1259388	100.06
TSHR agonist	Thyroid stimulating hormone receptor, agonist mode	1224843, 1224895	90.49
TSHR antagonist	Thyroid stimulating hormone receptor, antagonist mode	1259385, 1259395	93.27
TSHR wide type	Thyroid stimulating hormone receptor, wild type	1259384, 1259393	99.62
Shh-GLI3 agonist	Sonic hedgehog (Shh) signaling, agonist mode	1259368, 1259366, 1259390	90.36
Shh-GLI3 antagonist	Sonic hedgehog (Shh) signaling, antagonist mode	1259369, 1259367, 1259392	68.79
Real-time viability (HepG2)	Real-time cell viability in HepG2 cells	1224882, 1224876, 1224878, 1224883, 1224879, 1224890, 1224889, 1224885, 1224867, 1224877, 1224870, 1224873	N/A
Real-time viability (HEK293)	Real-time cell viability in HEK293 cells	1224869, 1224888, 1224875, 1224881, 1224871, 1224884, 1224880, 1224872, 1224886, 1224868, 1224874, 1224887	N/A
Luciferase biochemical	Firefly luciferase, biochemical	1224835	99.97

14.2 Method and Materials

14.2.1 Tox21 Compound Collection and Assay Data

The Tox21 compound library is a collection of over 10,000 environmental chemicals and drugs (approximately 8947 unique chemical entities) representing a wide range of structural diversity [6, 7]. Three Tox21 partners—EPA, NTP, and NCATS—each contributed over 3000 physical compounds to the collection primarily procured

from commercial sources. The main criteria for selection of the Tox21 compounds included, but were not limited to, known or perceived environmental hazards or exposure concerns, physicochemical properties (molecular weight, volatility, solubility, logP) suitability for qHTS, commercial availability, and cost. The types of compounds selected for constructing the library include pesticides, food additives, flame retardants, industrial chemicals, drinking water disinfection by-products, preservatives, household cleaning agents, and drugs [20]. In addition, a set of 88 diverse compounds was included as duplicates in every screening plate and designated as internal controls for assay reproducibility [8]. All of the compounds in the Tox21 10K collection are currently going through analytical quality control (QC) testing for purity and identity. The structures and annotations of the Tox21 10K library as well as the QC results are publicly available [6, 7]. The compound library was plated in 1536-well plates as 15 concentration series with $\sqrt{}$5-fold dilution in dimethylsulfoxide (DMSO), covering a concentration range up to four orders of magnitude. To assess the reproducibility of data, three physical copies of the library were prepared in three different formats, in which the same compound was plated in a different well location in each copy.

A large amount of data have been generated during the phase II screening of the Tox21 10K library against a panel of cell-based assays. Counter screens are also employed to minimize interferences from off-target, assay-specific artifacts, such as compound auto fluorescence and cytotoxicity. A standardized qHTS data analysis process has been developed at NCATS to integrate the data and characterize the activities observed from these assays [21, 22]. This process includes the following:

(1) plate level data normalization and correction;
(2) concentration response curve fitting and classification;
(3) data reproducibility evaluation; and
(4) assignment of activity outcome to compounds.

After further evaluation for quality and utility by all Tox21 partners, the datasets are released to the public domain in a number of public databases including PubChem (http://pubchem.ncbi.nlm.nih.gov/), the NCATS Tox21 Browser (https://tripod.nih.gov/tox21/), the NIEHS Chemical Effects in Biological Systems (CEBS) database (http://tools.niehs.nih.gov/cebs3/ui/), and EPA's Aggregated Computational Toxicology Online Resource (ACToR) (http://actor.epa.gov).

To date, over 85 million data points generated from more than 50 assays on the Tox21 10K collection have been deposited in PubChem (see Table 14.1 for list of PubChem assay IDs and assay performance scores) [18].

14.2.2 In Vivo Toxicity Modeling

The 10K library was clustered using the self-organizing map (SOM) algorithm [23] based on either structural similarity or similarity in its members' activity profiles across all the in vitro assays. Models were built for animal in vivo toxicity endpoints

using either the structure (structure-based models) or assay activity (activity-based models) SOM clusters or both [17]. To build models using both the structure and activity SOM clusters, each compound was reassigned to a "consensus cluster" such that only compounds that belong to the same structure cluster and the same activity cluster were assigned to the same "consensus cluster." The consensus clusters were used to build the structure–activity combined models. For each SOM cluster containing the training compounds, the enrichment of toxic compounds was determined by a Fisher's exact test. The $-\log_{10}$ p-value from the Fisher's exact test was used as a measure of the toxic potential (toxicity score) of the compounds in this cluster, and evaluated as a predictor of toxicity for test compounds that fall into the same cluster. More significant p-values (larger $-\log p$-values) indicate a larger probability of toxicity. If a cluster was deficient of toxic compounds, i.e., the fraction of toxic compounds in the cluster was smaller than the fraction of toxic compounds in the whole library, the $\log_{10}$ p-value was used instead. Here we denote the toxicity scores obtained from the activity SOM as p-activity, those from the structure SOM as p-structure, and those using both the activity and structure SOMs as p-both. To test model performance, the corresponding SOM cluster or consensus cluster was located for each test set compound, and p-activity, p-structure, or p-both obtained from the training set were retrieved. These statistics were compared with the true toxicity outcome of the test compound to determine if the test compound should be counted as a true positive (TP: toxic and score > cutoff), false positive (FP: non-toxic and score > cutoff), true negative (TN: non-toxic and score $\leq$ cutoff), or false negative (FN: toxic and score $\leq$ cutoff).

Models were built for the human adverse drug effects (ADEs) using assay activity (activity-based models), compound structure (structure-based models), combinations of structure and activity data with or without drug target annotations (DTAs), and animal toxicity endpoints [19]. The Weighted Feature Significance (WFS) method previously developed at NCATS [24] was applied to construct the models. Briefly, WFS is a two-step scoring algorithm. In the first step, a Fisher's exact test is used to determine the significance of enrichment for each feature in the drugs with a certain ADE compared to the ones without such ADE reported, and a p-value is calculated for all the features present in the dataset. For assay activity data, each assay readout was treated as a feature and the feature value was set to 1 for active compounds and 0 for inactive compounds. For animal in vivo toxicity data, each toxicity endpoint was treated as a feature, and the feature value was set to 1 for toxic compounds and 0 for non-toxic compounds. For structure data, the feature value was set to 1 for drugs containing that structural feature and 0 for drugs that do not have that feature. For DTA data, each DTA was treated as a feature, and the feature value was set to 1 for drugs that reported to have that DTA and 0 for drugs that not known to have the DTA (see Table 14.2). If a feature is less frequent in the active compound set than the non-active compound set, then its p-value is set to 1. These p-values form what we call a "comprehensive" feature fingerprint, which is then used to score each drug for its potential to cause a certain ADE according to Eq. (1), where p_i is the p-value for feature i; C is the set of all features present in a drug; M is the set of features encoded in the "comprehensive" feature fingerprint (i.e., features present in at least

Table 14.2 Transformation rules of datasets to 1/0 bits for modeling

	Dataset				
Drug type	Bit	Assay activity	In vivo toxicity	Chemical structure	DTA
	1	Active	Toxic	Feature present	Target
	0	Inactive	Non-toxic	Feature not present	Not target

one drug with that ADE); N is the number of features; and α is the weighting factor, which is set to 1 in all the models described here. A high WFS score indicates a strong potential for ADE.

$$\mathrm{WFS} = \frac{\sum \log(p_i)}{\min(\log(p_i)) \times (\alpha N_{C-M} + N_{M \cap C})} \tag{1}$$

For each model, compounds were randomly split into two groups of approximately equal sizes, with one used for training and the other for testing. Randomization was conducted 100 times to generate 100 different training and test sets to evaluate the robustness of the models. Model performance was assessed by calculating the area under the receiver operating characteristic (ROC) curve (AUC-ROC), which is a plot of sensitivity [TP/(TP+FN)] versus (1-specificity [TN/(TN+FP)]) [25]. A perfect model would have an AUC-ROC of 1 and an AUC-ROC of 0.5 indicates a random classifier. The random data split and model training and testing were repeated 100 times, and the average AUC-ROC values were calculated for each model.

14.3 Results/Case Studies

14.3.1 Modeling Tox21 Data for In Vivo Toxicity Prediction

One of the Tox21 goals is to establish in vitro compound activity signatures that are predictive of in vivo toxicity. To evaluate their utility in realizing this goal, the Tox21 10K data were applied to build predictive models for 72 in vivo toxicity endpoints of mostly animal and some human origins [17]. Models were built using either the compound structure (structure-based models) or assay activity (activity-based models) self-organizing map (SOM) clusters, or both (combined models). Model performance was measured by the area under the receiver operating characteristic (ROC) curve (AUC-ROC) [25]. The premise for these models is that compounds sharing similar in vitro signatures and/or structure features are likely to show similar in vivo effects. There are five human toxicity endpoints, including standard Draize test for human skin irritation, multiple dose toxicity data (TDLo) through oral exposure from human females, human males, and humans (gender not specified), and reproductive toxicity data (TDLo) through oral exposure from human females. The activity-based models

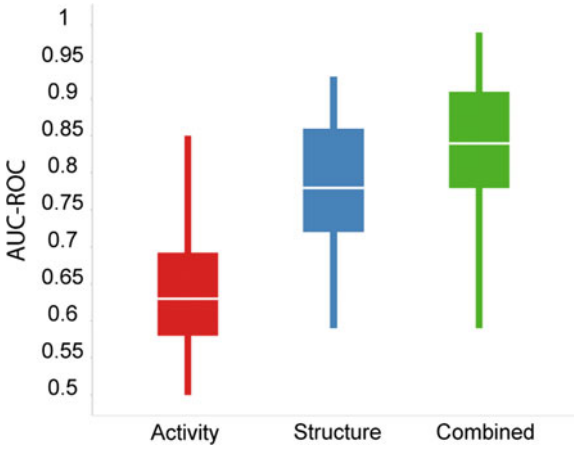

Fig. 14.1 Performance distribution of in vivo toxicity prediction models built with different datasets measured by AUC-ROC

built for the human endpoints performed significantly better than the models of the animal (mouse/rat/rabbit) toxicity endpoints. The compound structure-based models showed overall better performance than the activity-based models with an average AUC-ROC of 0.78 (Fig. 14.1), underlying the ongoing need to further expand the battery of in vitro assays. However, the performance of the structure-based models did not show any species difference. Compared to the activity-based models, the structure-based models performed significantly better for the mouse/rat/rabbit toxicity endpoints, but not as significantly for the human toxicity endpoints. Combining the compound activity and structure data resulted in significantly better models (average AUC-ROC of 0.84), than the models built with activity or structure alone (Fig. 14.1). Similar to the structure-based models, the species difference between the model performances disappeared.

14.3.2 Expanding Biological Space Coverage Improves Human Toxicity Prediction

A subset of the Tox21 10K library is composed of approved drugs [20], and human adverse effect data are publicly available for some of these drugs. To address the issue of species differences and re-evaluate the utility of the Tox21 in vitro human cell-based assay data, we collected adverse drug effect (ADE) data, a common manifest of human toxicity, and rebuilt models to predict this type of toxicity [19]. For comparison purposes, we also conducted the first meta-analysis to evaluate the performance of animal in vivo toxicity data in predicting human adverse outcomes in parallel with in vitro assay data [19]. Animal toxicity datasets do not seem to have a clear advantage over human cell-based data in predicting human in vivo effects based on these modeling results (Fig. 14.2). Models built with in vivo animal toxi-

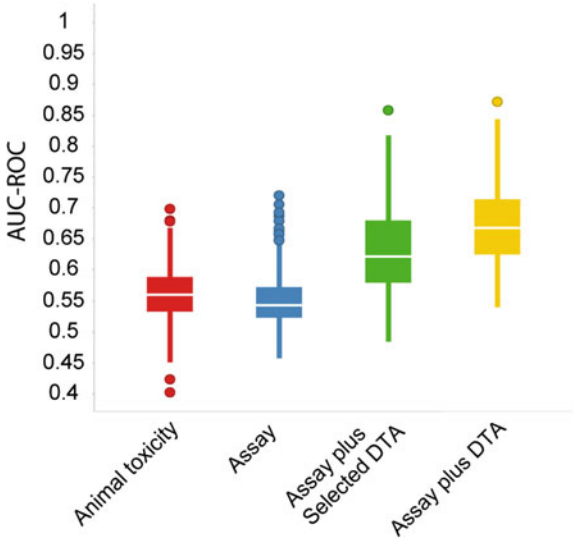

Fig. 14.2 Performance distribution of human adverse drug effect prediction models built with different datasets measured by AUC-ROC

city endpoints performed moderately (average AUC-ROC = 0.56), similar to those built with the in vitro assay data (average AUC-ROC = 0.55) for predicting ADEs in human. This result again confirms that species differences, as well as data sparsity and lack of consistency, limit the reliability of extrapolating animal in vivo toxicity data to human in vivo effects.

Similar to the animal toxicity-based models, most models built with in vitro human cell-based assay data did not show good predictive capacity of human ADEs either. This low performance may be due to the limited biological space covered by the current panel of Tox21 assays. Since many drugs in the 10K collection have target and/or mechanism of action annotations available, we collected drug target annotations (DTAs) from the literature (2370 DTAs) and combined them with in vitro assay data to build new models. These combined models showed remarkable improvements in predictive performance with average AUC-ROC for human ADE prediction increased from 0.55 to 0.67 (Fig. 14.2) [19]. In addition, we identified a small subset of 58 DTAs that contributed the most to the prediction. Adding this set of 58 DTAs to in vitro assay data significantly improved the model performance, increasing the average AUC-ROC to 0.63 for human ADE prediction (Fig. 14.2) [19]. This result shows that data on just a small set of additional DTAs (2% of the entire 2370 DTA set) can expand the biological space coverage sufficiently to produce predictive models of human toxicity when combined with in vitro assay data. While the entire DTA set improved the model performance by 22–28% on average, the selected set of 58 DTAs alone improved the model performance, on average, by 15–18%. In other words, 2% of the DTA information could account for ~70% of the improvement in the predictive capacity of the models. Most of the 58 selected targets/pathways are GPCR targets, which are important drug targets non-specific activity on which

can lead to undesirable side effects and other liabilities [26]. CYP3A4 and CYP2D6 (the two most significant cytochrome P450 (CYP) isozymes among the CYP family that are essential for drug metabolism) are also found within the set of 58 targets [27]. This selected set of targets contains several metabolic pathways, a number of cancer pathways, disease pathways, stress response pathways, and other signaling pathways. Most of these targets/pathways, including the GPCRs and CYPs, are not part of the current Tox21 suite of assays. This set of 58 targets can serve as a guide for assay development in order to generate in vitro data that can better predict human toxicity.

14.3.3 The Tox21 Data Challenge—New Methods for Data Modeling

The high-quality concentration response datasets generated by the Tox21 program on a wide spectrum of pathways and phenotypic toxicity endpoints provide a valuable resource for predictive toxicity modeling. These data can not only serve as in vitro signatures that could be used to predict in vivo toxicity endpoints [17, 19, 28–30] and to prioritize chemicals for more in depth toxicity testing [31] that help to fulfill the Tox21 goals, but also serve as a knowledge base to correlate chemical structures to their biological activities for the QSAR (quantitative structure–activity relationship) modeling community to build more robust models [24, 32]. There is no human exposure and/or hazard data for 95% of the >80,000 chemicals registered for use in the USA to inform society about their potential toxic effects [33]. In silico approaches, such as QSAR models that infer biological activity from chemical structure similarity, provide a viable alternative to fill in the experimental data gap [34, 35].

To encourage the mining and usage of the Tox21 data, NCATS launched the Tox21 Data Challenge 2014 [36], to "crowdsource" data analysis by independent researchers to obtain new models and methods that can predict the potential toxicity of compounds by disrupting cellular and biochemical pathways using chemical structure data. The Tox21 10K qHTS data from 12 assays, seven nuclear receptor and five stress response pathway assays, were selected based on data quality and hit rate for the challenge. The competition attracted 125 participants representing 18 different countries, with 378 model submissions from 40 teams received for final evaluation.

The winning models all achieved >80% accuracy (Fig. 14.3). Several models exceeded 90% accuracy. High-quality winning models serve as a confirmation of the ability of computational approaches to provide meaningful predictions of toxicity responses in terms of pathway disruption upon environmental compound exposure, and also as a validation of the quality of datasets produced from the Tox21 qHTS assays. Consensus models constructed by combining the individual models from all participating teams resulted in improved predictive performance, with some outperforming the winning models, showing the wisdom of the crowd. The winning models

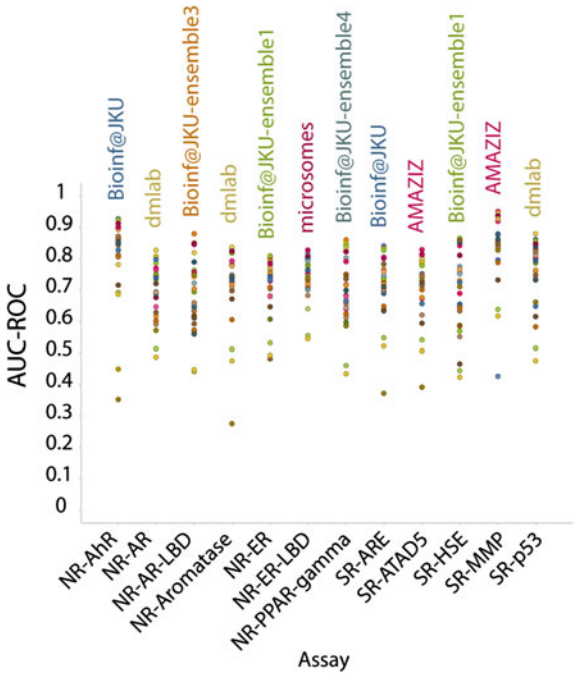

Fig. 14.3 Tox21 Challenge participant model performances measured by AUC-ROC. The winning model for each subchallenge assay is labeled with the team name

as well as methods used by some other challenge participants have been published in a special issue of the journal Frontiers in Environmental Science as Research Topic "Tox21 Challenge to build predictive models of Nuclear Receptor and stress response pathways as mediated by exposure to environmental toxicants and drugs" [37]. A number of new methods, such as associative neural networks (ASNN) with stratified bagging [38] and multi-tree ensemble (e.g., Random Forest, Extra Trees) with assorted feature selection [39], were employed by the winning teams to achieve high-performance models. In addition to the traditional machine learning methods, the grand challenge winning team applied novel Deep Learning [40] techniques to their winning models [41].

All winning models, or better performing consensus models, can be applied in parallel to establish activity/toxicity profiles for data poor environmental chemicals to obtain an estimate of their toxicity potential in a matter of hours of computational time. These computational models could become decision-making tools for government agencies in determining which environmental chemicals and drugs are of the greatest potential concern to human health. Chemicals estimated to have a high potential for toxicity, which would be a much smaller number, could be prioritized for experimental evaluation and validation. Combining these computational models with existing experimental data [7, 42] will make chemical prioritization more time and cost-efficient.

14.4 Discussion/Notes

The Tox21 10K collection has been screened against over 50 assays yielding high-quality datasets. When applied to predicting in vivo toxicity, models built with the current set of Tox21 assay activity profiles showed reasonable but less than ideal performance for most in vivo toxicity endpoints and appeared to be less predictive than the models built with chemical structures [17]. On the other hand, the assay data-based models performed markedly better in predicting human toxicity, though just a few endpoints, than animal toxicity, whereas the structure-based models did not show this species selectivity. As all of the Tox21 assays screened so far are cell-based assays constructed from human cells or cell lines, species difference may be a significant contributing factor to the less than ideal performance of the activity-based models on predicting animal toxicity endpoints. New drugs that passed animal testing are often known to fail in human clinical trials because of lack of effect or unexpected toxicity [43]. Other studies also showed that animal data only predicted human outcomes approximately half of the time [44]. Our analysis revealed that animal toxicity-based models did not perform significantly better than assay activity-based models in predicting human adverse drug effects, providing further evidence for this species-related issue [19]. This exemplifies the need to have in vivo human toxicity data, i.e., clinical toxicity data presently not readily available to the public, in order to better assess the predictive value of the human in vitro assay data. Nevertheless, combining activity data with structure information significantly improved the model performance for most of the in vivo endpoints, which served as a validation of the value of the in vitro assay data when applied to in vivo toxicity prediction. The model performance in addition highlighted the importance of data quality. The Tox21 in vitro qHTS data showed good reproducibility (Table 14.1) [17]. The in vivo data used in modeling were also evaluated for reproducibility, and a significant positive correlation was found between the reproducibility of the in vivo endpoint and the performance of the model built for that endpoint, suggesting that the performance of in vivo toxicity prediction models could be further improved if better quality in vivo data were available.

As the current Tox21 assays focused primarily on nuclear receptor signaling and stress response pathways, all aspects of biology involved in toxic response are not covered sufficiently, indicating the need to expand the coverage of the biological space by including assays that target additional pathways relevant for toxicity. As surrogates of assay data, adding drug target information significantly improved the performance of Tox21 in vitro assay data-based models in predicting human ADEs [19]. These DTAs have good coverage of the drug target space known in the literature and can be considered validated experimental or assay data, and thus produced good predictive models even with a small selected subset that provides sufficient expansion of the biological space. In addition to limited target space coverage, the current assay data used for modeling is primary HTS data without further validation and thus undoubtedly confounded with noise and assay artifacts. These results again highlight the importance of data quality and selecting the right assays. Validated DTA

data seem to be the best choice for ADE or human in vivo toxicity prediction. The DTA-based models, however, cannot be applied to predict new compounds without such annotations available. It is therefore important to generate high-quality assay data with good coverage of the biological space and validation of these datasets.

14.5 Conclusions and Future Directions

The Tox21 program is a multiagency federal collaboration to advance in vitro toxicological testing in the twenty-first century. During phase II, the production phase, a battery of in vitro assays with target-specific and mechanism-based readouts has been developed, validated, and adapted to a quantitative high-throughput screening platform. The Tox21 10K compound collection has been successfully screened against a panel of over 50 nuclear receptor and stress response pathway assays, generating over 85 million publicly available data points as a rich resource for toxicology.

These high-quality datasets have proven instrumental in identifying mechanisms of compound toxicity and developing models for predicting in vivo toxicity response. While in vitro assay data alone showed limited predictive power of adverse human effects, complementing the biological space coverage with additional targets, in the continuation of the Tox21 program, showed promise to significantly improve the performance of the assay data-based models resulting in robust models for human toxicity prediction. All results provide rich datasets to researchers for further data mining, generation of new hypotheses, and developing new methods for activity modeling. The predictive computational models generated from these high-quality datasets can help shed light on the potential of using in vitro assays as an alternative approach for assessing chemical toxicity.

References

1. NTP (2014) Current directions and evolving strategies
2. Collins FS, Gray GM, Bucher JR (2008) Toxicology. Transforming environmental health protection. Science 319(5865):906–907
3. Kavlock RJ, Austin CP, Tice RR (2009) Toxicity testing in the 21st century: implications for human health risk assessment. Risk Anal 29(4):485–487 (Discussion 492–487)
4. NRC (2007) Toxicity testing in the 21st century: a vision and a strategy. In: Council NR (ed), The National Academies Press, Washington, DC
5. Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121(7):756–765
6. PubChem (2013) Tox21 phase II compound collection [updated 2013; cited 4 Dec 2013]. Available from http://www.ncbi.nlm.nih.gov/pcsubstance/?term=tox21
7. NCATS (2016) Tox21 data browser [cited 2016]. Available from https://tripod.nih.gov/tox21/
8. Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, Austin CP, Shinn P, Simeonov A, Tice RR, Xia M (2013) The Tox21 robotic platform for the assessment of environmental chemicals—from vision to reality. Drug Discov Today 18(15–16):716–723

9. Hsu CW, Zhao J, Huang R, Hsieh JH, Hamm J, Chang X, Houck K, Xia M (2014) Quantitative high-throughput profiling of environmental chemicals and drugs that modulate farnesoid X receptor. Sci Rep 4:6437. https://doi.org/10.1038/srep06437
10. Huang R, Sakamuru S, Martin MT, Reif DM, Judson RS, Houck KA, Casey W, Hsieh JH, Shockley KR, Ceger P, Fostel J, Witt KL, Tong W, Rotroff DM, Zhao T, Shinn P, Simeonov A, Dix DJ, Austin CP, Kavlock RJ, Tice RR, Xia M (2014) Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. Sci Rep 4:5664. https://doi.org/10.1038/srep05664
11. Huang R, Xia M, Cho MH, Sakamuru S, Shinn P, Houck KA, Dix DJ, Judson RS, Witt KL, Kavlock RJ, Tice RR, Austin CP (2011) Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ Health Perspect 119(8):1142–1148
12. Lynch C, Zhao J, Huang R, Kanaya N, Bernal L, Hsieh JH, Auerbach SS, Witt KL, Merrick BA, Chen S, Teng CT, Xia M (2018) Identification of estrogen-related receptor alpha agonists in the Tox21 compound library. Endocrinology 159(2):744–753
13. Lynch C, Sakamuru S, Huang R, Stavreva DA, Varticovski L, Hager GL, Judson RS, Houck KA, Kleinstreuer NC, Casey W, Paules RS, Simeonov A, Xia M (2017) Identifying environmental chemicals as agonists of the androgen receptor by using a quantitative high-throughput screening platform. Toxicology 385:48–58
14. Attene-Ramos MS, Huang R, Michael S, Witt KL, Richard A, Tice RR, Simeonov A, Austin CP, Xia M (2015) Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. Environ Health Perspect 123(1):49–56
15. Nishihara K, Huang R, Zhao J, Shahane SA, Witt KL, Smith-Roe SL, Tice RR, Takeda S, Xia M (2015) Identification of genotoxic compounds using isogenic DNA repair deficient DT40 cell lines on a quantitative high throughput screening platform. Mutagenesis 31(1):69–81
16. Witt KL, Hsieh JH, Smith-Roe SL, Xia M, Huang R, Zhao J, Auerbach SS, Hur J, Tice RR (2017) Assessment of the DNA damaging potential of environmental chemicals using a quantitative high-throughput screening approach to measure p53 activation. Environ Mol Mutagen 58(7):494–507
17. Huang R, Xia M, Sakamuru S, Zhao J, Shahane SA, Attene-Ramos M, Zhao T, Austin CP, Simeonov A (2016) Modelling the Tox21 10K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat Commun 7:10425
18. PubChem (2016) Tox21 phase II data 2016 [cited 16 Nov 2013]. Available from http://www.ncbi.nlm.nih.gov/pcassay?term=tox21
19. Huang R, Xia M, Sakamuru S, Zhao J, Lynch C, Zhao T, Zhu H, Austin CP, Simeonov A (2018) Expanding biological space coverage enhances the prediction of drug adverse effects in human using in vitro activity profiles. Sci Rep 8(1):3783
20. Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, Nguyen DT, Austin CP (2011) The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. Sci Transl Med 3(80):80ps16
21. Huang R (2016) A quantitative high-throughput screening data analysis pipeline for activity profiling. In: Zhu H, Xia M (eds) High-throughput screening assays in toxicology. Methods in molecular biology, vol 1473. Humana Press
22. Wang Y, Huang R (2016) Correction of microplate data from high throughput screening. In: Zhu H, Xia M (eds) High-throughput screening assays in toxicology. Methods in molecular biology, vol 1473. Humana Press
23. Kohonen T (2006) Self-organizing neural projections. Neural Networks Official J Int Neural Network Soc 19(6–7):723–733

24. Huang R, Southall N, Xia M, Cho MH, Jadhav A, Nguyen DT, Inglese J, Tice RR, Austin CP (2009) Weighted feature significance: a simple, interpretable model of compound toxicity based on the statistical enrichment of structural features. Toxicol Sci 112(2):385–393
25. Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39(4):561–577
26. Allen JA, Roth BL (2011) Strategies to discover unexpected targets for drugs active at G protein-coupled receptors. Annu Rev Pharmacol Toxicol 51:117–144
27. Lynch T, Price A (2007) The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. Am Fam Physician 76(3):391–396
28. Martin MT, Knudsen TB, Reif DM, Houck KA, Judson RS, Kavlock RJ, Dix DJ (2011) Predictive model of rat reproductive toxicity from ToxCast high throughput screening. Biol Reprod 85(2):327–339
29. Sipes NS, Martin MT, Reif DM, Kleinstreuer NC, Judson RS, Singh AV, Chandler KJ, Dix DJ, Kavlock RJ, Knudsen TB (2011) Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. Toxicol Sci 124(1):109–127
30. Sipes NS, Wambaugh JF, Pearce R, Auerbach SS, Wetmore BA, Hsieh JH, Shapiro AJ, Svoboda D, DeVito MJ, Ferguson SS (2017) An intuitive approach for predicting potential human health risk with the Tox21 10K library. Environ Sci Technol 51(18):10786–10796
31. Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, Reif DM, Rotroff DM, Shah I, Richard AM, Dix DJ (2010) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485–492
32. Sun H, Veith H, Xia M, Austin CP, Tice RR, Huang R (2012) Prediction of cytochrome P450 profiles of environmental chemicals with QSAR models built from drug-like molecules. Mol Inform 31(11–12):783–792
33. Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E (2009) The toxicity data landscape for environmental chemicals. Environ Health Perspect 117(5):685–695
34. Muster W, Breidenbach A, Fischer H, Kirchner S, Muller L, Pahler A (2008) Computational toxicology in drug development. Drug Discov Today 13(7–8):303–310
35. Vedani A, Smiesko M (2009) In silico toxicology in drug discovery—concepts based on three-dimensional models. Altern Lab Anim 37(5):477–496
36. Huang R, Xia M, Nguyen D-T, Zhao T, Sakamuru S, Zhao J, Shahane SA, Rossoshek A, Simeonov A (2016) Tox21 challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. Front Environ Sci 3(85):1–9
37. Huang R, Xia M (2016) Research topic: Tox21 challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental toxicants and drugs. Front Environ Sci 2954
38. Abdelaziz A, Spahn-Langguth H, Schramm K-W, Tetko IV (2016) Consensus modeling for HTS assays using in silico descriptors calculates the best balanced accuracy in Tox21 challenge. Front Environ Sci 4(2):1–12
39. Barta G (2016) Identifying biological pathway interrupting toxins using multi-tree ensembles. Front Environ Sci 4:52
40. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
41. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. Front Environ Sci 3(80):1–15
42. USEPA (2017) ToxCast data. Available from http://www2.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data
43. FDA (2004) Innovation or stagnation: challenge and opportunity on the critical path to new medical products
44. Martic-Kehl MI, Schibli R, Schubiger PA (2012) Can animal data predict human outcome? Problems and pitfalls of translational animal research. Eur J Nucl Med Mol Imaging 39(9):1492–1496

Ruili Huang is the informatics group leader on the toxicity profiling team at the NIH National Center for Advancing Translational Sciences (NCATS). Dr. Huang and her group contribute to quantitative high-throughput screening (qHTS) data processing and interpretation, and development and implementation of software tools and algorithms that facilitate NCATS' data pipeline. As part of the Tox21 program, her group focuses on evaluating qHTS assay performance for prioritization, and analyzing compound toxicity profiling data to generate hypotheses on compound mechanisms of toxicity and to build predictive models for in vivo toxicity. Dr. Huang received her Ph.D. in chemistry from Iowa State University, trained as a computational biologist at the National Cancer Institute, and joined NCATS in 2006.

Chapter 15
In Silico Prediction of the Point of Departure (POD) with High-Throughput Data

Dong Wang

Abstract Determining the point of departure (POD) is a critical step in chemical risk assessment. Current approaches based on chronic animal studies are costly and time-consuming while being insufficient for providing mechanistic information regarding toxicity. Driven by the desire to incorporate multiple lines of evidence relevant to human toxicology and to reduce animal use, there has been a heightened interest in utilizing transcriptional and other high-throughput assay endpoints to infer the POD. In this review, we outline common data modeling approaches utilizing gene expression profiles from animal tissues to estimate the POD in comparison with obtaining PODs based on apical endpoints. Various issues in experiment design, technology platforms, data analysis methods, and software packages are explained. Potential choices for each step are discussed. Recent development for models incorporating in vitro assay endpoints is also examined, including PODs based on in vitro assays and efforts to predict in vivo PODs with in vitro data. Future directions and potential research areas are also discussed.

Keywords High-throughput assays · Microarrays · Point of departure · Predictive modeling · RNAseq · Toxicogenomics · Transcriptional profiling

Abbreviations

AC_{50}	Half-maximal effective concentration
AIC	Akaike information criterion
AOP	Adverse outcome pathway
BMD	Benchmark dose
BMDL	A statistical lower bound of BMD

D. Wang (✉)
Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration (FDA), Jefferson, AR, USA
e-mail: dong.wang@fda.hhs.gov

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_15

BMR	Benchmark risk
BRBZ	Bromobenzene
C_{max}	Peak plasma concentration
EPA	Environmental Protection Agency
EU	European Union
HCI	High content imaging
HZBZ	Hydrazobenzene
IVIVE	In vitro-in vivo extrapolation
KDMM	Kernel density mean of M-component
KE	Key event
KER	Key event relationship
LOAEL	Lowest-observed-adverse-effect level
MDMB	4,4′-Methylenebis (*N*,Ndimethyl) benzenamine
MIE	Molecular initiating event
MOA	Mode of action
MSigDB	Molecular Signature Database
NDPA	*N*-Nitrosodiphenylamine
NOAEL	No-observed-adverse-effect-level
POD	Point of departure
REACH	Registration, evaluation, authorization and restriction of chemical substances
RMA	Robust Multi-array Average normalization method
RPKM	Reads per kilobase per million mapped reads
TG-GATEs	Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System
TLR	Target learning region
TRBZ	1,2,4-Tribromobenzene
TTCP	2,3,4,6-Tetrachlorophenol

15.1 Introduction

The determination of the point of departure (POD) is an essential step for chemical risk assessment. Currently, the gold standard in POD estimation is chronic animal dosing studies. Rats or mice are commonly used for this purpose, though other animal species might also be utilized. Though this approach has been essential for the current toxicological testing regime, some shortcomings have been noted. One issue is that high-dose animal testing often poorly predicts human toxicity. Even when it provides meaningful results, it is difficult to obtain a mechanistic understanding of biological pathways leading to adverse effects from the organism level toxicological response as measured by apical endpoints. As a result, there has been a strong interest in incorporating multiple lines of evidence relevant to human toxicology in order to generate a more detailed understanding of toxicological properties for chemical risk assessment. Toxicogenomics and other high-throughput assays as well as in silico

modeling approaches are expected to play critical roles for this purpose. This vision was elaborated in several important publications, including two National Research Council reports—Toxicity Testing in the 21st Century [1] and Using 21st Century Science to Improve Risk-Related Evaluations [2].

The requirements regarding cost and speed provide another impetus to adopt high-throughput assays and in silico methods. There is a huge backlog of chemicals to be evaluated by regulatory agencies around the world. Chronic animal experiments alone are too costly and time-consuming to deal with this problem efficiently. On the other hand, genomics and high-throughput assays can potentially provide a comprehensive picture of perturbed pathways, which can then be used to guide targeted testing. The significant societal interest for reducing animal testing also calls for greater use of innovative assay methods and in silico modeling approaches [3]. The same principle has been advocated by the EU program for registration, evaluation, authorization and restriction of chemical substances (REACH) program and EU Cosmetic Directive [4].

Right now, it is very common to generate transcriptomic profiles for chemicals under consideration using either microarrays or next generation sequencing technology (RNAseq) to provide insights for toxicological mechanisms. An example of systematic data generation efforts is the Open TG-GATEs (Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System [5]), through which a large-scale database consisting of data for gene expression and pathology from both animal-tissue- and cell-line-based experiments on 170 compounds has been generated. It provides an excellent source for exploring transcriptomic changes caused by some important chemicals.

In the area of high-throughput in vitro assays, several large projects are ongoing to develop high-throughput-cell-based- or cell-free tests for toxicological evaluations and to establish data repositories with a diverse collection of chemicals. The Tox21 program and EPA's ToxCast are two important projects in this field. Phase I of the Tox21 program [6, 7] focused on more than 50 assays regarding cytotoxicity, mitochondrial toxicity, cell signaling, DNA damage, nuclear-receptor activation, among others; with testing on more than 2800 chemicals completed. The ToxCast program [8, 9] examines high-throughput assays covering a range of cell responses and over 300 signaling pathways [10]. More than 2000 chemicals have been evaluated in ToxCast Phase I and Phase II. Both the Tox21 and ToxCast programs have generated data in dose response format for chemical-endpoint combinations. Various models for hazard identification and prioritization for screening have been developed using these datasets [11–13].

Another important development is the coordinated effort to characterize adverse outcome pathways (AOPs, [14, 15]). An AOP links a molecular initiating event (MIE) to the adverse outcome (AO) via a series of key events (KEs), which is specified by key event relationships (KERs). Mechanistic information can thus be connected to apical endpoints in a formalized, quality-controlled, and transparent way. As MIEs and KEs are often associated with certain genes, proteins, or metabolites; AOPs provide a valuable framework to describe the biological context for mechanistic information regarding transcriptomic and in vitro assays. Here, it is also useful to

mention another concept, mode of action (MOA), which is similar in principle to AOP but with different emphasis [14]. Most discussions in this review are applicable to both AOP and MOA, and both terms are used in the literature cited.

Though our knowledge about mechanistic information for toxicity is far from complete, substantial efforts have been made to incorporate genomic and high-throughput assays into risk assessment with some notable successes. In this review, we will focus on the problem of obtaining estimates for PODs with genomic and high-throughput assay endpoints. Compared to hazard identification, determining the POD requires the characterization of the dose response relationship regarding the chemical under evaluation. If the dose response experiment has been performed under settings different from the targeted POD (e.g., in vitro assays relative to in vivo PODs), then there is the added question how to translate the concentration between different settings. For this, using transcriptomic data to infer in vivo POD with the same animal tissue is the most straightforward. The seminal paper by Thomas et al. [16] pioneered the strategy of using the most sensitive pathways, which has been further studied by a number of research groups. In the next section, we will review current methods for inferring in vivo PODs with transcriptomic data from the same animal tissues. PODs derived from in vitro assays will be then discussed, followed by a review of recent work using in vitro assay endpoints to predict in vivo PODs. We will conclude with some general discussions. Due to the size of the literature in this field, we will not cover every aspect of in silico modeling of PODs in detail. Rather, a general outline will be provided regarding the most often used approach in this area. Though the references cited do not constitute a complete list of relevant literature, readers can use them as a starting point for further reading.

15.2 Infer In Vivo PODs with Transcriptomic Data from the Same Tissues

Inferring the in vivo PODs with transcriptomic profiles from the same target tissue (usually from rats or mice) for apical endpoints is the most mature approach discussed in this review. Conceptually, a toxic chemical will trigger expression changes that underpin both direct and indirect toxicity responses. There exists significant literature comparing transcriptionally derived PODs with PODs based on apical endpoints. The general finding is that transcriptional PODs are usually consistent with apical-endpoint-based values. Since the work of Thomas et al. [16], a number of authors have proposed variations with this approach. To summarize these ideas, we will first review the approach of Thomas et al. [17] as an example to illustrate various steps taken in this type of studies. Then, we will discuss in detail important considerations and possible choices in each step.

15.2.1 An Example from Thomas et al. [17]

Thomas et al. [17] discussed the temporal concordance between apical and transcriptional PODs for several chemicals. The chemicals are 1,2,4-Tribromobenzene (TRBZ), 2,3,4,6-tetrachlorophenol (TTCP), Bromobenzene (BRBZ), 4,4′-Methylenebis (*N*,Ndimethyl) benzenamine (MDMB), Hydrazobenzene (HZBZ), and N-Nitrosodiphenylamine (NDPA). A specific strain and sex of rat as well as route of exposure were chosen for each chemical. Rats were exposed to each chemical at five dose levels for 5 days and 2, 4, or 13 weeks. Rats were randomly assigned to each dose groups with ten rats per group. Liver is the target tissue for TRBZ, BRBZ, TTCP, and HZBZ. Bladder is the target tissue for MDMB while thyroid is the target tissue for NDPA.

After exposure, the target tissue was harvested for both histological evaluation and transcriptional profiling. Typically, ten rats were evaluated per concentration per time point for histological changes. RNA was isolated from six rats per dose per time points. After purification, RNA from five rats per concentration per time point with the best quality was used for microarray analysis using the Affymetrix HT RG-230 PM Array Plate.

Transcriptional POD analysis. The transcriptional POD was determined with the benchmark dose (BMD) approach. Genewise expression levels were analyzed using BMDExpress (v. 1.41, [18]), which enables automatic model selection and integration with biological pathways. Here, it is worthwhile to detail the modeling parameters used in this study while different potential choices will be discussed in the next subsection. In this study, the microarray data were $\log_2$ transformed and normalized with the Robust Multi-Array Average normalization method (RMA [19]). The normalized intensity values were then fit into four different dose-response models with BMDExpress: linear, two-degree polynomial, three-degree polynomial, and power models. The BMD was calculated as the dose where the estimated response is 1.349 times the standard deviation of the response at dose zero. A statistical lower bound estimate of a confidence interval for the BMD (BMDL) was also derived. To select a single model for POD determination, the likelihood ratio test (for nested models: linear, 2-degree polynomial, 3-degree polynomial) and the Akaike information criterion (AIC) were used. The best fitting model was used to calculate the BMD and BMDL. To avoid the effect of probe sets with poorly fitting models, it was further required that the BMD value to be lower than the highest dose and the goodness-of-fit p-value <0.01.

Once the BMD and BMDL were calculated for each probe set, they were aggregated at the pathway level. Probe sets were mapped to unique genes. Those mapped to multiple genes were removed from analysis. For genes represented by more than one probe set, BMDs and BMDLs were averaged to derive values for the gene. Gene identifiers were matched to pathways using the GeneGo Metacore database. Pathways with fewer than five genes with BMDs of required quality were removed from analysis. Median values for BMDs and BMDLs for each pathway were used as pathway level values.

Comparison to PODs derived from apical endpoints. As the focus of this review is on genomic and high-throughput assays, POD analysis using apical endpoints will not be discussed in detail. For more information, readers can refer to [17] and the references therein. In general, the transcriptional BMDs based on the most sensitive pathways are quite consistent with apical BMD values (Fig. 15.1). The difference is less than twofold on average. The transcriptional BMDs are also relatively stable over different time points, though the most sensitive pathways are not consistent across time points.

The results reported in [17] are typical for studies comparing transcriptome-based PODs with the in vivo counterpart. As the same tissues are used both for RNA extraction and pathological evaluations, the PODs are directly comparable. A number of studies have reported good concordance between transcriptional PODs and in vivo

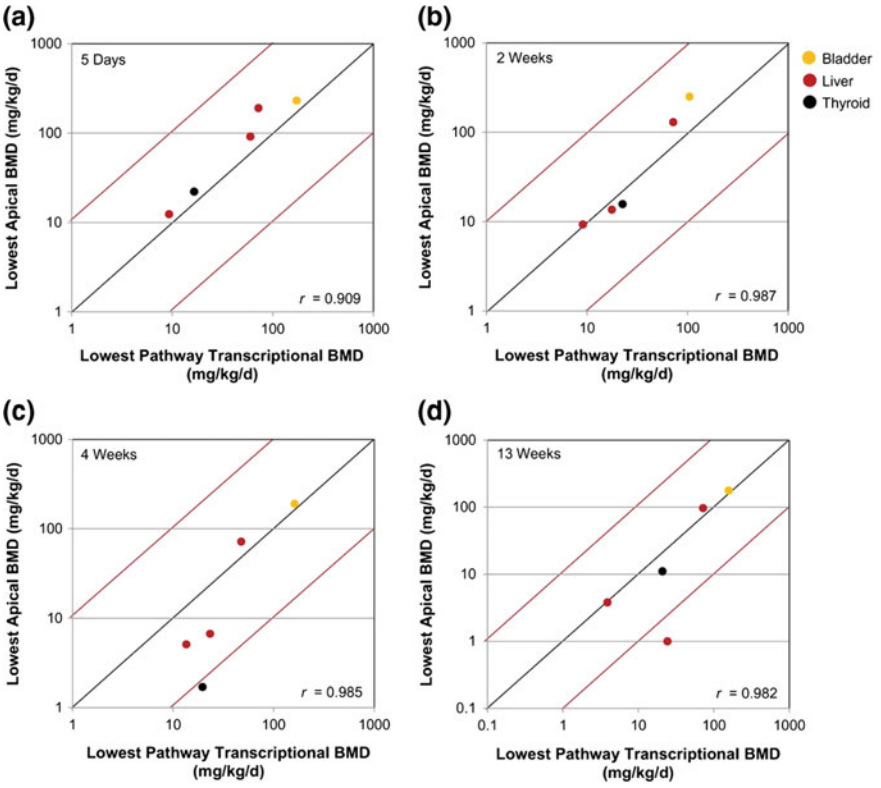

Fig. 15.1 Relationship between BMD values for noncancer apical endpoints and transcriptional BMD values for the most sensitive signaling pathway in [17] at **a** 5 days, **b** 2 weeks, **c** 4 weeks, and **d** 13 weeks of exposure. Data points are colored based on the target tissue. The black line dissecting the graph indicates equal apical and transcriptional BMD values. The red lines represent tenfold difference between the BMD values. This figure is adapted from [17] with permission from Toxicological Sciences, Oxford University Press; courtesy of Society of Toxicology

PODs using the most sensitive pathway approaches. However, different choices can be made for both experimental and modeling aspects for obtaining transcriptional PODs. Researchers should consider carefully how to proceed with these types of studies based on a number of considerations. In the next subsection, we will summarize some important issues to consider when inferring PODs from transcriptomic data.

15.2.2 Important Issues for Inferring PODs from Transcriptomic Data

After the work of Thomas et al. [16], a number of researchers have proposed variations to their approach. Though some systematic comparisons have been carried out and recommendations are given (e.g., [20]), a consensus has not been reached on all issues. In this review, we do not try to give comprehensive guidance for every question that might arise from a study. Rather, we will present a list of important issues and possible choices for solutions. Fortunately, experience so far in this field suggests that reasonable results can often be reached via different variations in approaches.

Experiment designs and technology considerations. Principles for designing dose response experiments aimed at determining PODs are well known. Though multiple approaches exist for determining PODs [e.g., lowest-observed-adverse-effect levels (LOAELs), no-observed-adverse-effect-levels (NOAELs)], generally the BMD approach is used for transcriptional PODs. The advantage of the BMD approach is that it can utilize all data points to obtain more stable results [21]. To accurately model the BMD, it is ideal to have doses covering the whole range of the dose response curve. However, since there are thousands of genes, it is impractical to expect the doses are ideal for modeling all genes. Usually, the doses are chosen based on knowledge about apical endpoints.

Besides doses, studies reported so far often include samples taken at multiple time points after the dosing started. This helps shed light on whether the transcriptional pattern changes with time. Reported studies show that the transcriptional PODs tend to be relatively stable even if the most sensitive pathways are not in a duration of several weeks. Considering the high cost associated with long-dosing periods, it is thus reasonable to focus on short-dosing periods (in a matter of days) when budget is limited. Due to cost considerations, the number of animals evaluated per dose per time point for transcription is often small (as low as three). It is suggested that more emphasis on the number of doses over the sample size at each dose might be preferable [22] when operating under budget constraints.

Another important factor to consider is technology. Though most studies published so far have used microarrays (usually from Affymetrix), RNAseq is expected to become more popular due to decreased cost and the improvement in quality. In the broad area of transcriptional profiling, RNAseq has been shown to provide higher precision, a great dynamic range, and the ability to detect novel transcripts. But

as RNAseq results in read count data rather than fluorescence intensity, different processing and normalization procedures will be required relative to microarrays. From studies done so far, RNAseq data can be used to model transcription-based PODs (e.g., [23]). Though the BMD value can be quite different between whether it is based on RNAseq or microarrays at the gene and pathway level, the PODs from most sensitive pathways are usually consistent [24]. As more studies based on RNAseq become available, we should have deeper appreciation of the strengths and special characteristics for RNAseq in the context of POD modeling. In this review, we will focus more on microarray data due to its usage in the majority of studies. We will mention specific aspects of RNAseq data when needed.

Data processing. As common for high-throughput data analysis, extensive preprocessing for both microarray and RNAseq data is usually needed before data can be used for BMD modeling. As mentioned, Affymetrix GeneChip is the platform of choice for most dose response microarray experiments. It is standard practice to transform the intensity value for each probe to $\log_2$ scale and perform RMA normalization. RMA has been reported to remove nonbiological effects between microarrays. Though other normalization methods exist for microarrays, they have not been extensively applied in the transcriptional modeling of PODs. For RNAseq data, the choice is not as clear. RNAseq experiments produce data as counts of short reads for each gene (after alignment and mapping). Commonly used software for differential expression analysis using RNAseq often models the count data directly using Poisson or negative binomial distributions. Popular software packages include edgeR [25], DESeq [26], and others. For BMD modeling, however, it is most convenient if the data can be analyzed by BMDExpress [18], which we will discuss later. As BMDExpress was originally developed for microarray data, this means transforming the RNASeq read count to a continuous variable, similar to microarray intensity measurement. One commonly used approach is to calculate reads per kilobase per million mapped reads (RPKM) followed by $\log_2$ transformation. Other methods like kernel density mean of M-component (KDMM) [24] have also been recommended. With more dose response experiments being carried out with the RNAseq technology, we should gain a better understanding of the appropriate data processing approaches in this setting. Once the transformation has been performed, the quantities like log-transformed RPKM are generally treated the same way as microarray intensity readings in dose response modeling.

Modeling BMDs (BMDLs) at the gene level. For each gene, the normalized expression level from microarrays or RNAseq can be used to derive a gene-level transcriptional BMD (BMDL) as other continuous endpoints like organ weight. A general discussion on benchmark calculations for continuous data can be found in [27]. The BMD is defined as the dose where the fitted dose response curve exceeds the background noise level seen in control samples. As there are usually only a limited number of doses tested, often several different models can fit the data reasonably well. Commonly used models are linear, polynomial models of different degrees, exponential, and Hill models. The preferred model can be selected using statistical criteria, while some researchers advocate averaging results from multiple models. Given that there are thousands of genes, performing model fitting and subsequent

pathway analysis can require significant efforts. Fortunately, this difficulty is alleviated by the availability of BMDExpress [18], which automated much of the work of gene-level modeling and subsequent analysis.

Since its introduction, BMDExpress has been widely adopted for modeling transcriptional BMDs. With further development by a team led by Scott S Auerbach of National Toxicology Program, BMDExpress 2.0 has been released with enhancements (https://github.com/auerbachs/BMDExpress-2/wiki). The work of Thomas et al. [17] described above provides a typical workflow of using BMDExpress. The user has the option to filter genes according to some criteria of differential expression (e.g., *p*-values from ANOVA). If this is to be done, the threshold should be sufficiently low in order not to exclude genes with real signal. The BMD is usually set as the dose where response exceeds the baseline level by a value of some constant times the standard deviation (s.d.) from the control samples. Usually, 1 s.d. or 1.349 s.d. are used, see [27] for explanation with regard to the relationship with benchmark risk (BMR). BMDExpress can fit several commonly used dose response models. Especially with the version 2.0, users can choose from linear, 2–4-degree polynomial, 2–5-parameter exponential, power, and Hill models. If only nested models are chosen (e.g., linear and polynomial models), the likelihood ratio test can be used to choose the preferred model; otherwise, the model with the lowest AIC is preferable. Once the preferred model is chosen, the probeset-wise BMD (BMDL) can be determined.

As BMDExpress was developed for microarray data, the probeset level results can be automatically mapped to gene identifiers. For RNAseq data, one will need to obtain read counts at the gene level and perform normalization as discussed before. Then, data can be fed into BMDExpress. Though good results have been reported using this approach (e.g., [23]), Black et al. [24] note different behaviors for genes with high expression and low expression levels. This is likely due to the well-known mean–variance relationship of read count data (genes with high mean expression levels tend to have high variance). Going forward, it might be sensible to incorporate the `voom` approach (in the `limma` Bioconductor package, [28]) in dose response modeling by introducing precision weights based on the predicted variance. Of course, a significant amount of work will be involved to produce a software package for easy practical use.

BMD modeling at the pathway or gene category level. The expression level for a single gene is notoriously noisy. Thus, integrating information from a biological pathway or gene category is more preferable in deriving the PODs with transcription data. This obviously leads to the question of how to define pathways or gene categories. Currently, there are a number of sources providing extensive listings of biological pathways or gene categories, mostly compiled through literature reviews to link genes with biological functions. BMDExpress 2.0 provides analysis with Gene Ontology (http://www.geneontology.org/), Reactome (http://www.reactome.org/), or categories defined by the user. In published studies, Ingenuity Pathways (QIAGEN), the Molecular Signature Database (MSigDB) [29], and other sources have also been used. Usually, only pathways with a certain number of genes (e.g., >5) with BMDs meeting quality requirements will be analyzed. Though it is common to use the

median or mean BMD of genes in the pathway to represent the pathway-wise BMD, other quantiles or summary statistics can also be used.

Determining the transcriptional PODs. After pathway-wise BMDs have been calculated, one needs to summarize them into a single value for the POD. The first question is whether all pathways in the consideration (use the most sensitive pathway or some other quantiles) or only pathways relevant to an AOP should be used. If there is solid knowledge linking the chemical under consideration with one or more AOPs, considering pathways or gene sets associated with the relevant AOPs will reduce noise and anchor the transcriptional PODs with specific biological mechanisms, which may lead to more accurate estimate. However, for a large number of chemicals, the existing knowledge is not sufficient to clearly identify relevant AOPs; though it might be possible to use transcriptomic data to construct biologically relevant pathways and potential AOPs (see [30–32] and others for more discussion) while estimating PODs. The alternative is to use an AOP-independent approach (e.g., relying on the most sensitive pathway), which can give valuable information on PODs without detailed knowledge of AOPs. This is especially true if the toxicity effects of the chemical are broad and disturbing many different pathways.

Early efforts using the AOP-independent approach have focused on the most sensitive pathways, i.e., using the lowest BMD or BMDL from all pathways (after filtering) as the transcriptomic POD. Though this approach has been shown to be successful with a number of studies, other alternatives can also be considered. Farmahin et al. [20] compared 11 different approaches for deriving transcriptomic PODs with different criteria for gene and pathway selection. They concluded that three approaches have the best performance: (1) using the 20 significantly enriched pathways with the lowest BMDs, (2) using the 20 genes with the largest fold changes relative to controls, and (3) using the 20 genes that contribute to the greatest number of enriched pathways. Reassuringly, other approaches, including the most sensitive pathway approach, usually also give reasonable results. From the author's experience, the most sensitive pathway approach is sometimes unstable when the number of time points and sample size is small, and the tenth percentile of all pathways can give good results. Thus, trying multiple approaches in determining the POD should be beneficial.

15.3 PODs Based on In Vitro Assays

Discussions so far have focused on transcriptional profiling with the same tissues used to derive apical endpoints. This provides the most direct comparison of transcriptional PODs with the apical endpoint based counterpart. On the other hand, a significant amount of data has become available using in vitro assays. The ToxCast project (including Tox21 assays) generated data for a diverse array of chemicals using a dose response format that usually involves a range of doses in triplets. ToxCast provides a specialized software package (the R package tcpl, [33]) to fit dose response curves with constant, Hill, or gain-loss models. It is therefore feasible to perform BMD-type

modeling with ToxCast or Tox21 data, see [34] for an example. The POD derived this way is concerning the dose in an in vitro assay environment. Shah et al. [35] studied the use of ToxCast data to reconstruct dynamic cell-state trajectories and estimate the in vitro POD. In this study, the authors evaluated the effects of 967 chemicals in multiple doses on HepG2 cells over a 72-h exposure period using high content imaging (HCI). For each chemical, HCI endpoints including various protein readings, mitochondrial properties, cell cycle indicators, and other cell properties were used to define a cell-state trajectory. It is posited that if the effect of the chemical is not intolerable, the cells tend to recover to their original states after a period of perturbation. Tipping points were identified as concentration-dependent transitions in system recovery, beyond which the potential for recovery will be lost. The authors argue that the tipping point can serve as a point of departure to provide information about the effects of new chemicals and about critical concentrations at which cellular responses fail to recover to the pre-perturbation levels. This is potentially useful for screening a large number of chemicals for prioritization.

15.4 Predict In Vivo PODs with In Vitro Assays

Though defining in vitro PODs as in [35] has significant potential for prioritization in screening for a large number of chemicals, it is often of interest to predict in vivo PODs. If this can be done directly with in vitro assay data, it will greatly advance the vision of Tox21. To do this, however, one needs to relate the concentration used in vitro assays to the oral dose in animals or humans. One approach is to use in vitro-in vivo extrapolation (IVIVE) techniques based on toxicokinetics. Sipes et al. [36] discussed a strategy in relating the peak plasma concentration ($C_{\max}$) to the half-maximal effective concentration (AC_{50}) from in vitro assays to assess chemical–biological interactions. The $C_{\max}$ value can be related to oral dose with models for toxicokinetics, which is implemented in the R package HTTK [37]. In principle, a similar approach can be applied with in vitro PODs instead of AC_{50}, though it has not been attempted.

Wang [34] took a predictive learning approach to this problem, i.e., using a large number of in vitro assay endpoints to infer the in vivo point of departure. In this paper, a robust learning approach was developed to infer the in vivo point of departure (POD) with in vitro assay endpoints from ToxCast and Tox21 projects. First, the in vitro dose response data were utilized to derive the in vitro PODs for several hundred chemicals following the BMD approach. These were combined with in vivo PODs from ToxRefDB regarding the rat and mouse liver to build a high-dimensional robust regression model. The advantage of this approach is the separation of chemicals into a majority, well-predicted set; and a minority, outlier set. Prominent relationships will then become apparent in the majority set (Fig. 15.2). For both mouse and rat-liver PODs, over 93% of chemicals have inferred values from in vitro PODs that are within ± 1 of the in vivo PODs on the $\log_{10}$ scale (the target learning region, or TLR) and with R^2 values of 0.80 (rats) and 0.78 (mice) for these chemicals.

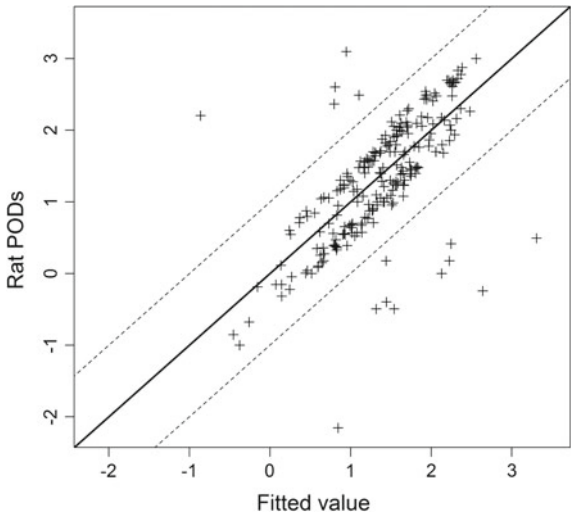

Fig. 15.2 Relationship between rat-liver PODs inferred from in vitro assays and those based on apical endpoints in [34]. The *x*-axis is the fitted value with robust learning. The *y*-axis represents the in vivo PODs from rat-liver chronic studies. Both axes are on the $\log_{10}$ scale. The solid line at the diagonal is the identity line. The two dashed lines indicate the region where the prediction is within ± 1 of the in vivo PODs. This figure is adapted from [34] with permission from Archives of Toxicology, Springer

This is comparable with extrapolation between related species (mouse and rat) with in vivo PODs, which results in 93% chemicals within the TLR and the R^2 being 0.78. Chemicals in the outlier set tend to also display large discrepancies between mouse and rat. This demonstrates that predictive modeling can provide a valuable route to infer in vivo PODs though allowance for a certain portion of outliers has to be made due to the deficiency in the available data.

15.5 Conclusions

Though apical endpoints based on animal studies are still routinely required for toxicological evaluations, there has been increased acceptance and demand to use genomic toxicology to complement traditional approaches regarding POD determination. It has been confirmed in various studies that transcriptional profiles using animal tissues after short-term exposure, when combined with suitable mathematical models, can provide consistent estimates regarding PODs. As discussed in this review, there is a myriad of issues including experiment design, technology platform, statistical filtering of features, BMD modeling, and pathway integration that a researcher has to consider. Though the choice is not always clear cut, there are a number of studies, some discussed in this paper, which provide reasonable guides

on how to perform this type of study. Usually, multiple approaches will arrive at PODs within ±1 of the apical-endpoint-based PODs on the $\log_{10}$ scale, which lends confidence for using transcriptomic data.

A more ambitious and more difficult task is to utilize in vitro assay endpoints to infer in vivo PODs. As shown in [34], the ToxCast data already provide valuable information that can be used to build predictive models for this purpose. However, due to the incompleteness in coverage of important toxicity pathways, some chemicals have to be treated as outliers. Fortunately, more complete data sets may soon be available. The Tox21 program is working to screen a large collection of chemicals with a set of toxicologically relevant "sentinel" genes. The S1500+ sentinel gene list has been created [38], containing 1500 genes designed to comprehensively cover toxicologically relevant pathways by taking advantage of the co-expression patterns between genes. Upon sufficient accumulation of data along these lines, there will be opportunity to develop more powerful models to infer in vivo PODs with high-throughput assays.

Acknowledgements The author would like to thank the editor and an anonymous reviewer for valuable suggestions. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the US Food and Drug Administration.

References

1. NRC (National Research Council) (2007) Toxicity testing in the 21st century: a vision and a strategy. The National Academies Press, Washington, DC
2. NRC (National Research Council) (2017) Using 21st century science to improve risk-related evaluations. The National Academies Press, Washington, DC
3. Locke PA, Westphal M, Tischler J, Hessler K, Frasch P, Myers B, Krewski D (2017) Implementing toxicity testing in the 21st century: challenges and opportunities. Int J Risk Assess Manage 20(1–3):198–225
4. Rudén C, Hansson SO (2010) Registration, evaluation, and authorization of chemicals (REACH) is but the first step. How far will it take us? Six further steps to improve the European chemicals legislation. Environ Health Perspect 118(1):6–10
5. Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T, Yamada H (2015) Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res 43(D1):D921–D927
6. Collins FS, Gray GM, Bucher JR (2008) Transforming environmental health protection. Science 319(5865):906–907
7. Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121(7):756–765
8. Judson R, Houck K, Martin M, Knudsen T, Thomas R, Sipes N, Shah I, Wambaugh J, Crofton K (2014) In vitro and modeling approaches to risk assessment from the U.S. environmental protection agency ToxCast program. Basic Clin Pharmacol Toxicol 115(1):69–76
9. Richard A, Judson R, Houck K, Grulke C, Volarath P, Thillainadarajah I et al (2016) The ToxCast chemical landscape—paving the road to 21st century toxicology. Chem Res Toxicol 29(8):1225–1251
10. Silva M, Pham N, Lewis C, Iyer S, Kwok E, Solomon G, Zeise L (2015) A comparison of ToxCast test results with in vivo and other in vitro endpoints for neuro, endocrine, and developmental toxicities: a case study using endosulfan and methidathion. Birth Defects Res B Dev Reprod Toxicol 104(2):71–89

11. Liu J, Mansouri K, Judson RS, Martin MT, Hong H, Chen M, Xu X, Thomas RS, Shah I (2015) Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem Res Toxicol 28(4):738–751
12. Huang R, Xia M, Sakamuru S, Zhao J, Shahane SA, Attene-Ramos M, Zhao T, Austin CP, Simeonov A (2016) Modelling the Tox21 10K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat Commun 7:10425. https://doi.org/10.1038/ncomms10425
13. Kleinstreuer N, Ceger P, Watt E, Martin M, Houck K, Browne P, Thomas R, Casey W, Dix D, Allen D, Sakamuru S, Xia M, Huang R, Judson R (2017) Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol 30(4):946–964
14. Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD et al (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741
15. Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH et al (2014) Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci 142(2):312–320
16. Thomas RS, Allen BC, Nong A, Yang L, Bermudez E, Clewell HJ III, Andersen ME (2007) A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. Toxicol Sci 98(1):240–248
17. Thomas RS, Wesselkamper SC, Wang NCY, Zhao QJ, Petersen DD et al (2013) Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. Toxicol Sci 134(1):180–194
18. Yang L, Allen BC, Thomas RS (2007) BMDExpress: a software tool for the benchmark dose analyses of genomic data. BMC Genom 8(1):387
19. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4(2):249–264
20. Farmahin R, Williams A, Kuo B, Chepelev NL, Thomas RS, Barton-Maclaren TS, Curran IH, Nong A, Wade MG, Yauk CL (2017) Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. Arch Toxicol 91(5):2045–2065
21. Kodell RL (2009) Replace the NOAEL and LOAEL with the BMDL01 and BMDL10. Environ Ecol Stat 16(1):3–12
22. National Toxicology Program (2018) NTP research report on national toxicology program approach to genomic dose-response modeling. U.S. Department of Health and Human Services, Washington, DC
23. Zhou YH, Cichocki JA, Soldatow VY, Scholl EH, Gallins PJ, Jima D et al (2017) Comparative dose-response analysis of liver and kidney transcriptomic effects of trichloroethylene and tetrachloroethylene in B6C3F1 mouse. Toxicol Sci 160(1):95–110
24. Black MB, Parks BB, Pluta L, Chu TM, Allen BC, Wolfinger RD, Thomas RS (2014) Comparison of microarrays and RNA-seq for gene expression analyses of dose-response experiments. Toxicol Sci 137(2):385–403
25. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140
26. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11(10):R106. https://doi.org/10.1186/gb-2010-11-10-r106, http://genomebiology.com/2010/11/10/R106/
27. Crump K (2002) Critical issues in benchmark calculations from continuous data. Crit Rev Toxicol 32(3):133–153
28. Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15(2):R29. http://genomebiology.com/2014/15/2/R29
29. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P (2015) The molecular signatures database hallmark gene set collection. Cell Syst 1(6):417–425

30. Moffat I, Chepelev NL, Labib S, Bourdon-Lacombe J, Kuo B et al (2015) Comparison of toxicogenomics and traditional approaches to inform mode of action and points of departure in human health risk assessment of benzo [a] pyrene in drinking water. Crit Rev Toxicol 45(1):1–43
31. Labib S, Williams A, Yauk CL, Nikota JK, Wallin H et al (2015) Nano-risk science: application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. Part Fibre Toxicol 13(1):15
32. Dean JL, Zhao QJ, Lambert JC, Hawkins BS, Thomas RS, Wesselkamper SC (2017) Application of gene set enrichment analysis for identification of chemically induced, biologically relevant transcriptomic networks and potential utilization in human health risk assessment. Toxicol Sci 157(1):85–99
33. Filer DL, Kothiya P, Setzer RW, Judson RS, Martin MT (2016) tcpl: the ToxCast pipeline for high-throughput screening data. Bioinformatics 33(4):618–620
34. Wang D (2018) Infer the in vivo point of departure with ToxCast in vitro assay data using a robust learning approach. Arch Toxicol 92(9):2913–2922
35. Shah I, Setzer RW, Jack J, Houck KA, Judson RS, Knudsen TB et al (2016) Using ToxCast™ data to reconstruct dynamic cell state trajectories and estimate toxicological points of departure. Environ Health Perspect 124(7):910–9
36. Sipes NS, Wambaugh JF, Pearce R, Auerbach SS, Wetmore BA et al (2017) An intuitive approach for predicting potential human health risk with the Tox21 10K library. Environ Sci Technol 51(18):10786–10796
37. Pearce RG, Setzer RW, Strope CL, Sipes NS, Wambaugh JF (2017) Httk: R package for high-throughput toxicokinetics. J Stat Softw 79(4):1–26
38. Mav D, Shah RR, Howard BE, Auerbach SS, Bushel PR et al (2018) A hybrid gene selection approach to create the S1500+ targeted gene sets for use in high-throughput transcriptomics. PLoS ONE 13(2):e0191105

Dong Wang received his Ph.D. in Genetics from Iowa State University in 2003 and received his Ph.D. in Statistics in 2006 also from Iowa State University. He was a faculty member in the Department of Statistics, University of Nebraska-Lincoln from 2006 to 2014 with the rank of assistant professor and later associate professor. He worked for three years as the leader of Statistics and Mathematics Group at Dow AgroSciences between 2014 and 2016. He joined National Center for Toxicological Research in December 2016 as a senior statistician in the Division of Bioinformatics and Biostatistics. He has research interest in risk assessment, statistical genomics, statistical machine learning, high-dimensional modeling, and Bayesian methods.

Chapter 16
Molecular Modeling Method Applications: Probing the Mechanism of Endocrine Disruptor Action

Xianhai Yang, Huihui Liu and Rebecca Kusko

Abstract The potential endocrine-related detrimental effects of endocrine-disrupting chemicals (EDCs) on humans and wildlife are a growing worldwide concern. The mechanism of action (MOA) of EDCs induced endocrine-related diseases and endocrine dysfunction can be summarized as the interactions between EDCs and biomacromolecules in endocrine system. Thus, insights into the endocrine-linked MOA of EDCs with corresponding targets will pave the way for developing screening methods of EDCs, prioritizing, and constructing endocrine-related adverse outcome pathways. To date, batteries of laboratory bioassays have been developed and employed to distinguish whether EDCs activate/inhibit/bind to a target or not. However, such test methods poorly assess the underlying molecular mechanisms. Molecular modeling methods are an essential and powerful tool in deciphering the mechanism of endocrine disruptor action. In this chapter, several critical processes related to performing the molecular modeling are described. Topics include preparing 3D biomacromolecules and EDCs structures, obtaining and refining the EDC–biomacromolecule complex, and probing the underlying interaction mechanism. Among these topics, we have emphasized revealing the underlying mechanism by analyzing binding patterns and noncovalent interactions and calculating binding energy. Lastly, future directions in molecular modeling are also proposed.

Keywords Endocrine-disrupting chemicals (EDCs) · Mechanism of endocrine disruptor action · Molecular modeling · Homology modeling · Molecular

X. Yang (✉) · H. Liu
Jiangsu Key Laboratory of Chemical Pollution Control and Resources Reuse, School of Environmental and Biological Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
e-mail: xhyang@njust.edu.cn

H. Liu
e-mail: hhliu@njust.edu.cn

R. Kusko
Immuneering Corporation, Cambridge, MA, USA
e-mail: bkusko@immuneering.com

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_16

docking · Molecular dynamics simulation · Binding pattern · Noncovalent interactions · Binding energy

Abbreviations

4′-HO-BDE 121	4′-HO-2,3′,4,5′,6-pentabrominated diphenyl ether
AOP	Adverse outcome pathway
diBBPA	Dibromobisphenol A
E	Binding energy
E_{total}	Total binding energy
E_{ele}	Electrostatic energy
E_{vdw}	Van der Waals interaction energy
EADB	Estrogenic Activity Database
ECHA	European Chemicals Agency
EDCs	Endocrine-disrupting chemicals
EDSP	Endocrine Disruptor Screening Program
EDSP21	Endocrine Disruptor Screening Program in the twenty-first century
ERα	Estrogen receptors alpha
ERβ	Estrogen receptors beta
hAR	Human androgen receptor
hTTR	Human transthyretin
LBD	Ligand binding domain
logRBA	Median relative binding affinity
MD	Molecular dynamics
MIEs	Molecular-initiating events
MOA	Mechanisms of action
monoBBPA	Monobromobisphenol A
NRs	Nuclear hormone receptors
OECD	Organization for Economic Co-operation and Development
PDB	Protein Data Bank
PFHpA	Perfluoroheptanoic acid
PFOS	Perfluorooctane sulfonic acid
PPARγ	Proliferator-activated receptor gamma
QM/MM	Quantum mechanics/molecular mechanics
(Q)SAR	(Quantitative) structure–activity relationship
RAAF	Read-Across Assessment Framework
RMSD	Root-mean-square deviation
SHBG	Sex hormone-binding globulin
T4	Tetraiodothyronine
TBBPA	3,3′,5,5′-tetrabromobisphenol A
TRβ	Thyroid receptor beta
TCBPA	Tetrachlorobisphenol A

triBBPA	Tribromobisphenol A
UNEP	United Nations Environment Programme
US EPA	United States Environmental Protection Agency
WHO	World Health Organization

16.1 Introduction

Since the 1990s, the scientific community, regulators, and the public have been increasingly concerned by the harmful effects of endocrine-disrupting chemicals (EDCs) on humans and wildlife [1]. The observed adverse effects include suppression of gene expression or enzyme activities, alteration of protein concentrations or hormones homeostasis, disruption of brain or immune system development, reproductive dysfunction and/or hormone-dependent cancers, feminization effects or demasculinization problems, and so on [2–6]. In response to this pressing issue, several EDCs screening programs and national/international actions have been launched and implemented in USA [7], European Union [8], China [9], Japan [10], Organization for Economic Co-operation and Development (OECD) [11], World Health Organization (WHO) [12], and United Nations Environment Programme (UNEP) [13, 14] since 1996. Clarification of the endocrine-linked toxic mechanism of action (MOA) is needed to implement EDC screening programs and actions [15, 16].

To date, there are various mechanisms through which EDCs can exert their effects on the endocrine system [17–20]: (a) impacting macromolecule regulatory function in the hypothalamic-pituitary-gonad/thyroid/adrenal axis, (b) inhibiting hormone synthesis-related enzymes, (c) disrupting hormone transport proteins, (d) activating/inhibiting hormone receptors, and/or (e) inhibiting hormone metabolism-related enzymes. Thus, the endocrine-linked MOA of EDCs causing endocrine-related diseases and endocrine dysfunction can be summarized as the interactions between small molecules (toxicant) and biomacromolecule (target) [21, 22]. Furthermore, the interaction between EDCs and biological targets was also the critical molecular-initiating event (MIEs) of the endocrine-specific adverse outcome pathway (AOP) [23, 24]. Thus, studying the interaction of EDCs with endocrine system targets (hormone receptors, synthesis, and metabolism-related enzymes of hormones, hormone transport proteins, and so on) will pave the way for developing screening methods, prioritizing, and elucidating the endocrine-related AOP.

A variety of bioassays have been developed and used to test if a given EDC activates, inhibits, or binds to a target up to now [11, 18, 25]. However, current experimental methods poorly probe the underlying molecular mechanisms. For example, it is thought that compounds with common structural features exhibit and elicit similar toxicological effects as well as share similar interaction mechanisms. As shown in Fig. 16.1, the 4′-HO-2,3′,4,5′,6-pentabrominated diphenyl ether (4′-HO-BDE 121) and 3,3′,5,5′-tetrabromobisphenol A (TBBPA) are structurally similar to tetraiodothyronine (T4). In contrast, the structure of pentabromophenol, perfluoroheptanoic acid (PFHpA), perfluorooctane sulfonic acid (PFOS) is greatly different

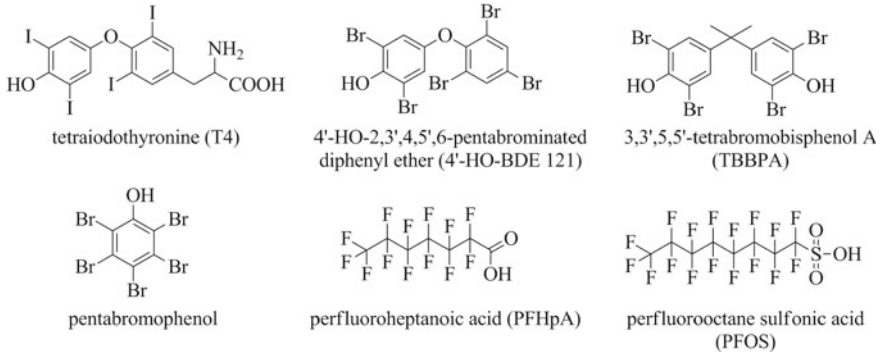

Fig. 16.1 Chemical structure of tetraiodothyronine (T4) and some human transthyretin disruptors

from that of T4. As expected, the experimental results documented that the binding affinity of 4′-HO-BDE 121 and TBBPA to human transthyretin (hTTR) was similar with that of T4 [26]. However, it was puzzling that the hTTR binding potency of pentabromophenol was higher than that of T4 [26]. PFHpA and PFOS also exhibited binding affinity to hTTR [27, 28]. Not only T4 mimics were potential hTTR binders, but other structurally dissimilar binders exist too. Which compounds should be considered as potential hTTR binders? Answering this question is paramount for screening potential EDCs or prioritizing. Before answering this question, we need to first discuss other questions, such as why does pentabromophenol have comparable hTTR binding potency to T4? What is the underling binding mechanism between EDCs and hTTR? It was difficult to clarify the underlying molecular mechanism by employing aforementioned laboratory test methods only.

Computational toxicology methods have been an essential and powerful tool for querying environmental endocrine-disrupting effects [11, 15, 25, 29, 30]. For example, in order to implement the Endocrine Disruptor Screening Program (EDSP) in the twenty-first century (EDSP21), the United States Environmental Protection Agency (US EPA) has been moving toward computational models and high-throughput screening assays to help prioritize and screen chemicals for endocrine activity [31]. When leveraged appropriately, computational toxicology methods can: (1) reveal the interaction mechanism between EDCs and biomacromolecules, (2) fill the data gap for EDCs on their endocrine-disrupting activity, (3) set priority and (4) screen. In practice, the predictive methods used in this field could be crudely divided into two basic types: toxicant-based (also called ligand-based) and target-based (also called structure-based) [32].

The toxicant-based methods customarily derive a quantitative or qualitative relationship among various attributes (e.g., molecular descriptors and/or physicochemical properties) of EDCs and a given biological targets activities (end points). In this method, only toxicant structures are involved in modeling. To date, there is extensive literature on endocrine activity modeling, e.g., the (quantitative) structure–activity relationship ((Q)SAR) models for nuclear receptors (NRs) [33–39],

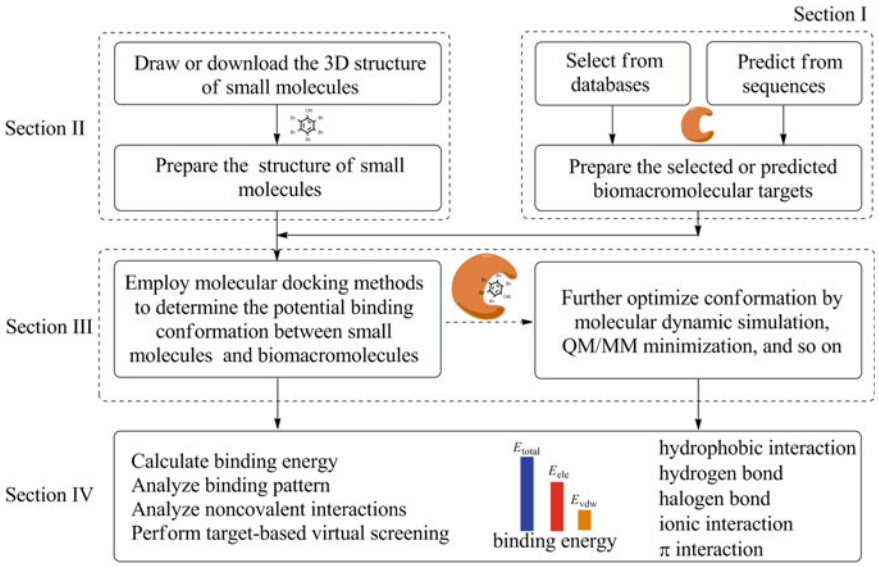

Fig. 16.2 Overall flow diagram of molecular modeling method

hormone transport proteins [40–42], steroidogenesis [43], and so on. Some predictive models have even been integrated into existing software such as VEGA (https://www.vegahub.eu/) and OECD QSAR Toolbox (https://www.qsartoolbox.org/) [44]. In addition, some EDC-related databases including Estrogenic Activity Database (EADB) [45], EDCs DataBank [46], EDSP21 Dashboard (https://actor.epa.gov/edsp21/), the potential endocrine disruptors list in Endocrine Disruption Exchange (https://endocrinedisruption.org/), were also constructed. Different from the toxicant-based methods, both the 3D structure of the interested biomacromolecule and toxicant are involved in the modeling process of target-based methods [47, 48]. Figure 16.2 depicts an overall diagram of target-based modeling method.

In this chapter, we zoom in on methods to probe the mechanism of endocrine disruptor action by employing the target-based molecular modeling methods. Generally speaking, there are four components. Parts one and two involve obtaining and preparing the 3D structure of target biomacromolecules and model compounds, respectively. In part three, the biomacromolecule model compound complex is obtained. During part four, the underlying binding mechanism of action is probed and target-based virtual screening is performed. Among them, we have been put emphasis on revealing the underlying mechanism by analyzing binding patterns, noncovalent interactions, and binding energy.

16.2 Preparation of 3D Biomacromolecule Structures

There are two major sources for researchers to obtain 3D biomacromolecule structures in the endocrine system, i.e., selecting the crystal structure from various databases or predicting the 3D structure from corresponding amino acid sequence. Generally, various databases, e.g., Protein Data Bank (PDB, https://www.rcsb.org/) [49] are the preferred sources for researchers to obtain their desired target structure. However, there are many challenges and disadvantages in selecting the crystal structure of a biomacromolecule from public databases. If an inappropriate 3D target structure was selected during the molecular modeling, the accuracy and trustworthiness of the modeling results will inevitably be affected. Recently, Yang et al. [50] proposed six principles to guide appropriate crystal structure selection. The principles included (1) species differences, (2) the MOA, (3) mutant amino acid residues, (4) protein chain number, (5) the degree of structural similarity between the ligand in crystal structure and the model compounds, and (6) others factors, e.g., the experimental pH conditions of the crystal structure determined and resolution [50]. After considering the aforementioned factors, appropriate 3D structure can be selected more successfully.

As it is difficult and time-consuming to determinate the crystal structures from experimental methods such as X-ray crystallography and nuclear magnetic resonance for every target of interest, the available crystal structures of biomacromolecules are unfortunately still limited. On the contrary, determination of the macromolecule sequences is easier than structure determination. As of July 2018, the UniParc database contained more than 221 million protein sequences (http://www.uniprot.org/uniparc/). In contrast, as of July 2018 there were only 142,379 structures solved experimentally in PDB. A huge gap between known annotated sequences and available 3D structures existed [51]. To bridge the gap between the demand for 3D structures of biomacromolecules and limited experimental structures, computational methods, e.g., homology modeling, protein threading, and ab initio methods can be employed to predict the 3D structure of targets from corresponding amino acid sequence [52–55]. Many tools could be used to perform this task, such as modeler [56], Swiss-PdbViewer [57], and so on.

Before proceeding, the obtained 3D target structure should undergo treatment as follows. Generally, the treatment steps include but are not limited to:

(1) Insert missing atoms in incomplete residues;
(2) Remove the ions, water molecules, and other substructures;
(3) Standardize the atom and/or residue names;
(4) Protonate or deprotonate the ionizable residues under given pH conditions;
(5) Find and define the binding sites. It deserves mentioning that some important water molecules in the structure of targets should be kept. For example, it was well documented that there were water molecules taking part in forming the conserved hydrogen bond with Glu 353 and Arg 394 in the ligand-binding domain (LBD) of estrogen receptors alpha (ERα) [58]. In this case, those water molecules were recommended to be retained.

16.3 Preparation of the Molecular Structure of EDC Molecules

The molecular structure of EDCs can be downloaded from various software or databases such as PubChem (https://pubchem.ncbi.nlm.nih.gov/) and ZINC (http://blaster.docking.org/zinc). For the compounds not contained in those sources, their molecular structures must be drawn directly. Then, the possible ionization states can be generated at normal physiological or experimental pH conditions. Lastly, their molecular structures can be optimized by various methods.

16.4 Obtaining the EDC—Biomacromolecule Complexes

16.4.1 Performing Molecular Docking

As shown in Fig. 16.2, a key prerequisite for successful application of molecular modeling to probe the binding mechanism lies in obtaining the toxicant–target complex. The complex is a conformation demonstrating where (binding site location) and how (binding orientation) the small molecule bound within the LBD of a given macromolecule. Undoubtedly, the experimental methods like NMR spectroscopy [59, 60] are the most accurate methods to investigate toxicant–target interactions. The time and cost limitations of NMR precipitate the need for alternative technologies.

Molecular docking is a powerful alternative technology for predicting the toxicant–target complex [61–64]. Generally speaking, molecular docking places a given ligand into the binding site of a given target. The molecular docking protocol usually contains two components: a search algorithm and a scoring function [65]. The search algorithms determine the precise ligand-binding site location and binding orientation, which also referred to as a "pose." The scoring function addresses the question, "how well does the ligand bind to the protein?" by ranking the pose. In the past thirty years, many search algorithms and scoring functions were developed. Detailed discussion about the search algorithms and scoring functions are reviewed elsewhere [66, 67] and are not considered herein.

To date, more than 60 software and Web servers for docking are available [68], e.g., Autodock [69] and GOLD [70]. Previous studies document that the predictive performance of the docking tools varies greatly between targets [62, 68]. It is therefore difficult to readily conclude which tool is more appropriate than others for a given modeling system. Thus, each given docking tools should be validated before use. In practice, the target/ligand crystal structures are typically utilized for validation. After extracting the 3D crystal ligand structure, it will be docked back into the binding site of that target. Lastly, the root-mean-square deviation (*RMSD*) between the heavy-atom positions of the crystal ligand and that of the predicted ligand is calculated [71]. If the *RMSD* value is <2 Å, the selected docking tool is considered reliable [61, 63].

16.4.2 Refining the Complex

The biomacromolecule is conformationally altered during the binding process. For example, the position of helix 12 in the human ER underwent a large rearrangement when the ligand changed from an agonist to an antagonist [72]. Not accounting for such structural changes during docking may prevent identification of the true binding pose. Docking which fully considers target flexibility is very computationally expensive. Instead, semi-flexible docking where the flexibility ligands are taken in account while the protein is kept rigid is more commonly used to prepare the toxicant–target complex.

To overcome this problem, molecular dynamics (MD) simulation or hybrid quantum mechanics/molecular mechanics (QM/MM) simulation can further refine the complex generated from docking [73–75]. MD simulation leverages classical molecular mechanics force fields to predict particle motions as a function of time [76]. In contrast, QM/MM simulation queries the site of interest (QM region) with quantum mechanics. The rest of the system (MM region) is studied with classical molecular mechanics force fields [77]. Those two provide detailed information on the conformational changes and fluctuations of the molecules in the complex, and both are now routinely employed to refine molecular structures, investigate the dynamics of a given molecular system, and elucidate atomic-level interactions [78]. Various software including AMBER [79], GROMACS [80], NAMD [81], CHARMM [82] could be employed to refine the complex.

As described so far, the desired toxicant–target complex is prepared on the basis of molecular docking and/or MD simulation or QM/MM simulation. Next, the types of intermolecular interactions, bond distances, and binding affinities can be inferred and used to reveal the underlying interaction mechanism between EDCs and endocrine system targets.

16.5 Probing the Underlying Binding Mechanism of Action

16.5.1 Analyzing Binding Patterns

As stated above, the binding conformation informs the binding site location and binding orientation of ligands within the LBD of a given target. Thus, by analyzing the binding pattern, we can determine the dominant orientation and preferential binding region for the functional groups in EDCs. Then, we would be able to further analyze the binding conformation similarities and differences between endocrine hormone and EDCs, and/or among the different EDCs. For example, by analyzing simulated ligand conformations and hTTR crystal structures, we found that the ionized function group (e.g. O^-, COO^-) in the ligand had a dominant orientation and pointed toward the entry port of the binding site. However, no dominant orientation was observed for the neutral form of the ionizable group [83, 84].

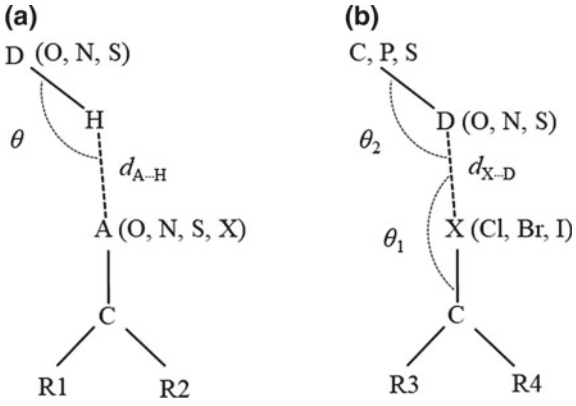

Fig. 16.3 Schematic diagram of hydrogen bond and halogen bond

16.5.2 Analyzing Noncovalent Interactions

Noncovalent interaction formation is the intrinsic driving force for molecular recognition between EDCs and endocrine system targets [85]. Detailed understanding of those interactions is of vital importance to reveal the binding mechanism. The reported noncovalent interactions included hydrophobic interaction, hydrogen bond, halogen bond, electrostatic interaction or ionic interaction, and π interaction [86]. It is important to note that not all noncovalent interaction types will be formed in every system. In this case, a comprehensive noncovalent interaction analysis is required to identify the dominant interactions for a given system.

(1) *Hydrophobic interaction*

Hydrophobic interactions arise from the close contact between lipophilic groups in a ligand and nonpolar amino acid side chains in a target [87]. Hydrophobic interactions are one of the most basic noncovalent interactions between a small molecule and its target. In some cases, the hydrophobic interaction is the only dominant noncovalent interaction [88, 89]. For example, Avvakumov et al. [90] reported that the binding of non-steroidal ligands with sex hormone-binding globulin (SHBG) was governed by hydrophobic interactions.

(2) *Hydrogen bond*

The hydrogen bond is another of the most common noncovalent interactions between a small molecule and its target [88, 91]. A hydrogen bond is formed between an electron acceptor (A = O, N, X) in one molecule and a hydrogen atom on an electron donor (D = O, N, S) (Fig. 16.3a). The criteria for a hydrogen bond are:

(a) The distance between a hydrogen atom and an electron acceptor atom ($d_{A\cdots H}$) is < their sum of van der Waals radii;
(b) The D-H···A angle is >135° [92].

For many targets, there may be conserved hydrogen bonds. For example, three residues involved in forming hydrogen bonds are particularly important for ligand binding in ERα LBD, namely, Glu 353, Arg 394, and His 524. While in ERβ LBD, the three residues were Glu 305, Arg 346, and His 475. Two residues (Gln 725 and Arg 766) formed conserved hydrogen bonds between the ligand and the human progesterone receptor [93]. Forming the conserved hydrogen bonds network is the critical factor triggering the subsequent active or inactive conformation transition [94].

(3) *Halogen bond*

The halogen bond is a type of noncovalent interaction between a halogen atom (Cl, Br, I) in one molecule and an electron donor (D = O, N, S) in another molecule (Fig. 16.3b) [95]. The criteria for a halogen bond are:

(a) The distance between halogen atom and electron donor atom ($d_{X\cdots D}$) is < their sum of van der Waals radii;
(b) The C-X···D angle is >140° [96].

It was recognized that halogen bond plays important roles in the molecular recognition processes between organohalogens and target [91, 97, 98]. It has been reported that many EDCs contained halogen moieties in their structure. For example, among the 250 tested hTTR disruptors, 198 compounds (79%) were halogenated [99]. This fact indicates that the hTTR disruptor halogen moieties may drive the interactions between those compounds and hTTR. Indeed, our results implied that the halogen moieties in hTTR binders could directly or indirectly affect the binding interactions [100]. On one hand, the halogen atom could form halogen bonds and halogen–hydrogen bonds with the residues in hTTR directly. On the other hand, the halogen atom could affect binding through inductive effects and hydrophobic effects. Additionally, the results from Zhuang et al. [101] indicated that the bromine and chlorine atoms in TBBPA, tribromobisphenol A (triBBPA), and tetrachlorobisphenol A (TCBPA) could form halogen bonds with the residues in PPARγ (proliferator-activated receptor gamma). The bromine atoms in Monobromobisphenol A (monoBBPA) and dibromobisphenol A (diBBPA) could form halogen bonds with the residues in ERα.

(4) *Ionic interactions*

An ionic interaction is formed between a charged group in the ligand and an oppositely charged group in the target. In some cases, a hydrogen bond would be superimposed onto an ionic interaction, which is called a charge-assisted hydrogen bond [87]. If there exist stable ionic interactions, the distance between the two oppositely charged groups is ≤5 Å [102]. For example, we analyzed the possibility of forming an orientational ionic interaction between the model compounds (phenolic compounds and poly-/perfluorinated chemicals) and hTTR. We calculated the distance (d) between the anionic groups in the ligands and the $-NH_3^+$ group in Lys 15 based on the conformations from the simulation and the hTTR crystal structures. Among the 82 simulated anionic ligands, there were 67 compounds with $d \leq 5$ Å. For the 64 crystal complexes with anionic ligands, the d for 49 structures is ≤5 Å. The result

confirmed that the anionic groups of the ligands formed ionic interactions with the $-NH_3^+$ group of Lys15 in hTTR [83, 84]. On the basis of the obtained conformations (200 frames) from MD simulation for each compound, we also found that the *d* for >91% conformations was ≤5 Å, indicating that the formed ionic interaction was stable [84].

(5) *π interactions*

The π interaction included π–π, cation–π, anion–π, and sigma–π interactions. For the endocrine system targets, π–π and cation–π interaction have been observed up to now. For example, Li et al. [103] observed π–π interactions between the phenyl group of hydroxylated polybrominated diphenyl ethers (HO-PBDEs) and Phe272, Phe442, and Phe455 of TRβ. Yang et al. [83] observed cation–π interaction between the phenyl group of phenolic compounds and $–NH_3^+$ group of Lys15 in hTTR.

As a best practice, it is recommended to analyze the binding pattern and noncovalent interaction of a crystal ligand with the corresponding target before performing molecular modeling in order to validate the reliability of simulation results. For example, the noncovalent interaction analysis indicated that the bisphenols only formed hydrogen bonds and hydrophobic interactions with human androgen receptor (hAR). There were four amino acid residues (Asn 705, Gln 711, Arg 752 and Thr 877) involved in forming hydrogen bonds (Fig. 16.4a). Among them, the Gln 711 and Asn 705 were identified as the most important amino acid residues by calculating the hydrogen bond formation rates. After analyzing the 76 hAR crystal structures, a similar result was observed (Fig. 16.4b), which confirmed the reliability of the simulation results [104].

16.5.3 Calculating Binding Energy

Based on the simulated conformation, the binding energy (*E*) of EDCs with a target or other scores could be calculated. Theoretically, there was a significant correlation between *E* or other scores and the biological targets activities, e.g., estrogenic activity and thyroid hormone activity. A linear correlation of *E* or other scores with biological target activities could be derived, which could be further used to screen potential EDCs or fill the data gap. For example, Ng et al. [78] developed a good linear relationship between the median relative binding affinity values (log*RBA*) of estrogenic activity for bisphenol A replacement compounds and *E* ($\log RBA = -7.719 - 0.0860\ E$ ($p = 0.007$)). Then, the missing estrogenic activity of other bisphenol A replacement compounds was filled by employing the developed equation.

In addition, lots of software and tools could decompose the total binding energy (E_{total}) into different components, such as electrostatic energy (E_{ele}), van der Waals interaction energy (E_{vdw}), and so on. After analyzing the energy components, we could identify the dominant driving force. For example, Lu et al. [106] decomposed the E_{total} into three components using Amber software, and the results indicated that the E_{vdw} was the major component of the total binding energy. This implied that

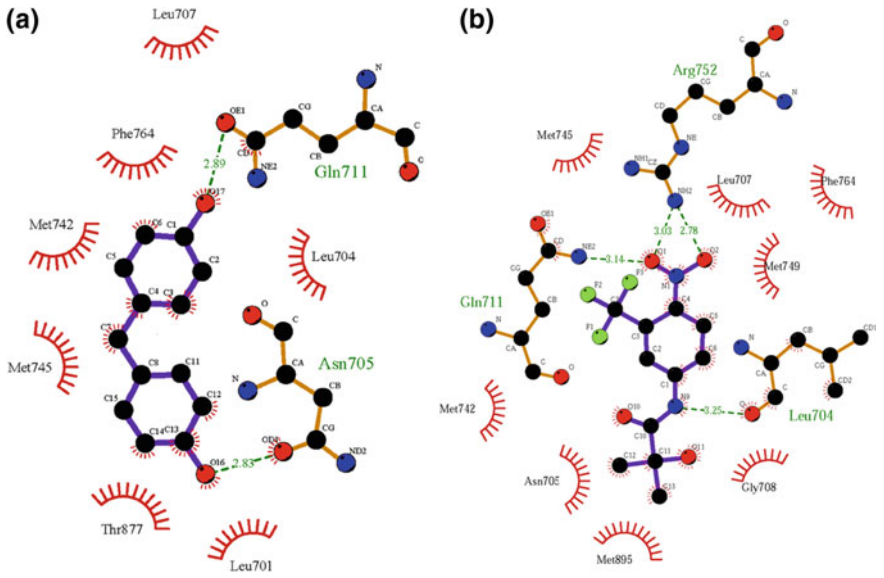

Fig. 16.4 Binding interactions of simulated Bisphenol F (**a**) and crystal hydroxyflutamide (**b**, PDB ID: 4OIU) with human androgen receptor in the ligand-binding site. (ligand bond, receptor bond, hydrogen bonding, His 53 receptor residues involved in hydrophobic interactions, Corresponding atoms involved in hydrophobic interactions). This was illustrated by LigPlot+ program [105]

critical driving force of the binding between the bisphenol S analogues and TRβ was van der Waals interactions.

So far, we have introduced methods of analyzing binding patterns, noncovalent interactions, and binding energy. How does one make use of that information to unveil the underlying binding mechanism? We will employ the interaction of EDCs with hTTR as an example for the reader. The experimental results indicated that ionizable function groups, aromatic ring, and halogen in EDCs were critical structural alerts that can affect the binding potency between EDCs and hTTR [26, 27, 107, 108]. What is the role of those structural alerts in this molecular recognition process? Our molecular modeling results indicated that the aromatic ring could form cation–π interaction with the $-NH_3^+$ group of Lys15 in hTTR. The halogen could form halogen bonds and halogen–hydrogen bonds with the residues in TTR directly. In addition, the halogen also could affect the binding through inductive effects and hydrophobic effects. For the ionizable group in EDCs, their anionic form binds more strongly to hTTR than a corresponding neutral form. Thus, the ionization of the ionizable groups was non-negligible. The anionic form of the ionizable groups could form ionic interaction and hydrogen bond interaction with hTTR. Forming those dominant and orientational noncovalent interactions lead to the anionic form of ionizable functional groups in EDCs orienting toward the entry port of hTTR (Fig. 16.5) [83, 84, 100].

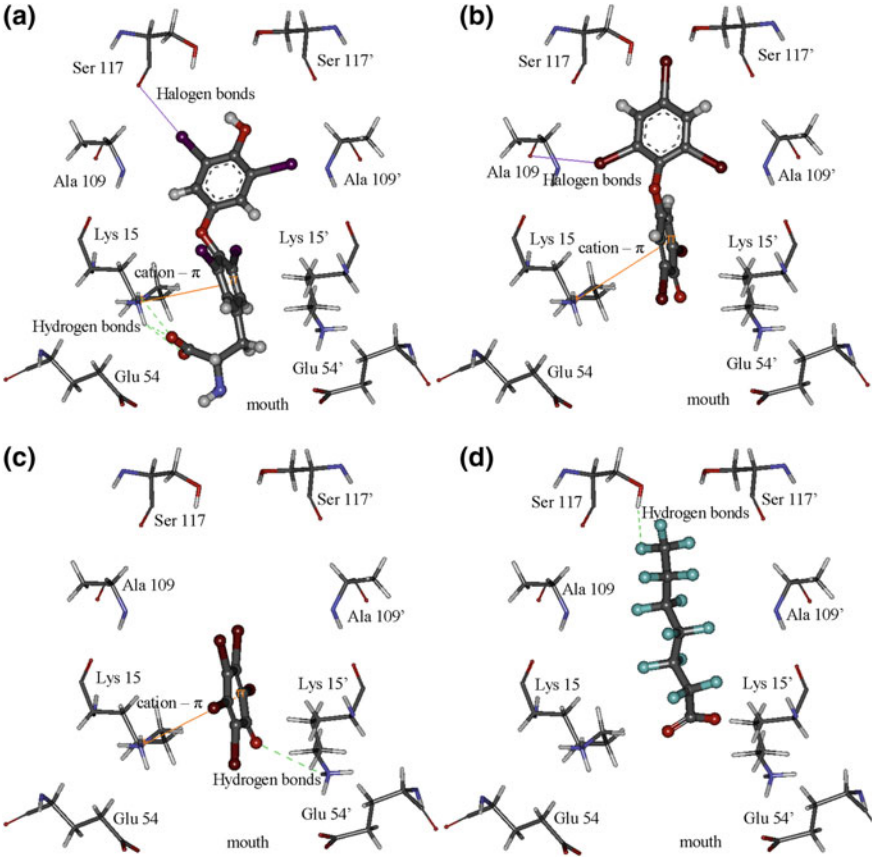

Fig. 16.5 Conformations of T4 (**a**) 4′-HO-2,3′,4,5′,6-pentabrominated diphenyl ether (**b**) pentabromophenol (**c**), and perfluoroheptanoic acid (**d**) adopted in the binding site of human transthyretin (1ICT). The green line, purple line, and orange line represent hydrogen bonds, halogen bonds, and π–cation interactions, respectively. The figures were prepared by using Discovery Studio Visualizer 2017 (BIOVIA, http://accelrys.com/)

As a summary, T4 formed hydrophobic interaction, hydrogen bond, halogen bond, ionic interaction, and cation–π interaction with hTTR. The quantitative contribution of those five noncovalent interactions in the molecular recognition between EDCs and hTTR was not completely clear. However, the hydrogen bond, halogen bond, ionic interaction, and cation–π interaction still could be identified as the critical noncovalent interactions based on this comparative analysis. If one compound could form the critical noncovalent interactions with hTTR, it may be considered as a potential hTTR binder. For example, pentabromophenol was one of the most potent hTTR binders and could form a hydrophobic interaction, hydrogen bond, ionic interaction, and cation–π interaction with hTTR. Compared to T4, the ionizable halogenated aliphatic compounds do not have aromatic rings in their molecular structures. Thus,

four types of noncovalent interactions except for cation–π interactions were observed between the aliphatic compounds and hTTR. The missing cation–π interaction can explain why those compounds exhibit moderate hTTR binding potency.

To date, many target-based virtual screening strategies and methods have been developed [78, 108–112]. In addition, some target-based virtual screening software has also been developed, including VirtualToxLab™ [113]. We will not attempt here to cover these advances in detail.

16.6 Conclusions and Future Directions

In this contribution, we have described how to reveal the underlying mechanism of endocrine disruptor action by employing the molecular modeling methods. Various aspects were discussed including (1) preparation of the 3D structure of biomacromolecules and EDC molecules, (2) obtaining EDC–biomacromolecule complexes, and (3) probing the mechanism. A variety of examples were included in the presentation.

Some future directions are proposed as below:

Improving the modeling methods. Tremendous progress had been made for the target-based methods. However, the efficiency and accuracy of the modeling methods (e.g., the search algorithms and scoring functions) need to be improved in the future.

Nonreceptor-mediated toxicity pathways. It was clear that EDCs could interfere with multiple steps of hormone regulation, including biosynthesis and metabolism, plasma binding, receptor activation/inhibition. Thus far, most researches in EDCs have focused on modeling the disruption effects of nuclear receptors. There is a need to pursue more efforts in modeling the interaction between EDCs and nonreceptor-mediated target in the future.

Attention to species diversity. The structure and function of biomacromolecules in the endocrine system were conserved across different species. However, the sequence similarity showed considerable species variation, which may result in different binding affinity and interaction mechanisms of EDCs with target from different species.

Molecular modeling protocols or guidelines. Lots of software or tools for target-based methods were available. In this situation, how to correctly use those tools and obtain consistent, reproducible results became a priority problem. Chen conducted a critical survey on whether we should trust the results of docking studies [68]. His results indicated that questionable docking results could be observed, even in high-profile journals. How do we increase the accuracy and scientific rigorousness of the modeling results? Issuing a general molecular modeling protocol or guideline may be one solution. To date, several official guidelines for toxicant-based modeling methods have been issued to guide the development and use of ligand-based predictive model. For example, the OECD has published a guidance document on the validation of (quantitative) structure–activity relationship ((Q)SAR) models [114] and guidance on grouping of chemicals [115]. Meanwhile, the European Chemicals Agency (ECHA) has issued Read-Across Assessment Framework (RAAF) [116]. However, no official

molecular modeling protocols or guidelines were available for target-based methods to date. The lack of such protocols or guidelines may lead to inconsistent use of such methods [117]. Thus, in order to ensure the molecular modeling processes are performed and evaluated in a consistent, reproducible, and well-documented manner, a molecular modeling protocol or guideline should be issued as quickly as possible.

Acknowledgements The study was supported by National Natural Science Foundation of China (No. 21507038, No. 41671489, No. 21507061) and Natural Science Foundation of Jiangsu Province (No. BK20150771).

References

1. Kwiatkowski CF, Bolden AL, Liroff RA, Rochester JR, Vandenbergh JG (2016) Twenty-five years of endocrine disruption science: remembering Theo Colborn. Environ Health Perspect 124(9):A151–A154
2. Diamanti-Kandarakis E, Bourguignon JP, Giudice LC, Hauser R, Prins GS, Soto AM, Zoeller RT, Gore AC (2009) Endocrine-disrupting chemicals: an endocrine society scientific statement. Endocr Rev 30(4):293–342
3. Soto AM, Sonnenschein C (2010) Environmental causes of cancer: endocrine disruptors as carcinogens. Nat Rev Endocrinol 6(7):363–370
4. United Nations Environment Programme/World Health Organization (2013) State of the science of endocrine disrupting chemicals. United Nations Environment Programme/World Health Organization (UNEP/WHO), Geneva
5. Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, Toppari J, Zoeller RT (2015) EDC-2: the endocrine society's second scientific statement on endocrine-disrupting chemicals. Endocr Rev 36(6):E1–E150
6. Matthiessen P, Wheeler JR, Weltje L (2018) A review of the evidence for endocrine disrupting effects of current-use chemicals on wildlife populations. Crit Rev Toxicol 48(3):195–216
7. U.S. Environmental Protection Agency (1998) Addendum endocrine disruptor screening and testing advisory committee. U.S. Environmental Protection Agency, Washington
8. Commission of the European Communities (1999) Communication from the commission to the council and the european parliament, community strategy for endocrine disrupters, a range of substances suspected of interfering with the hormone systems of humans and wildlife. Commission of the European Communities, Brussels
9. The State Council of Pepole's Republic of China (2015) Notice of the state council on issuing the action plan for prevention and control of water pollution. GF[2015] No. 17. The State Council of Pepole's Republic of China, Beijing
10. Ministry of the Environment of Japan (1998) Environment agency's basic policy on environmental endocrine disruptors, strategic programs on environmental endocrine disruptors SPEED' 98. Environmental Policy Bureau, Tokyo
11. Organization for Economic Co-Operation and Development (2012) Guidance document on standardised test guidelines for evaluating chemicals for endocrine disruption. Technical report for OECD environment, health and safety publications series on testing and assessment no. 150. OECD, Paris
12. Damstra T, Barlow S, Bergman A, Kavlock R, Van Der Kraak G (2002) Global assessment of the state-of-the-science of endocrine disruptors. World Healthe Organization, Geneva
13. International Conference on Chemicals Management Third Session (2012) Report of the international conference on chemicals management on the work of its third session. International Conference on Chemicals Management Third Session, Nairobi

14. International Conference on Chemicals Management Fourth Session (2015) Emerging policy issues and other issues of concern (report by the secretariat). International Conference on Chemicals Management Fourth Session, Geneva
15. Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, Lucier G, Luster M, Mac MJ, Maczka C, Miller R, Moore J, Rolland R, Scott G, Sheehan DM, Sinks T, Tilson HA (1996) Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. Environ Health Perspect 104(S4):715–740
16. Schug TT, Johnson AF, Birnbaum LS, Colborn T, Guillette LJ Jr, Crews DP, Collins T, Soto AM, Vom Saal FS, McLachlan JA, Sonnenschein C, Heindel JJ (2016) Minireview: endocrine disruptors: past lessons and future directions. Mol Endocrinol 30(8):833–847
17. Dayan CM, Panicker V (2009) Novel insights into thyroid hormones from the study of common genetic variation. Nat Rev Endocrinol 5(4):211–218
18. Murk ATJ, Rijntjes E, Blaauboer BJ, Clewell R, Crofton KM, Dingemans MM, Furlow JD, Kavlock R, Köhrle J, Opitz R, Traas T, Visser TJ, Xia M, Gutleb AC (2013) Mechanism-based testing strategy using in vitro approaches for identification of thyroid hormone disrupting chemicals. Toxicol Vitro 27(4):1320–1346
19. Gilbert ME, Rovet J, Chen Z, Koibuchi N (2012) Developmental thyroid hormone disruption: prevalence, environmental contaminants and neurodevelopmental consequences. Neurotoxicology 33(4):842–852
20. Li F, Yang XH, Chen JW, Lyakurwa F (2015) QSARs on the thyroid hormone effects of polybrominated diphenyl ether (PBDE) derivatives (Chapter 17), In: Zeng EY (ed) Persistent organic pollutants (POPs): analytical techniques, environmental processes and biological effects. Compr Anal Chem 67:547–586
21. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl 41(15):2644–2676
22. Rabinowitz JR, Little SB, Laws SC, Goldsmith MR (2009) Molecular modeling for screening environmental chemicals for estrogenicity: use of the toxicant-target approach. Chem Res Toxicol 22(9):1594–1602
23. Allen TE, Goodman JM, Gutsell S, Russell PJ (2014) Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem Res Toxicol 27(12):2100–2112
24. Ankley GT, Edwards SW (2018) The adverse outcome pathway: a multifaceted framework supporting 21st century toxicology. Curr Opin Toxicol 9:1–7
25. Willett CE, Bishop PL, Sullivan KM (2011) Application of an integrated testing strategy to the U.S. EPA endocrine disruptor screening program. Toxicol Sci 123(1):15–25
26. Meerts IA, van Zanden JJ, Luijks EA, van Leeuwen-Bol I, Marsh G, Jakobsson E, Bergman A, Brouwer A (2000) Potent competitive interactions of some brominated flame retardants and related compounds with human transthyretin in vitro. Toxicol Sci 56(1):95–104
27. Weiss JM, Andersson PL, Lamoree MH, Leonards PE, van Leeuwen SP, Hamers T (2009) Competitive binding of poly- and perfluorinated compounds to the thyroid hormone transport protein transthyretin. Toxicol Sci 109(2):206–216
28. Ren XM, Qin WP, Cao LY, Zhang J, Yang Y, Wan B, Guo LH (2016) Binding interactions of perfluoroalkyl substances with thyroid hormone transport proteins and potential toxicological implications. Toxicology 366–367:32–42
29. Kavlock R, Dix D (2010) Computational toxicology as implemented by the U.S. EPA: providing high throughput decision support tools for screening and assessing chemical exposure, hazard and risk. J Toxicol Environ Health B Crit Rev 13:197–217
30. Browne P, Casey WM, Dix DJ (2018) Use of high-throughput and computational approaches for endocrine pathway screening. In: Garcia-Reyero N, Murphy C (eds) A systems biology approach to advancing adverse outcome pathways for risk assessment. Springer, Cham
31. U.S. Environmental Protection Agency (2015) Use of high throughput assays and computational tools; endocrine disruptor screening program; notice of availability and opportunity for comment federal register, pp 35350–35355

32. Śledź P, Caflisch A2 (2018) Protein structure-based drug design: from docking to molecular dynamics. Curr Opin Struct Biol 48:93–102
33. Hong H, Tong W, Fang H, Shi L, Xie Q, Wu J, Perkins R, Walker JD, Branham W, Sheehan DM (2002) Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts. Environ Health Perspect 110(1):29–36
34. Li JZ, Gramatica P (2010) Classification and virtual screening of androgen receptor antagonists. J Chem Inf Model 50(5):861–874
35. Li F, Chen JW, Wang ZJ, Li J, Qiao XL (2009) Determination and prediction of xenoestrogens by recombinant yeast-based assay and QSAR. Chemosphere 74(9):1152–1157
36. Yu H, Wondrousch D, Li F, Chen J, Lin H, Ji L (2015) In silico investigation of the thyroid hormone activity of hydroxylated polybrominated diphenyl ethers. Chem Res Toxicol 28(8):1538–1545
37. Mansouri K, Abdelaziz A, Rybacka A, Roncaglioni A, Tropsha A, Varnek A, Zakharov A, Worth A, Richard AM, Grulke CM, Trisciuzzi D, Fourches D, Horvath D, Benfenati E, Muratov E, Wedebye EB, Grisoni F, Mangiatordi GF, Incisivo GM, Hong H, Ng HW, Tetko IV, Balabin I, Kancherla J, Shen J, Burton J, Nicklaus M, Cassotti M, Nikolov NG, Nicolotti O, Andersson PL, Zang Q, Politi R, Beger RD, Todeschini R, Huang R, Farag S, Rosenberg SA, Slavov S, Hu X, Judson RS (2016) CERAPP: collaborative estrogen receptor activity prediction project. Environ Health Perspect 124(7):1023–1033
38. Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, Thomas RS, Casey WM, Dix DJ, Allen D, Sakamuru S, Xia M, Huang R, Judson R (2017) Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol 30(4):946–964
39. Yin C, Yang X, Wei M, Liu H (2017) Predictive models for identifying the binding activity of structurally diverse chemicals to human pregnane X receptor. Environ Sci Pollut Res 24(24):20063–20071
40. Papa E, Kovarich S, Gramatica P (2013) QSAR prediction of the competitive interaction of emerging halogenated pollutants with human transthyretin. SAR QSAR Environ Res 24(4):333–349
41. Liu H, Yang X, Lu R (2016) Development of classification model and QSAR model for predicting binding affinity of endocrine disrupting chemicals to human sex hormone-binding globulin. Chemosphere 156:1–7
42. Liu H, Yang X, Yin C, Wei M, He X (2017) Development of predictive models for predicting binding affinity of endocrine disrupting chemicals to fish sex hormone-binding globulin. Ecotoxicol Environ Saf 136:46–54
43. Breen MS, Breen M, Terasaki N, Yamazaki M, Conolly RB (2010) Computational model of steroidogenesis in human H295R cells to predict biochemical response to endocrine-active chemicals: model development for metyrapone. Environ Health Perspect 118(2):265–272
44. Coady KK, Biever RC, Denslow ND, Gross M, Guiney PD, Holbech H, Karouna-Renier NK, Katsiadaki I, Krueger H, Levine SL, Maack G, Williams M, Wolf JC, Ankley GT (2017) Current limitations and recommendations to improve testing for the environmental assessment of endocrine active substances. Integr Environ Assess Manag 13(2):302–316
45. Shen J, Xu L, Fang H, Richard AM, Bray JD, Judson RS, Zhou G, Colatsky TJ, Aungst JL, Teng C, Harris SC, Ge W, Dai SY, Su Z, Jacobs AC, Harrouk W, Perkins R, Tong W, Hong H (2013) EADB: an estrogenic activity database for assessing potential endocrine activity. Toxicol Sci 135(2):277–291
46. Montes-Grajales D, Olivero-Verbel J (2015) EDCs DataBank: 3D-structure database of endocrine disrupting chemicals. Toxicology 327:87–94
47. Rabinowitz JR, Goldsmith MR, Little SB, Pasquinelli MA (2008) Computational molecular modeling for evaluating the toxicity of environmental chemicals: prioritizing bioassay requirements. Environ Health Perspect 116(5):573–577
48. Chen QC, Tan H, Yu H, Shi W (2018) Activation of steroid hormone receptors: shed light on the in silico evaluation of endocrine disrupting chemicals. Sci Total Environ 631–632:27–39
49. Burley SK, Berman HM, Christie C, Duarte JM, Feng Z, Westbrook J, Young J, Zardecki C (2018) RCSB protein data bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. Protein Sci 27(1):316–330

50. Yang XH, Liu HH, Liu JN, Li F, Li XH, Shi LL, Chen JW (2016) Rational selection of the 3D structure of biomacromolecules for molecular docking studies on the mechanism of endocrine disruptor action. Chem Res Toxicol 29(9):1565–1570
51. Xiang ZX (2006) Advances in homology protein structure modeling. Curr Protein Pept Sci 7(3):217–227
52. Guo JT, Ellrott K, Xu Y (2008) A historical perspective of template-based protein structure prediction. Methods Mol Biol 413:3–42
53. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T, Maniatis T, Califano A, Honig B (2012) Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature 490(7421):556–560
54. Keskin O, Tuncbag N, Gursoy A (2016) Predicting protein-protein interactions from the molecular to the proteome level. Chem Rev 116(8):4884–4909
55. Orry A, Abagyan R (eds) (2012) Homology modeling: methods and protocols, methods in molecular biology, vol 857. Humana Press
56. Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. Curr Protoc Bioinform 54(1):5.6.1–5.6.37
57. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18(15):2714–2723
58. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engström O, Ohman L, Greene GL, Gustafsson JA, Carlquist M (1997) Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 389(6652):753–758
59. Furukawa A, Konuma T, Yanaka S, Sugase K (2016) Quantitative analysis of protein-ligand interactions by NMR. Prog Nucl Magn Reson Spectrosc 96:47–57
60. Nitsche C, Otting G (2018) NMR studies of ligand binding. Curr Opin Struct Biol 48:16–22
61. Morris GM, Lim-Wilby M (2008) Molecular docking. In: Kukol A. (eds) Molecular modeling of proteins. Methods molecular biology™, vol 443. Humana Press
62. Sousa SF, Fernandes PA, Ramos MJ (2006) Protein-ligand docking: current status and future challenges. Proteins 65(1):15–26
63. Bortolato A, Fanton M, Mason JS, Moro S (2013) Molecular Docking Methodologies. In: Monticelli L, Salonen E (eds) Biomolecular simulations. Methods in molecular biology (methods and protocols), vol 924. Humana Press, Totowa
64. Trisciuzzi D, Alberga D, Leonetti F, Novellino E, Nicolotti O, Mangiatordi GF (2018) Molecular docking for predictive toxicology. In: Nicolotti O (ed) Computational toxicology. Methods in molecular biology, vol 1800. Humana Press, New York
65. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein-small molecule docking methods. J Comput Aided Mol Des 16(3):151–166
66. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3(11):935–949
67. Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47(4):409–443
68. Chen YC (2015) Beware of docking! Trends Pharmacol Sci 36(2):78–95
69. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791
70. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52(4):609–623
71. Ng HW, Zhang W, Shu M, Luo H, Ge W, Perkins R, Tong W, Hong H (2014) Competitive molecular docking approach for predicting estrogen receptor subtype α agonists and antagonists. BMC Bioinform 15(Suppl 11):S4
72. Celik L, Davey J, Lund D, Schiøtt B (2008) Exploring interactions of endocrine-disrupting compounds with different conformations of the human estrogen receptor alpha ligand binding domain: a molecular docking study. Chem Res Toxicol 21(11):2195–2206
73. Lindahl E (2015) Molecular dynamics simulations. In: Kukol A (eds) Molecular modeling of proteins. Methods in molecular biology (Methods and protocols), vol 1215. Humana Press, New York

74. Mortier J, Rakers C, Bermudez M, Murgueitio MS, Riniker S, Wolber G (2015) The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. Drug Discov Today 20(6):686–702
75. Steinbrecher T, Elstner M (2013) QM and QM/MM Simulations of Proteins. In: Monticelli L, Salonen E (eds) Biomolecular simulations. Methods in molecular biology (Methods and protocols), vol 924. Humana Press, Totowa
76. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652
77. Field MJ, Bash PA, Karplus M (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem 11(6):700–733
78. Ng HW, Shu M, Luo H, Ye H, Ge W, Perkins R, Tong W, Hong H (2015) Estrogenic activity data extraction and in silico prediction show the endocrine disruption potential of bisphenol A replacement compounds. Chem Res Toxicol 28(9):1784–1795
79. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package. WIREs Comput Mol Sci 3(2):198–210
80. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718
81. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802
82. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30(10):1545–1614
83. Yang XH, Xie HB, Chen JW, Li XH (2013) Anionic phenolic compounds bind stronger with transthyretin than their neutral forms: nonnegligible mechanisms in virtual screening of endocrine disrupting chemicals. Chem Res Toxicol 26(9):1340–1347
84. Yang XH, Lyakurwa F, Xie HB, Chen JW, Li XH, Qiao XL, Cai XY (2017) Different binding mechanisms of neutral and anionic poly-/perfluorinated chemicals to human transthyretin revealed by In silico models. Chemosphere 182:574–583
85. Johnson ER, Keinan S, Mori-Sánchez P, Contreras-García J, Cohen AJ, Yang W (2010) Revealing noncovalent interactions. J Am Chem Soc 132(18):6498–6506
86. Mahadevi AS, Sastry GN (2016) Cooperativity in noncovalent interactions. Chem Rev 116(5):2775–2825
87. Klebe G (2013) Protein-ligand interactions as the basis for drug action. In: Klebe G (ed) Drug design. Springer, Berlin, Heidelberg
88. Ferreira de Freitas R, Schapira M (2017) A systematic analysis of atomic protein-ligand interactions in the PDB. Medchemcomm 8(10):1970–1981
89. Gallina AM, Bork P, Bordo D (2014) Structural analysis of protein-ligand interactions: the binding of endogenous compounds and of synthetic drugs. J Mol Recognit 27(2):65–72
90. Avvakumov GV, Cherkasov A, Muller YA, Hammond GL (2010) Structural analyses of sex hormone-binding globulin reveal novel ligands and function. Mol Cell Endocrinol 316(1):13–23
91. Rowe RK, Ho PS (2017) Relationships between hydrogen bonds and halogen bonds in biological systems. Acta Crystallogr B Struct Sci Cryst Eng Mater 73(Pt 2):255–264
92. Glättli A, Daura X, Seebach D, van Gunsteren WF (2002) Can one derive the conformational preference of a beta-peptide from its CD spectrum? J Am Chem Soc 124:12972–12978
93. Williams SP, Sigler PB (1998) Atomic structure of progesterone complexed with its receptor. Nature 393(6683):392–396
94. Celik L, Lund JD, Schiøtt B (2007) Conformational dynamics of the estrogen receptor alpha: molecular dynamics simulations of the influence of binding site structure on protein dynamics. Biochemistry 46(7):1743–1758

95. Politzer P, Lane P, Concha MC, Ma Y, Murray JS (2007) An overview of halogen bonding. J Mol Model 13(2):305–311
96. Zhang Q, Xu Z, Shi J, Zhu W (2017) Underestimated Halogen bonds forming with protein backbone in protein data bank. J Chem Inf Model 57(7):1529–1534
97. Auffinger P, Hays FA, Westhof E, Ho PS (2004) Halogen bonds in biological molecules. Proc Natl Acad Sci USA 101(48):16789–16794
98. Zhang Q, Xu Z, Zhu W (2017) The underestimated halogen bonds forming with protein side chains in drug discovery and design. J Chem Inf Model 57(1):22–26
99. Weiss JM, Andersson PL, Zhang J, Simon E, Leonards PE, Hamers T, Lamoree MH (2015) Tracing thyroid hormone-disrupting compounds: database compilation and structure-activity evaluation for an effect-directed analysis of sediment. Anal Bioanal Chem 407(19):5625–5634
100. Yang XH, Cai XY, Chen JW, Li XH, Xie HB, Qiao XL (2014) Effects of halogenation on binding interaction between halogenated thyroid disrupting chemicals and transthyretin (in Chinese). Chin Sci Bull (Chin Ver) 59:2673–2680
101. Zhuang S, Zhang C, Liu W (2014) Atomic insights into distinct hormonal activities of Bisphenol A analogues toward PPARγ and ERα receptors. Chem Res Toxicol 27(10):1769–1779
102. Kumar S, Nussinov R (2002) Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys J 83(3):1595–1612
103. Li F, Xie Q, Li XH, Li N, Chi P, Chen JW, Wang ZJ, Hao C (2010) Hormone activity of hydroxylated polybrominated diphenyl ethers on human thyroid receptor-beta: in vitro and in silico investigations. Environ Health Perspect 118(5):602–606
104. Yang XH, Liu HH, Yang Q, Liu JN, Chen JW, Shi LL (2016) Predicting anti-androgenic activity of bisphenols using molecular docking and quantitative structure-activity relationships. Chemosphere 163:373–381
105. Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51:2778–2786
106. Lu L, Zhan T, Ma M, Xu C, Wang J, Zhang C, Liu W, Zhuang S (2018) Thyroid disruption by bisphenol S analogues via thyroid hormone receptor β: in vitro, in vivo, and molecular dynamics simulation study. Environ Sci Technol 52(11):6617–6625
107. Ren XM, Guo LH (2012) Assessment of the binding of hydroxylated polybrominated diphenyl ethers to thyroid hormone transport proteins using a site-specific fluorescence probe. Environ Sci Technol 46(8):4633–4640
108. Zhang J, Begum A, Brännström K, Grundström C, Iakovleva I, Olofsson A, Sauer-Eriksson AE, Andersson PL (2016) Structure-based virtual screening protocol for in silico identification of potential thyroid disrupting chemicals targeting transthyretin. Environ Sci Technol 50(21):11984–11993
109. Hong HX, Harvey BG, Palmese GR, Stanzione JF, Ng HW, Sakkiah S, Tong W, Sadler JM (2016) Experimental data extraction and in silico prediction of the estrogenic activity of renewable replacements for bisphenol A. Int J Environ Res Public Health 13(7):705
110. Montes-Grajales D, Bernardes GJ, Olivero-Verbel J (2016) Urban endocrine disruptors targeting breast cancer proteins. Chem Res Toxicol 29(2):150–161
111. Wang X, Zhang X, Xia P, Zhang J, Wang Y, Zhang R, Giesy JP, Shi W, Yu H (2017) A high-throughput, computational system to predict if environmental contaminants can bind to human nuclear receptors. Sci Total Environ 576:609–616
112. Trisciuzzi D, Alberga D, Mansouri K, Judson R, Novellino E, Mangiatordi GF, Nicolotti O (2017) Predictive structure-based toxicology approaches to assess the androgenic potential of chemicals. J Chem Inf Model 57(11):2874–2884
113. Vedani A, Dobler M, Hu Z, Smieško M (2015) OpenVirtualToxLab–a platform for generating and exchanging in silico toxicity data. Toxicol Lett 232(2):519–532
114. Organisation for Economic Co-Operation and Development (2007) Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. Technical report for series on testing and assessment no. 69. OECD, Paris
115. Organisation for Economic Co-Operation and Development (2014) Guidance on grouping of chemicals, 2nd edn. Technical report for series on testing and assessment no. 194. OECD, Paris

116. European Chemicals Agency (2015) Read-across assessment framework (RAAF). European Chemicals Agency, Helsinki
117. Myatt GJ, Ahlberg E, Akahori Y, Allen D, Amberg A, Anger LT, Aptula A, Auerbach S, Beilke L, Bellion P, Benigni R, Bercu J, Booth ED, Bower D, Brigo A, Burden N, Cammerer Z, Cronin MTD, Cross KP, Custer L, Dettwiler M, Dobo K, Ford KA, Fortin MC, Gad-McDonald SE, Gellatly N, Gervais V, Glover KP, Glowienke S, Van Gompel J, Gutsell S, Hardy B, Harvey JS, Hillegass J, Honma M, Hsieh JH, Hsu CW, Hughes K, Johnson C, Jolly R, Jones D, Kemper R, Kenyon MO, Kim MT, Kruhlak NL, Kulkarni SA, Kümmerer K, Leavitt P, Majer B, Masten S, Miller S, Moser J, Mumtaz M, Muster W, Neilson L, Oprea TI, Patlewicz G, Paulino A, Lo Piparo E, Powley M, Quigley DP, Reddy MV, Richarz AN, Ruiz P, Schilter B, Serafimova R, Simpson W, Stavitskaya L, Stidl R, Suarez-Rodriguez D, Szabo DT, Teasdale A, Trejo-Martin A, Valentin JP, Vuorinen A, Wall BA, Watts P, White AT, Wichard J, Witt KL, Woolley A, Woolley D, Zwickl C, Hasselgren C (2018) In silico toxicology protocols. Regul Toxicol Pharmacol 96:1–17

Xianhai Yang is Lecturer at the Nanjing University of Science and Technology (NJUST). His major areas of research include: (a) testing and evaluating the endocrine-disrupting effect of emerging organic pollutants; (b) development of in silico models and software for chemical risk assessment, classification, and labeling; (c) ecological and health risk assessment of emerging organic pollutants. Dr. Yang received his Ph.D. in Environmental Engineering from Dalian University of Technology in 2014. He worked in Nanjing Institute of Environmental Science, Ministry of Environmental Protection during 2014–2017, and joined NJUST in 2018.

Huihui Liu is Associate professor at the Nanjing University of Science and Technology (NJUST). Her research interests include: (a) development and application of new passive sampling technology; (b) environmental behavior and toxicological effects of microplastics; (c) computational simulation study on the eco-toxicological effects of organic pollutants. In 2014, Dr. Liu received her Ph.D. in Environmental Sciences from University of Chinese Academy of Sciences and then joined NJUST.

Rebecca Kusko is the SVP and Head of R&D at Immuneering Corporation. She completed her undergraduate degree at MIT in Biological Engineering and earning a Ph.D. in Genetics and Genomics from BU before joining Immuneering in 2014. At Immuneering, she has been accelerating pharma R&D using advanced algorithms. Additionally, she is a board member of the MAQC Society, an initiative around best practices for big data analysis.

Chapter 17
Xenobiotic Metabolism by Cytochrome P450 Enzymes: Insights Gained from Molecular Simulations

Zhiqiang Fu and Jingwen Chen

Abstract Accurate chemical risk assessment requires consideration of the metabolism functioned by the vast majority of enzymes, since neglecting these metabolic pathways (and toxic metabolites) may lead to inaccurate evaluation of their adverse effects on human health. Traditional in vivo or in vitro methods toward this end can be confronted with obstacles, e.g., the huge and ever-increasing number of chemicals, cost and labor-intensive tests, and lack of chemical standards in analysis. Instead, molecular simulations (in silico) are deemed as a promising alternative, which has gradually proven to be feasible for gaining insights into toxicological disposition of xenobiotic chemicals. In this chapter, we review recent progress in molecular simulations of xenobiotic metabolism catalyzed by the typical phase I enzyme: cytochrome P450 enzymes (CYPs). The first section describes the significance of xenobiotic metabolism in chemical risk assessment. Then, the versatile functionality of CYPs in xenobiotic metabolism is briefly summarized by introducing some of the fundamental reactions, e.g., C–H hydroxylation, phenyl oxidation, and heteroatom (N, P, S) oxidation. The last section presents case studies of molecular simulations for metabolism of typical environmental contaminants (e.g., brominated flame retardants, chlorinated alkanes, substituted phenolic compounds), with an emphasis on mechanistic insights gained from quantum chemical density functional theory (DFT) calculations with the active species of CYPs.

Keywords Xenobiotic chemicals · P450 enzymes · Density functional theory · Metabolic mechanisms · Compound I · Environmental contaminants

Z. Fu · J. Chen (✉)
Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology,
Dalian University of Technology, Dalian 116024, China
e-mail: jwchen@dlut.edu.cn

Z. Fu
e-mail: fuzq@dlut.edu.cn

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_17

Abbreviations

ADMET	Absorption, distribution, metabolism, excretion and toxicity
AOPs	Adverse outcome pathways
$ArNH_2$	Primary aromatic amines
BDEs	Bond dissociation energies
Compound I	The active species of P450 enzymes
CYPs	Cytochrome P450 enzymes
DFT	Density functional theory
HAT	Hydrogen atom transfer
HO·	Hydroxyl radical
(HO–)PBDEs	(Hydroxylated) polybrominated diphenyl ethers
IP	Ionization potential
KIEs	Kinetic isotope effects
MD	Molecular dynamics
MIEs	Molecular initiating events
NADPH	Nicotinamide adenine dinucleotide phosphate
NIH shift	National Institute of Health shift
OAR	Oxygen addition rearrangement
PAHs	Polycyclic aromatic hydrocarbons
PBDDs	Polybrominated dibenzo-p-dioxins
PCDD/Fs	Polychlorodibenzo-p-dioxins/furans
PCM	Polarizable continuum model
PFOS	Perfluorooctane sulfonate
QM/MM	Quantum mechanics/molecular mechanics
SET	Single electron transfer
SOM	Site of metabolism
SPCs	Substituted phenolic compounds
SSR	Spin-selective reactivity
TSR	Two-state reactivity
$\Delta E_{ST}(\pi\pi^*)$	Singlet–triplet excitation energy
ΔE	Activation barriers

17.1 Introduction

17.1.1 Significance of Xenobiotic Metabolism in Chemicals Risk Assessment

Organisms are increasingly exposed to numerous xenobiotic chemicals (e.g., environmental contaminants, drugs) via food intake, inhalation, and dermal contacts. It is now generally recognized that these xenobiotics have the potential to disrupt physiological homeostasis, thus threatening human health. Within an organism, xenobiotic

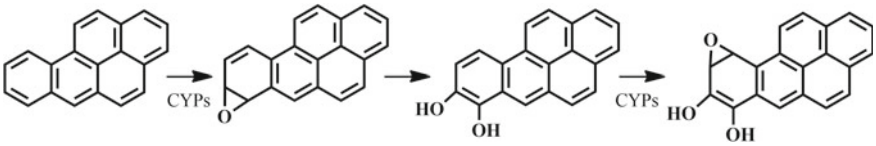

Fig. 17.1 Possible mechanism for P450 bioactivated carcinogenesis of benzo(a)pyrene

chemicals undergo process of "absorption, distribution, metabolism, excretion and toxicity (ADMET)." In this process, metabolism contributes primarily to the elimination of xenobiotics from the biota and thus is one of the major determinants for the biological fate and potential toxicology of xenobiotics. In general, xenobiotics are metabolized via two fundamental routes: phase I and phase II transformation. Phase I transformation is commonly referred to as "catabolism" that breaks down large xenobiotic chemicals into smaller units, comprising reactions of oxidation, reduction, hydrolysis, etc. Cytochrome P450 enzymes (CYPs) have been recognized as being dominant among the vast majority of phase I enzymes (peroxidase, hydrolase, dehydrogenase, amine oxidase, xanthine oxidase, etc.) responsible for the catabolic processes.

Phase I transformations render pollutant molecules more water soluble by introducing polar functional groups (e.g., hydroxyl, carboxyl group), thus facilitating their excretion from the body; otherwise, intermediates or products generated therefrom can be more reactive than their parent compounds. These species are prone to covalently bind with bio-macromolecules (e.g., proteins, nucleic acid), inducing "toxicity enhancement" effects. One case demonstrating this point is polycyclic aromatic hydrocarbons (PAHs), which are typical pollutants known for their carcinogenic, teratogenic, and mutagenic effects. It is reported that the carcinogenicity of PAHs is mainly attributed to the bioactivation by CYPs [1]. Figure 17.1 depicts one possible mechanism via which a typical PAH, benzo(a)pyrene, is bioactivated to be carcinogenic by CYPs. Benzo(a)pyrene is first transformed into phenyl epoxides by CYPs, which is then hydrolyzed to catechol analogues via a ring-opening process. Subsequently, these catechol intermediates are subjected to a secondary bioactivation by CYPs. The resultant dihydrodiol epoxides presumed the ultimate products that induce carcinogenesis. As another example, the persistent organic pollutants polybrominated diphenyl ethers (PBDEs) become more potent endocrine disruptors when transformed into hydroxylated PBDEs (HO-PBDEs) by CYPs [2].

In short, enzymatic transformations serve as a significant determinant for the distribution, fate, and toxicological effects of xenobiotics. Broadly, enzymatic transformation of xenobiotics pertains to one genre of molecular initiating events (MIEs) that subsequently lead to toxicological effects or adverse outcome pathways (AOPs). Therefore, investigation of the metabolic mechanisms is vital for toxicology and health risk assessment of xenobiotic environmental contaminants.

17.1.2 Molecular Simulation of Typical Xenobiotics Metabolism Catalyzed by P450 Enzymes

17.1.2.1 Introduction of P450 Enzymes and Related Metabolic Reactions

P450 enzymes are commonly found in tissues of humans, wildlife, and microorganisms. In mammals, these enzymes reside mainly in the endoplasmic reticulum and mitochondrial inner membranes of liver cells. P450 enzymes represent an enzyme superfamily characterized by a heme-containing active center, which consists of a protoporphyrin substituted with four methyl, two ethenyl, two ionizable propionate groups as well as certain axial/vertical ligands (Fig. 17.2). The catalytic cycle of P450 enzymes starts from a resting (reduction) state, wherein the porphyrin iron binds with H_2O to give a metastable state. When H_2O is replaced by CO, the complex displays maximum absorbance at a wavelength of 450 nm; thus, the enzymes are termed "P450". P450 enzymes are known as the nature's most versatile biological catalyst [3]. The catalytic capability of P450 s covers a broad range of chemicals, involving the functional groups –OH, –CHO, –COOH, $–NH_2$, –CN, phenyl, and halogens. The reactions mediated by P450 enzymes are mostly oxidations (Fig. 17.2), including alkane C–H hydroxylation, alkene C=C epoxidation, and heteroatom (N, P, S) oxidation. In these oxidations, P450 enzymes function as "monooxygenases" that insert an oxygen atom into substrates. Other reactions may also display the exceptional functionality of P450 enzymes, e.g., reductive dehalogenation of halogenated alkanes and C–C bond coupling [4].

With the completion of the human genome project, it is known that human beings have 57 P450 enzyme genes. Animals generally have more P450 genes than humans; for instance, the number of P450 genes for a mouse is 101 and even up to 120 for a sea urchin. The P450 superfamily is named with a nomenclature that comprises family, subfamily, and isoforms. Enzymes in the same family require an amino acid homology >40%, which is marked with Arabic numerals, e.g., the CYP 1 family; enzymes in the subfamily have >55% gene sequences in common and are labeled with capital letters, e.g., the "CYP1A subfamily"; the enzyme subfamily constitutes various isoforms, labeled with Arabic numerals, e.g., "CYP1A2". Different P450 isoforms accommodate their specific substrates; e.g., P4501A2 binds primarily with the aromatic compounds [5]. Exposure to environmental contaminants can also induce special P450 isoforms; e.g., the flame retardant PBDEs are metabolized mainly by CYP2B6 in organisms [6]. Of all the catalytic reactions by P450 enzymes, about 90% is accomplished by isoforms of CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4, with CYP3A4 being dominant [7].

Early studies on P450 enzymes were mainly focused on the metabolism and disposition of carcinogens, pharmaceuticals, and steroids. Two schemes are adopted in these studies, namely (a) in vivo and (b) in vitro experiments. In the in vivo scheme, enzyme inducers (e.g., phenobarbital) are fed to animals and possible metabolites in the tissues, urine, and feces from these animals are then studied; the in vitro scheme

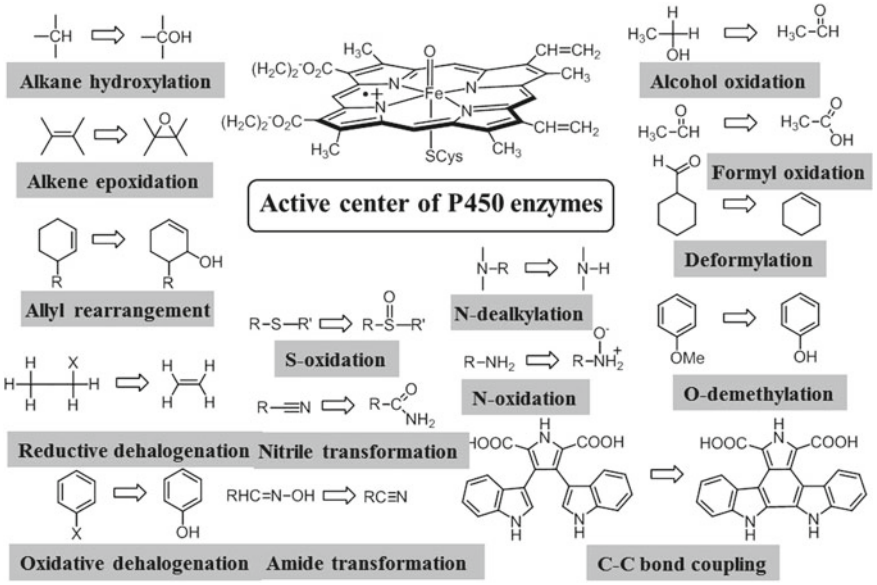

Fig. 17.2 Structure of the active center of P450 enzymes and related catalytic reactions

is mainly based on incubation with liver microsomes, slices, or recombinant P450 enzymes. Although the in vivo scheme is the closest to the metabolic scenario in mammals and humans, it suffers from several interfering factors, e.g., the adsorption and entrapment of xenobiotics by other proteins (serum albumin, etc.), the participation of other metabolizing enzymes, and the deficiency in detecting phase I products. Another major holdback for in vivo testing is animal ethical concerns, namely the widely known 3R principles of "Replacement, Reduction and Refinement." Although microsomal incubations contain minute amount of hydrolytic and phase II enzymes, it is generally accepted that P450 enzymes are dominant. In vitro experiments require simple conditions, e.g., constant temperature and pH, and shaking bath, whereas the detection of metabolites is restricted by a lack of chemical standards. In addition, reaction mechanisms cannot be unveiled through in vitro testing since the transient information of the reaction process cannot be captured with the current instruments (UV–Vis spectra, electronic paramagnetic resonance, Mossbauer spectroscopy, etc.). In fact, P450 enzymes are species- and isoform-different, which are distinct from each other in residue numbers, types, and even in secondary and tertiary structures. These differences lead to distinctive reaction kinetics and product distributions for oxidation of one single substrate by different species, which further brings about difficulties in metabolic studies of P450 enzymes. In recent years, the advancement of quantum chemical theory and molecular simulation techniques has made it possible to simulate the metabolic mechanisms of xenobiotics catalyzed by P450 enzymes. These simulation techniques have gradually garnered interest as important research

alternatives. In fact, the existing knowledge of the catalytic cycle of P450 enzymes has been complemented or supported by quantum chemical calculations.

17.1.2.2 Catalytic Cycle of P450 Enzymes and Related Common Reaction Genres

P450 enzymes commonly act as monooxygenases that transform exogenous chemicals (e.g., alkanes, RH) into oxidative products and H_2O using an O_2 molecule and nicotinamide adenine dinucleotide phosphate (NADPH). The overall reaction can be formulated as: $RH + O_2 + NADPH + H^+ \rightarrow ROH + NADP^+ + H_2O$. Through more than half a century's work, much of the mystery in P450 chemistry has been unveiled, among which the catalytic cycle is the most eye-catching part [4, 8]. As seen in Fig. 17.3, the catalytic cycle starts from a resting state (**1**), wherein the heme iron is hexa-coordinated with H_2O and cysteine in the axial direction and four porphyrin nitrogen atoms in the radial direction. This trivalent iron atom is almost coplanar with the porphyrin ring, rendering the whole system a low-spin doublet state. The binding water molecule is then expelled as the substrate enters the active site, giving a pentacoordinate Fe^{III} complex (**2**) that transports the iron from the heme plane to a position below the porphyrin ring. Complex **2** is in a high-spin sextet state that displays a strong electron-withdrawing capacity. Subsequently, this complex would receive an electron from NADPH and synchronously bind with one O_2 molecule, leading to singlet species Fe^{III}–O–O^- (**3**). **3** is still a good electron receptor and further reduced to the anion peroxo Fe^{III}–O–O^{2-} species (**4**) by absorbing a second electron. **4** is a strong Lewis base that readily gets protonated to the Fe^{III}–O–OH^- species (**5**, Compound 0) by capturing a H^+ from the active site. Compound 0 remains a Lewis base that undergoes a coupling reaction (Coupling-I) via receiving another H^+ from the enzyme, and produces a H_2O molecule and high-valent iron-oxo (Fe^{IV}=O) complex, which is also called Compound I (**6**). Compound I is accepted as the terminal oxidant that transforms substrates to oxidized products (SubO). The release of SubO from the active site and rebinding of H_2O accomplish the catalytic cycle and return the system back to the resting state. Overall, P450 enzymes require two electrons, two H^+ as well as one O_2 to fulfill the catalytic oxidative reactions, with one oxygen atom incorporated into H_2O and the other inserted into the substrate.

It deserves mentioning that the mechanism for protonation of Compound 0 is still controversial. As shown in Fig. 17.3, when the protonation initiates on the distal O atom, Compound I is yielded concomitant with O–O bond scission; otherwise, an iron hydrogen peroxide Fe^{III} (O_2H_2) complex (**8**) is produced when the protonation takes place on the proximal O atom. This Fe^{III} (O_2H_2) complex is susceptible to an uncoupling reaction by losing H_2O_2 and returning back to **2** in the case that H_2O_2 binds loosely with the iron. In a similar way, the resting state **1** can be reactivated to **8** via a shunt pathway. When H_2O_2 binds tightly with Fe in complex **8** (e.g., stabilized by the surrounding residues via hydrogen bonding), the enzyme takes on another coupling reaction scenario (Coupling-II) wherein one H_2O is detached [9] and gives rise to Compound I.

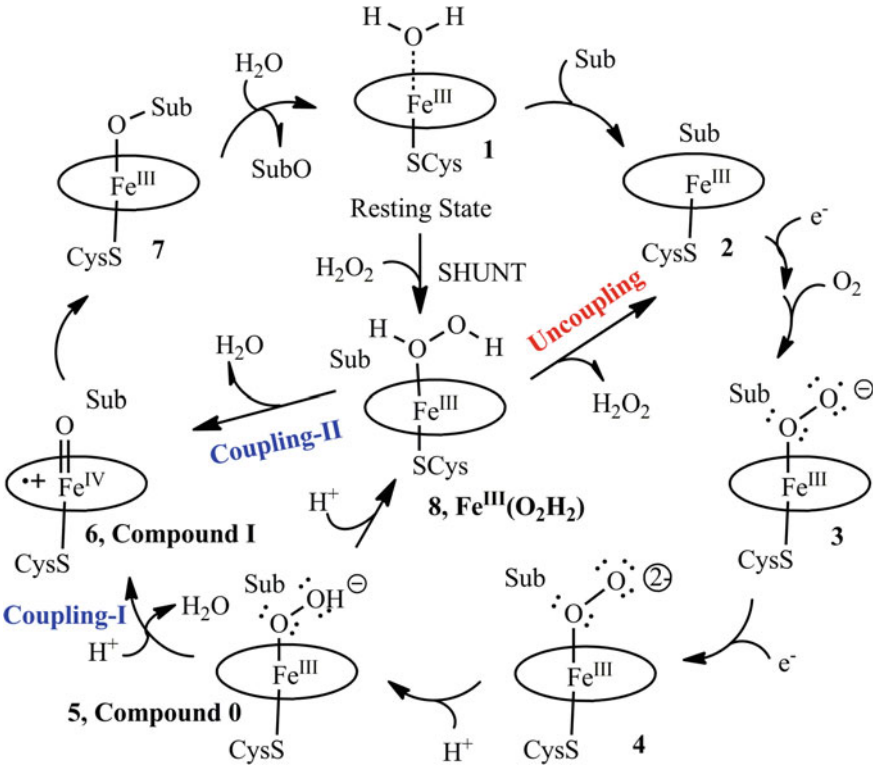

Fig. 17.3 Illustration of the catalytic cycle of P450 enzymes (adapted from Wang et al. [41]; the porphyrin is represented with an ellipse, CysS represents the cysteine residue, Sub is the symbol for the substrate, and the species in the cycle is labeled with 1, 2, 3, …)

Although several species in the cycle (e.g., **4**, **5**, **8**) have been reported to be involved, Compound I is currently the most widely accepted active species in P450 catalyzed reactions. Compound I is known to have a high oxidative reactivity and transient character, which makes it hard to capture by techniques such as electron paramagnetic resonance (EPR), UV–Vis, and Mössbauer spectroscopy. In 2010, Rittle and Green [10] successfully isolated and characterized Compound I for the first time in CYP119 from a thermophilic bacteria, using stopped-flow and rapid freeze–quench techniques. On the whole, there is a lack of direct experimental knowledge on the formation and catalytic mechanisms of Compound I. Instead, the electronic structure and unique reactivity of Compound I are better described with quantum chemical simulations.

17.2 Method and Materials

With the fast advancement of quantum chemical theories and high-performance computing capacities, quantum chemical methods such as density functional theory (DFT) calculations can now accurately reproduce the electronic configuration and properties of various systems and have demonstrated advantages in predicting chemical reactivity. However, DFT methods are restricted by the system size, and converged results obtained on a system with >200 atoms are less reliable [11]. Therefore, common practice is simulating P450 enzymes by adopting a simplified model of the active center (Cpd I in Fig. 17.4), also known as a cluster model in DFT calculations. The simplification is done by truncating the adjacent substituents (Fig. 17.2) to protoporphyrin and replacing the vertical cysteine residue with –SH, $–SCH_3$, or –SCys. The enzyme environment is then mimicked with implicit solvation models, e.g., the polarizable continuum model (PCM). The basic idea for probing P450 catalyzed reactions using the cluster model lies in a general fact: Although P450 enzyme structures vary with species and isoforms, they share analogous active centers that are intrinsically responsible for their metabolic reactions. Thus, simulating the active center with relatively accurate quantum chemical methods would serve to effectively answer chemistry-related questions such as reaction conformations, electronic structures, and reactivity.

As the computing capacity improves, cluster models can better describe more complex systems by incorporating sufficiently large numbers of atoms. Nevertheless, small models based on the active site retain their superiority in dealing with reaction mechanisms at the early stage. Firstly, small models are suitable for quick probing of various reaction routes because of a low computational cost. Secondly, employment of simple models avoids artifacts and tends to receive more accurate results. One general rule in computational biological chemistry is that when a large discrepancy occurs between large and small model results, results from the smaller models are more likely to be correct [11]. It remains difficult to obtain computationally correct results for simulations using large models.

Harris et al. [12] first investigated the electronic structures for the resting state of P450cam, an isoform which specifically binds with camphor, using combined quantum chemical Hartree–Fock and molecular dynamics (MD) calculations. Consistent with the electron spin echo envelope modulation (ESEEM) spectroscopic data, a low-spin doublet state was characterized for the resting state, which is stabilized by the electrostatic interactions with residues surrounding the active site and the ligated water molecule. Subsequent work by Shaik et al. [8] further probed the electronic structures and properties of other intermediates in the P450 catalytic cycle and unveiled several key factors that determine the catalytic reactivity of P450 enzymes. One factor concerns the donor ability of the –SH substituent, also termed "push effects," since the Fe–S distance in Compound I would affect the electronic configuration and thus its oxidative capability. Secondly, the protonation mechanisms of Compound 0 would decide the productivity of Compound I, in a way that an ineffective protonation at the proximal oxygen atom would possibly activate O_2 into

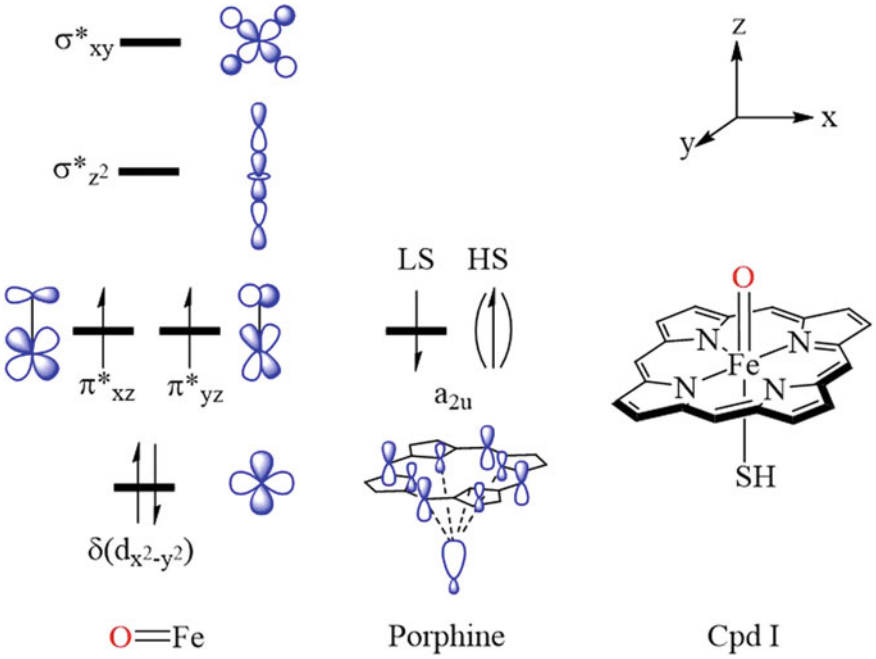

Fig. 17.4 Geometric structure and orbital occupations of Compound I (LS and HS represent, respectively, the low-spin doublet and high-spin quartet states, Cpd I is the abbreviation of Compound I, and porphine means the porphyrin macrocycle)

H_2O_2, which failed to oxidize the substrate while consuming an equivalent amount of reducing agents NADH/NADPH. The accessibility of water molecules to the active site is also an important determinant as water molecules can facilitate proton delivery and protonation of the intermediates; otherwise, redundant water molecules would lead to ineffective protonation in the case when the substrate binds loosely with the enzyme pocket. For the last two decades, simulations using cluster models have successfully uncovered the catalytic mechanisms for P450 enzymes, with contributions from Shaik and Yoshizawa [13–16] who investigated the C–H hydroxylation mechanisms using dozens of alkane substrates, de Visser and Kamachi [17, 18] who probed the C=C bond epoxidation, Shaik and Harvey [19] who uncovered the mechanisms for benzene hydroxylation, and Sharma et al. [20] who elucidated the mechanism for sulfur oxidation. Therefore, quantum chemical calculations based on DFT and cluster model for the active site have made a critical difference in unveiling the mechanisms for P450 catalytic reactions.

In addition, thermodynamics and kinetics of enzymatic reactions would be susceptible to tertiary structures of proteins [21]. For example, amino acid residues surrounding the active site would exert influences on the entry and binding mode of substrates and thus determine the site of metabolism (SOM); weak interactions between the substrate and the residues/water molecules would serve to stabilize the

reaction intermediates; the fast release of the products would facilitate the recovery of the resting state for the next cyclic process [4]. Taking together, the surrounding enzymatic matrix can influence the computational results in two main aspects: steric and electrostatic effects. Thus, these effects should not be neglected in studying P450 catalytic reactions, which demands on simulations with multiscalar techniques, such as the combined quantum mechanics/molecular mechanics (QM/MM) method.

In 2002, Thiel et al. [22] first performed a DFT(B3LYP)/MM study which delved into properties of intermediates in the catalytic cycle and mechanisms for Compound 0 protonation and C–H bond hydroxylation by P450 enzymes. Thereafter, Guallar [23, 24] conducted QM/MM investigations with a restricted open-shell DFT method, termed DFT(ROB3LYP)/MM, to probe the *H*-abstraction mechanism by P450cam. They found that the electrostatic interactions between the negatively charged propionate of the porphyrin side chain and the positively charged surrounding residues would facilitate C–H hydroxylation. In the past decade, QM/MM investigations on P450 enzymes have gradually been increasing, with P450cam being the dominant isoform. For example, to resolve the controversy of a "second oxidant" in P450cam catalytic cycle, Shaik et al. [9] employed QM/MM calculations to dissect into the reactivity of three intermediate species, i.e., Compound I, Compound 0, and the ferric hydrogen peroxide Fe^{III} (O_2H_2). It was shown that the persistence and oxidative reactivity of Fe^{III} (O_2H_2) depend primarily on the interplay of camphor and protein. The presence of camphor in the active site blocks the release of H_2O_2 (the uncoupling pathway in Fig. 17.3), which expedites its homolytic O–O bond cleavage to a HO· radical and Fe^{IV}-OH species. The HO· radical is further adjusted by hydrogen bonding with adjacent amino acid residues to an appropriate position that facilitates the *H*-abstraction from Fe^{IV}–OH by the HO· radical, leading to Compound I and H_2O. Hence in this case, QM/MM simulations serve to better understand the protonation of intermediates in the catalytic cycle.

In other cases, QM/MM investigations have adequately unveiled some of the long-standing puzzles in P450 catalytic cycle and contributed to understanding of the regio- and stereoselectivity in oxidative reactions. By using MD and QM/MM techniques, Ramanan et al. [25] simulated the binding of P450-BM3 with fatty acids and revealed that the substrate's binding with active site residues (e.g., Arg241 and Pro242) determines the regio- and enantioselective reactivity for fatty acid hydroxylation. These studies further verify that quantum chemical calculations are capable of predicting SOM and products based on molecular structures (i.e., ab initio). It is anticipated that QM/MM and MD calculations, along with future improvements in computational theory and capability, could pave the way for accurate prediction for metabolic mechanisms of xenobiotics catalyzed by P450 enzymes.

Through half a century's endeavors, people have acquired in-depth knowledge of the reaction mechanisms and metabolic profile of P450 enzymes. As noted above, the truncated cluster model, Compound I, contributed greatly in unveiling the reaction mechanisms. The geometric structure and electronic orbital occupations for Compound I are depicted in Fig. 17.4. The typical structure for Compound I (Fig. 17.4) comprises a protoporphyrin with a high-valent iron (IV) atom and –SH representing cysteine, an Fe=O bond of ca. 1.65 Å and Fe–S bond of ca. 2.30 Å.

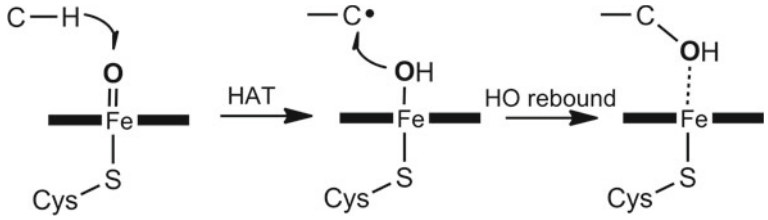

Fig. 17.5 Consensus mechanism for C–H hydroxylation by Compound I

Compound I is a cation radical, for which the electron spin states are decided by three unpaired electrons, i.e., two on the antibonding orbitals π^*_{xz} and π^*_{yz} of Fe=O and one on the delocalized a_{2u} orbital of the porphyrin macrocycle. For a low-spin doublet state, two spin-up electrons on π^*_{xz} and π^*_{yz} orbitals were combined with one spin-down electron in the a_{2u} orbital to give a count of one single unpaired electron; otherwise, Compound I would be in the high-spin quartet state with the three unpaired electrons in the same spin direction. In some cases, the system would even get excited by moving two paired electrons from a bonding orbital of Fe=O to two antibonding orbitals higher in energy, forming a high-spin sextet state. The doublet and quartet states are energetically degenerated (equivalent) states for Compound I and equally considered in the reaction mechanisms described below.

17.3 Results/Case Studies

17.3.1 Common Reaction Genres Mediated by P450 Enzymes and Related Mechanisms

17.3.1.1 Mechanisms for C–H Bond Hydroxylation

C–H bond hydroxylation pertains to the most thoroughly studied reaction functioned by P450 enzymes. In 2000, Ogliaro et al. [26, 27] first investigated C–H hydroxylation of methane catalyzed by Compound I with DFT calculations. Subsequently, numerous studies have probed dozens of alkane substrates and reached a consensus mechanism for C–H hydroxylation, i.e., the hydrogen atom transfer (HAT) followed by hydroxyl rebound scenario. First proposed by Groves et al., the C–H hydroxylation scenario (Fig. 17.5) starts from a *H*-abstraction from the C–H bond by the O atom of iron-oxo (Fe–O), giving rise to a carbon-centered radical and iron hydroxyl (Fe–OH) intermediate. After that, Fe–OH reoriented with the O atom pointing to the carbon radical, which is followed by a final OH rebound to give an alcohol product.

Computational studies [28] have also revealed that C–H bond hydroxylation by Compound I proceeds via a two-state reactivity (TSR) scenario which involves both doublet and quartet states in the reaction. The first HAT step has almost equivalent

activation barriers (with differences ≤2 kcal/mol) on both spin states and thus is rate-determining in the hydroxylation process. However, the second hydroxyl rebound step on the doublet state is barrierless, in contrast to a slight rebounding barrier needed for the quartet state reaction.

The rate-determining HAT step in C–H hydroxylation is generally characterized by large kinetic isotope effects (KIEs). KIEs refer to the change in reaction rates when one atom (typically the H atom) of the reactant is replaced by one of its isotopes (e.g., deuterium) [29]. According to the study by Li et al. [30], KIEs can serve as a spin-state reactivity probe because the TSR possesses similar KIEs values, whereas significantly discrepant KIEs for two spin states indicate a spin-selective reactivity (SSR) scenario in *H*-abstractions by Compound I. In addition, Shaik et al. [15] found a linear correlation between the activation barriers of alkane hydroxylation by Compound I with the corresponding bond dissociation energies (BDEs) of C–H bonds. BDEs are defined as the energy difference between the optimized substrate molecule (Sub-H) and the optimized substrate radical (Sub·) and H atom, i.e., BDE(Sub-H) = E(Sub-H) − E(Sub·) − E(H). However, this linear correlation becomes insignificant for C–H bonds with lower BDEs, e.g., Cα-H bonds adjacent to strong electron-withdrawing groups.

17.3.1.2 Mechanisms for Heteroatom (N, S, P) Oxidation

Heteroatom (N, S, P) oxidation pertains to another type of important reaction catalyzed by P450 enzymes. *N*-containing amines can be activated by P450 enzymes into products that may induce genotoxicity and carcinogenicity. Metabolism of secondary and tertiary amines by P450 enzymes would primarily proceed via *N*-dealkylation for which two possible mechanisms (Fig. 17.6) are reported [29]. Analogous to the HAT scheme for C–H bond hydroxylation, one mechanism suggests an initial Cα–H abstraction by Compound I with the yield of an alkane radical and $PorFe^{IV}$–OH intermediate, which is followed by hydroxyl rebound to form alcohol amine products. The other mechanism hypothesizes that a single electron transfer (SET) from the amine N atom to Compound I results in a N cation radical, which is then deprotonated to the alcohol amine. Bifurcate HAT and SET mechanisms have long been a controversial issue; only recently, quantum chemical calculations have ascertained that HAT is generally lower in activation barriers and thus more favorable than SET.

N-dealkylation of secondary and tertiary amines would yield primary amines as products. Further oxidation of primary aromatic amines ($ArNH_2$) can proceed via the following four mechanisms as shown in Fig. 17.7, three of which adopt Compound I as the oxidant, i.e., HAT followed by hydroxyl rebound, oxygen addition rearrangement (OAR), and SET followed by proton transfer (PT). The other pathway, however, involves the iron-superoxide species ($FeOO^{2-}$) in the catalytic cycle as the active oxidant. Computational studies [31] have revealed that the HAT and HO-rebound pathways require the lowest energy barrier and hence are the principal pathways for $ArNH_2$ oxidation by P450 enzymes. Nevertheless, significant energy barrier increases were observed for the rate-limiting HAT with the enhancement

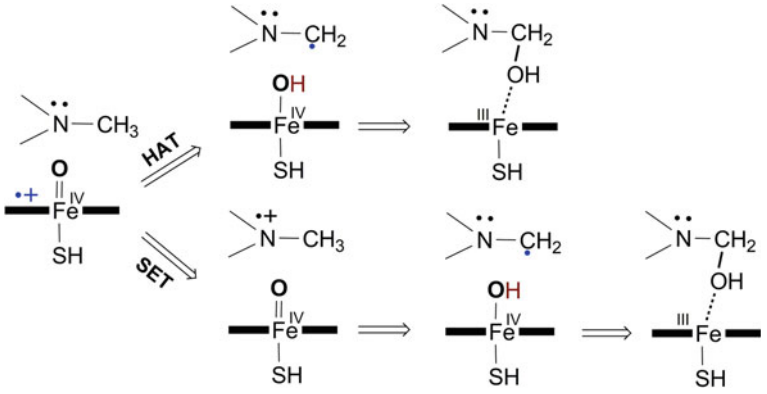

Fig. 17.6 Bifurcate mechanisms for *N*-dealkylation catalyzed by P450 enzymes

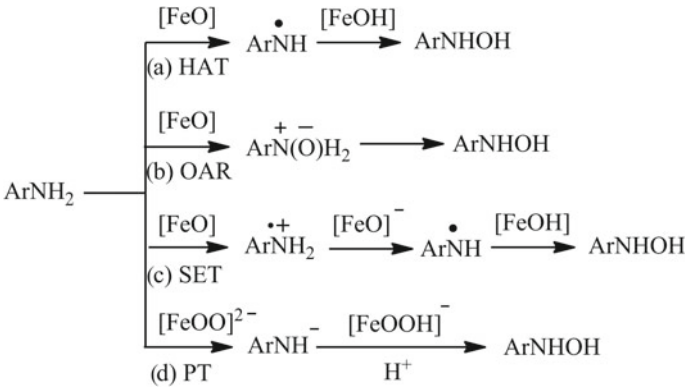

Fig. 17.7 Possible mechanisms for *N*-hydroxylation of aromatic primary amines by P450 enzymes. [FeO] is a simplification of Compound I, ArNH2 means primary aromatic amines, and pathways are labeled with **a**, **b**, **c,** and **d**

of electron-withdrawing abilities of substituents at the para position, indicating the inertness of such $ArNH_2$ toward oxidation by P450 enzymes. The mechanism of oxidation of aliphatic primary amines differs from that of $ArNH_2$ in two aspects: Firstly, *N*-oxidation of aliphatic primary amines takes on only two possible pathways of OAR and HAT; secondly, the oxygen rebound rather than *H*-abstraction becomes rate-determining in the feasible HAT pathway that leads to alcohol amine formation, which is also discrepant from the C–H hydroxylation mechanism.

Oxidation of sulfur-containing compounds (e.g., thiolethers) catalyzed by P450 enzymes produces sulfoxide or sulfone as the main products. *S*-oxidation involves an initial *O*-addition of Compound I to the sulfur atom to form sulfoxide, which is successively converted to sulfone by the secondary *O*-addition. The mechanistic elusiveness for *S*-oxidation mainly focuses on the possible "second oxidant" that participates in the reaction. Three reactive species in the catalytic cycle, i.e., Com-

pound I, Compound 0, and $FeOO^{2-}$, were previously evaluated for their reactivity toward *S*-oxidation. A cluster model study by Li et al. [32] concluded Compound I as the preponderant oxidant in *S*-oxidation of dimethyl sulfide. The same reaction was revisited [33] by QM/MM simulations, which also indicated that Compound 0 is a sluggish oxidant and less competent than Compound I for sulfoxidation.

17.3.1.3 Mechanisms for Phenyl Hydroxylation

Phenyl groups are prevalent in molecules of environmental contaminants, e.g., polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), and polychlorodibenzo-p-dioxins/furans (PCDD/Fs). When subjected to metabolism by P450 enzymes, the phenyl group would likely be transformed into epoxides that are reactive electrophilic agents toward bio-macromolecules, or yield hydroxylated products which are suspected endocrine disruptors. As a consequence, mechanisms for phenyl hydroxylation by P450 enzymes have attracted much attention. Owing to the strong electronic π-conjugation, phenyl C–H bonds possess BDE values larger than those of alkyl C–H bonds, which renders the disobedience of the phenyl oxidation mechanisms from alkane C–H hydroxylation.

Using DFT calculations, Shaik et al. first explained the mechanism for benzene hydroxylation [19] and clarified that the SET (as in *N*-hydroxylation) and HAT (as in alkyl hydroxylation) schemes are less viable and the reaction proceeds primarily on the low-spin doublet state. Three feasible mechanisms are illustrated in Fig. 17.8. According to previous studies, these mechanisms share a preceding rate-determining electrophilic π-addition of Compound I to the phenyl carbon, leading to a radical- or cation-like tetrahedral adduct. The radical-type intermediate is prone to subsequent epoxidation via addition of the O atom to the adjacent C atom, while the cation tetrahedral intermediate is liable to arrange via a National Institute of Health (NIH) shift that delivers the ipso-H atom to the neighboring C atoms, giving a cyclohexanone product. Moreover, if the tetrahedral intermediate takes on a "side-on" conformation wherein the phenyl is perpendicular to the porphyrin macrocycle, a proton shuttle mechanism can be viable that shifts the H atom first to porphyrin N, followed by H bounce back to the carbonyl O or neighboring atoms, resulting in phenol or cyclohexanone, respectively. QM/MM calculations [34] have revealed that epoxides and ketones are the two products most likely to be formed from "face-on" transition state conformations wherein the phenyl is parallel to the porphyrin, whereas epoxide and phenol products are favorable for "side-on" conformations.

For halogenated phenyl groups, the π-addition of Compound I to the halogen-substituted phenyl carbons requires increased activation barriers due to steric hindrance. In this case, the π-addition would initiate a NIH shift of the halogen atom, yielding a cyclohexanone product. Furthermore, when the phenyl is perhalogenated or the halogen NIH shift is hampered by steric hindrance, the oxidative reaction would lead to dehalogenated products. For substituted benzenes, DFT studies [35] have revealed lower activation barriers for π-addition of Compound I to phenyl car-

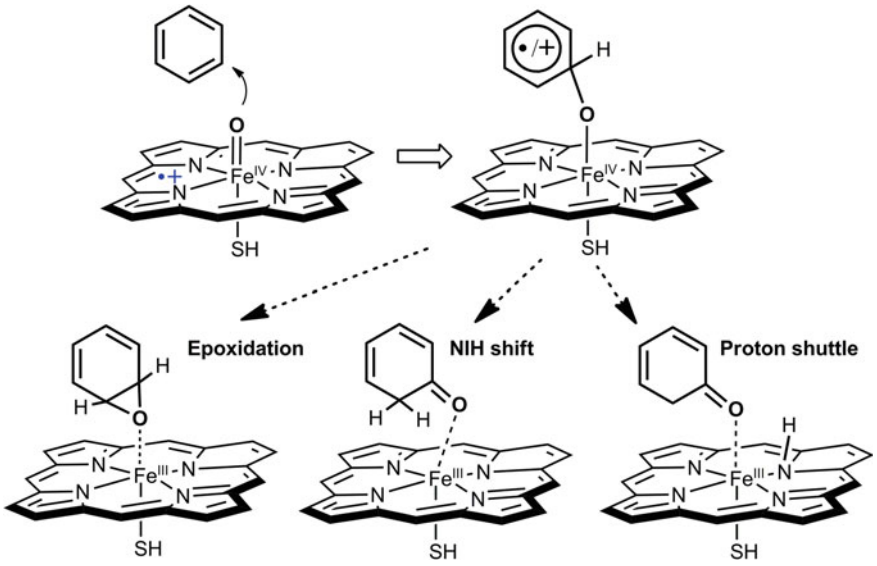

Fig. 17.8 Three principal reaction pathways (epoxidation, NIH shift, and proton shuttle) for benzene hydroxylation catalyzed by P450 enzymes

bons with substituents at the para position, which are in linear correlation with the Hammett parameters (ρ) of the substituents. The π-addition reactions at phenyl carbon para to the substituents are generally more facile than those at meta positions. Shaik et al. [36] uncovered that the π-addition barriers of Compound I with substituted benzenes were related to the ionization potential (IP) and singlet–triplet excitation energy $\Delta E_{ST}(\pi\pi^*)$ of substrate molecules. IP refers to the electronic energy difference of optimized structures between the substrate with one electron removed and the neutral substrate molecule, i.e., IP $= E(\text{Sub}^+) - E(\text{Sub})$. $\Delta E_{ST}(\pi\pi^*)$ is electronic energy difference between the triplet and the singlet states of substrates, namely $\Delta E_{ST}(\pi\pi^*) = E(\text{triplet}) - E(\text{singlet})$. According to a recent kinetic study [37], oxidation of aromatics, especially those with redox potentials lower than Compound I, would first initiate an electron transfer in the solvent cage to yield phenyl radical cations, coupled with the subsequent C–O bond formation step to give the tetrahedral adducts. This could intrinsically explain the reason why the π-addition barriers for Compound I are linearly correlated with the IPs of substrate molecules.

17.3.2 Case Studies Simulating Xenobiotic Chemical Metabolism by P450 Enzymes

17.3.2.1 Metabolism of PBDEs by P450 Enzymes: DFT Studies Using the Active Species of P450s

Existing computational case studies on P450 metabolism focused primarily on conventional chemical substrates (e.g., alkanes, alkenes, and benzene), endogenous hormones, pharmaceuticals, etc. By contrast, the metabolic profile of vast majority of environmental pollutants catalyzed by P450 enzymes is yet to be detailed. Flame retardants are typical pollutants that attract worldwide attention, especially polybrominated diphenyl ethers (PBDEs), demonstrated to be of environmental persistence (P), bioaccumulation potential (B), and toxicity potential (T) by numerous studies. As the primary metabolite of PBDEs, HO-PBDEs have been reported to possess enhanced endocrine disrupting potency and mitochondrial toxicity compared with PBDEs. Recent in vitro studies indicated that apart from HO-PBDEs, metabolism of PBDEs by P450 enzymes would lead to dihydroxylated and even to the notorious polybrominated dibenzo-p-dioxins (PBDDs) as products [38, 39], for which the underlying molecular mechanisms are unclear. In this section, DFT studies predicting the mechanisms for PBDEs metabolism by P450 enzymes are reviewed [40–42].

Based on the metabolic mechanisms for halogenated benzenes [35], Wang et al. [40] proposed oxidation of PBDEs by Compound I that would proceed via the pathways shown in Fig. 17.9, using 2,2′,4,4′-tetraBDE (BDE-47) as the model compound. The preceding π-addition of Compound I (a) to the non-Br-substituted phenyl carbons of BDE-47 leads to tetrahedral adducts that are subjected to further rearrangements (b). Otherwise, reaction with the Br-bonded phenyl carbons drives the NIH shift of Br to cyclohexanones that undergo reduction (d) in the non-enzymatic medium. For the rearrangements of tetrahedral adducts, three pathways are possible, which involve the NIH shift to cyclohexanones, ring closure to epoxides, and proton shuttle to phenol products. The epoxides are then rearranged (c) via proton-assisted ring-opening, giving rise to multiple products including HO-PBDEs and bromophenols.

According to the computational study by Wang et al. [40], π-addition of Compound I to Br-substituted phenyl carbons (C_2 and C_4) was more energy-demanding than non-substituted carbons (C_3, C_5, C_6), except for the C_1 site that possessed an increased barrier compared with non-substituted carbons due to large steric hindrances. The π-addition of Compound I to non-substituted carbons is endothermic with the resultant tetrahedral intermediate lying higher on the potential energy surface than the reactant. However, π-addition of Compound I to Br-substituted carbons is strongly exothermic and yields cyclohexanones via NIH shift of the Br atoms, a reaction analogous to the dehalogenation of hexachlorobenzene. Particularly, π-addition to the C_2 position leads to the expelling of a bromide ion due to the steric effects. The tetrahedral adducts formed from π-additions at the C_1, C_3, C_5, C_6 sites

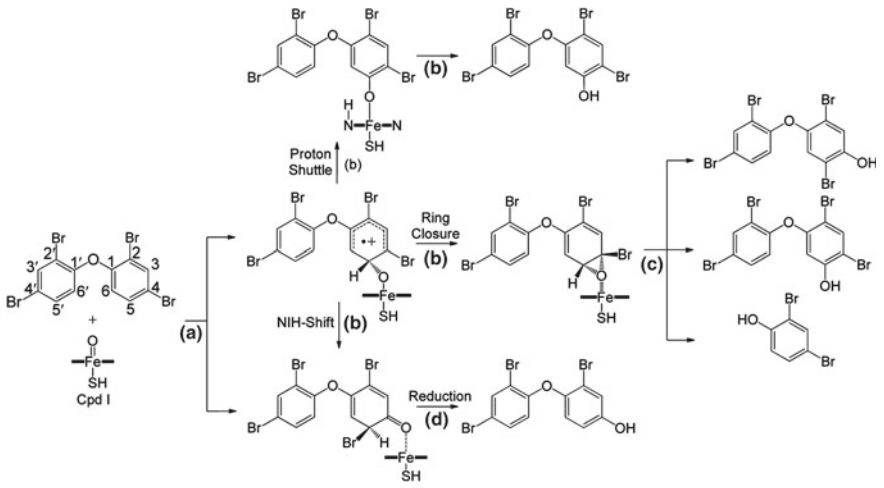

Fig. 17.9 Possible pathways for the reaction BDE-47 with Compound I

further rearrange via ring closure to six possible epoxides. As the epoxidation step generally requires low activation barriers (<5 kcal/mol), the precedent π-addition is rate-determining in BDE-47 oxidation catalyzed by Compound I. Formation of cyclohexanones and epoxides manifests the accomplishment of the monooxygenation functionality of Compound I and retrieves the resting state of P450 enzymes.

Since the formed BDE-47 epoxides are weakly bound with the porphyrin iron, they are easily released out of the active site to the bulk polar environment. These epoxides are prone to protonation in the non-enzymatic environment, which cleaves the epoxy bond barrierless to two types of ring-opened hydroxyl-cyclohexadienyl products, i.e., with –OH attached to Br-substituted and non-substituted phenyl carbons, respectively. These hydroxyl-cyclohexadienyl products are subjected to further rearrangement, with a scheme shown in Fig. 17.10.

When –OH is attached to Br-substituted carbons (#1), the hydroxyl-cyclohexadienyls rearrange via shifting the Br atom to neighboring carbons, followed by a proton loss to yield HO-PBDEs. Hydroxyl-cyclohexadienyls with –OH attached to non-substituted carbons (#2) would shift the H atom to proximate carbons, which subsequently deprotonate to form HO-BDE-47. Computational results have demonstrated that the NIH shift of Br is more preferable than the H atom, which leads to an alteration of the brominated pattern in HO-PBDEs. It is noteworthy that attachment of –OH to the ether carbon (C_1) induces homolytic or heterotic fission of the ether bond (#3), resulting exclusively in bromophenol products. Calculated activation barriers reveal that heterotic fission of the ether bond is more feasible than the homolysis reaction. The predicted products, e.g., 4-HO-BDE-42, 4′-HO-BDE-49, 5-HO-BDE-47, and 2,4-dibromophenol, are in line with the metabolites detected in in vitro experiment. Thus, DFT calculations are effective tools for elucidating

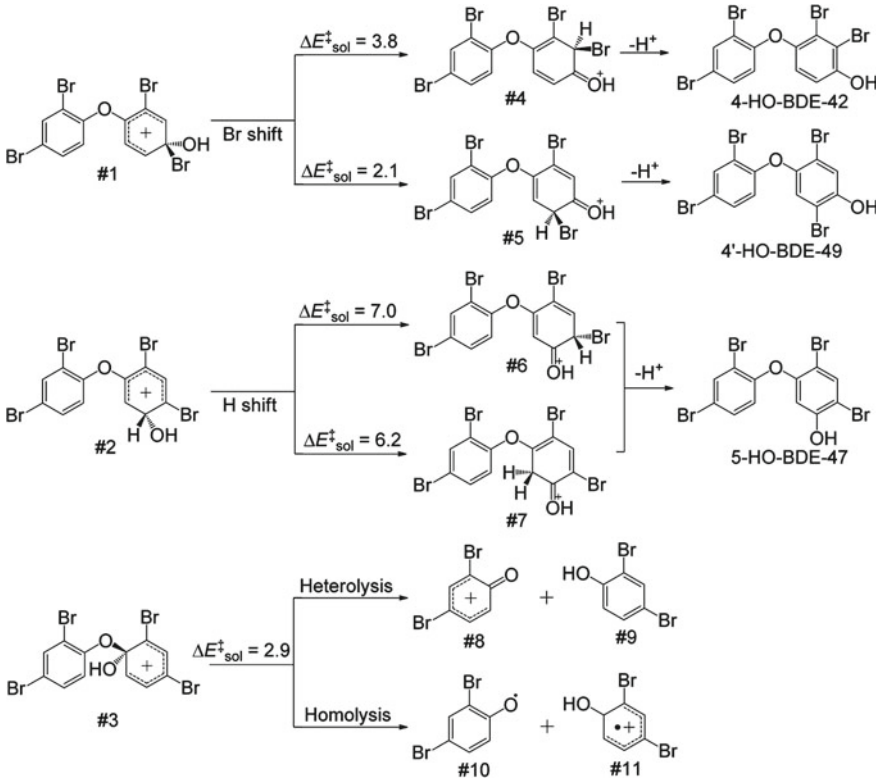

Fig. 17.10 Rearrangement scheme of the ring-opened hydroxyl-cyclohexadienyl products

the mechanisms and product distribution for BDE-47 hydroxylation catalyzed by Compound I.

The incipient π-addition is rate-determining in hydroxylation of BDE-47 by Compound I, as the subsequent rearrangement reactions generally have lower activation barriers. A comparison of the π-addition barriers for three PBDE congeners (BDE-15, -47, and -153) revealed that a higher degree of bromination decreased the potential of PBDEs to be oxidized by Compound I [42]. Lupton et al. [43] investigated the metabolism of three PBDEs (BDE-47, -99, and -153) by human liver microsomal incubations and also observed that the highly brominated BDE-153 was more inert to P450 metabolism relative to lower brominated ones. For one single PBDE congener, the most accessible reaction sites resided on the non-brominated and non-bridged carbons. However, this site-selective trend diminished with the increase of Br substitutions.

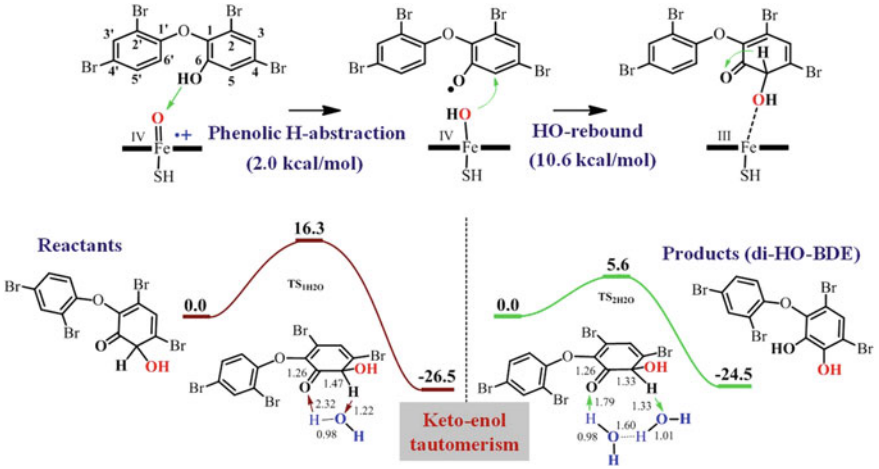

Fig. 17.11 Enzymatic (Compound I catalyzed) and non-enzymatic (keto–enol tautomerism) pathways for dihydroxylation of HO-PBDE to di-HO-BDE

17.3.2.2 Mechanisms for Dihydroxylation of PBDEs and Dioxin Evolution

Apart from HO-PBDEs, in vitro incubations of PBDEs also identified dihydroxylated PBDEs (di-HO-BDEs) and dioxin (PBDDs) as minor products. One incubation experiment utilizing rat liver microsomes revealed that HO-PBDEs are the precursor compounds for di-HO-BDE formation [44]. Taking 6-HO-BDE-47 as an example, Fu et al. [42] proposed two possible pathways for the reaction of HO-PBDEs with Compound I. One pathway concerned the π-addition of Compound I to phenyl carbons of 6-HO-BDE-47, a process analogous to the hydroxylation of PBDEs, resulting in tetrahedral adducts that further rearrange to di-HO-BDEs. In the other pathway, a successive phenolic *H*-abstraction and hydroxyl rebound scenario converted HO-PBDE to the hydroxyl cyclohexanone intermediate catalyzed by Compound I. Subsequently, keto–enol tautomerism of the hydroxyl cyclohexanone leads to the di-HO-BDE product (Fig. 17.11).

DFT calculations [42] revealed that HO-PBDEs acquire much lower π-addition barriers (15.9 kcal/mol) than those of PBDEs, implying the introduction of –OH increases electron densities on the phenyl group, thus facilitating the electrophilic π-addition. For the second pathway (Fig. 17.11), phenolic *H*-abstraction of 6-HO-BDE-47 is quite facile with a tiny barrier of 2.0 kcal/mol, while the succedent hydroxyl rebound had relatively a high barrier of 10.6 kcal/mol and thus was rate-determining in the whole reaction. Keto–enol rearrangement of the hydroxylated cyclohexanones to di-HO-BDEs proceeded with the assistance of H_2O in the non-enzymatic environment. Increasing the number of water molecules was shown to significantly decrease activation barriers of the keto–enol rearrangement. The participation of two water

molecules greatly expedited the rearrangement reaction in the physiological environment by lowering the barrier to 5.6 kcal/mol. By comparing the activation barriers for the rate-limiting steps in the two pathways, it can be concluded that the phenolic *H*-abstraction and hydroxyl rebound pathway were responsible for dihydroxylation of PBDEs catalyzed by Compound I. These results support the absence of epoxide hydrolase in in vitro observation [45]. For different HO-PBDE congeners (6-HO-BDE-15, -47, -153), the rate-limiting hydroxyl rebound proceeds preferably on the phenyl carbons ortho and para to the phenol, leading to the formation of catechol- or hydroquinone-like di-HO-BDEs, which are in line with the identified structures in the experiment.

PBDDs can have different sources in the environment, e.g., pyrolysis of PBDEs, photochemical transformation of HO-PBDEs, free radical oxidation of HO-PBDEs in the atmosphere, and biochemical oxidative coupling of bromophenols mediated by marine organisms (e.g., sponge and red algae) [46]. Molecular mechanisms for several cases, e.g., PBDD evolution from PBDEs pyrolysis and hydroxyl radical (HO·)-initiated oxidation of HO-PBDEs, have been elucidated based on previous DFT calculations [47, 48]. Though perceived as feasible, these mechanisms clearly possess certain prerequisites; e.g., the direct dissociation of an ortho Br or H atom in PBDE pyrolysis requires a tremendous amount of energy; the HO·-initiated oxidation of HO-PBDEs involves the participation of multiple HO· species; the photolysis of HO-PBDEs to yield PBDD is based on the existence of excited triplet states for the reactants. It is therefore obvious that the inaccessibility of these prerequisites precludes the viability of the aforementioned mechanisms for PBDD evolution in P450 enzymatic reactions.

Structurally, only ortho-hydroxylated PBDEs can potentially be transformed into PBDDs. Using 6-HO-BDE-47 as an example, two conventional pathways were proposed [42], including: (a) phenolic *H*-abstraction and successive O–C cyclization of 6-HO-BDE-47 to form PBDD; (b) a secondary hydroxylation of 6-HO-BDE-47 yields 6,6′-di-HO-BDE-47, which then dehydrates to PBDD. The calculated results reveal that the O–C cyclization step has a high activation barrier and the resultant PBDD radical requires multiple assisted water molecules to fulfill the *H*-rearrangement, which is tentatively inviable in the enzymatic environment. In pathway (b), hydroxylation of 6-HO-BDE-47 possesses a barrier approximate to PBDEs hydroxylation, indicating that the dihydroxylation is viable if the substrate is properly orientated. However, the dehydration of 6,6′-di-HO-BDE-47 to PBDD is highly energy-demanding; thus, this pathway can also be ruled out.

As indicated in the previous studies, bromophenols can be photo-transformed into HO-PBDEs [49], following a scheme of aryl radical coupling. Specifically, bromophenols are firstly converted to phenoxyl radicals that possess electronically three resonance structures, i.e., one *O*-centered phenoxyl radical, two C-centered radicals ortho and para to the phenoxyl. Coupling of the *O*-centered phenoxyl radical with the C-centered radicals leads to the formation of HO-PBDEs products. Inspired by the radical coupling scheme of the phenoxyl, it is anticipated that only heterocyclic di-HO-BDEs with –OH substituted on phenyl carbons ortho and meta to the ether bond can serve as precursors for PBDD.

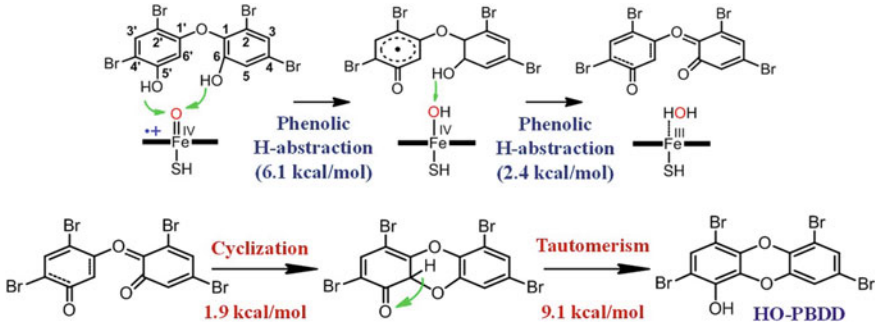

Fig. 17.12 Feasible pathway for HO-PBDD evolution from HO-PBDE metabolism catalyzed by Compound I

DFT studies have elucidated a feasible pathway (Fig. 17.12) for PBDD evolution using 5′,6-di-HO-BDE as the example substrate. Compound I first oxidizes 5′,6-di-HO-BDE to a diketone intermediate via two consecutive phenolic *H*-abstractions. The diketone is electronically in resonance with a biradical structure with the radicals resided on O_6 and $C_{6'}$. Aryl biradical coupling of the diketone intermediate generates a PBDD ketone isomer, which undergoes keto–enol tautomerism non-enzymatically to HO-PBDD. In the initial step, the phenolic *H*-abstraction is facilitated by hydrogen bonding between the O atom of Compound I and the hydroxyl H atom of 5′,6-di-HO-BDE. Calculated barriers indicate that phenolic *H*-abstractions, cyclization of the diketone intermediate, and tautomerism of the PBDD isomer are quite facile (<10 kcal/mol). However, PBDDs are observed as minor products in in vitro incubations of PBDEs, possibly due to a low efficient yield of di-HO-PBDEs or less substrate accessibility to the active site.

17.3.2.3 Simulation of the Metabolism of Perfluorooctane Sulfonate Precursors Catalyzed by P450 Enzymes

Perfluorooctane sulfonate (PFOS) pertains to one representative of persistent organic pollutants, yet its exposure profile in biota remains implicit. As indicated by in vivo and in vitro experiments, metabolism of PFOS precursors (PreFOS) is an important indirect exposure pathway for PFOS, whereas the underlying molecular mechanisms are largely unclarified. Fu et al. [50] investigated the metabolism of one typical PreFOS, *N*-ethyl perfluorooctane sulfonamide (*N*-EtPFOSA) catalyzed by Compound I. As shown in Fig. 17.13a, the metabolism of *N*-EtPFOSA proceeds via *N*-deethylation, which comprises a precedent Cα-H hydroxylation (a) and subsequent ethanolamine decomposition. Initially, the Cα-H abstraction of *N*-EtPFOSA catalyzed by Compound I produces a carbon-centered radical species, which combines with the rebounding hydroxyl to an ethanolamine intermediate. Subsequently,

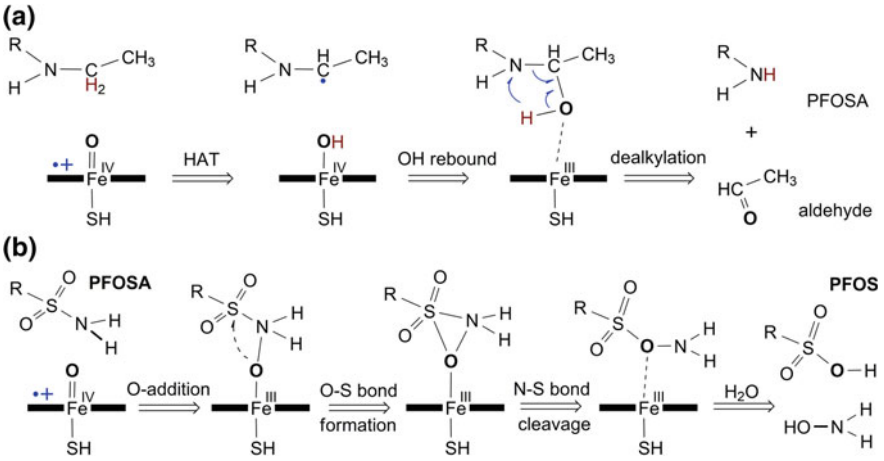

Fig. 17.13 Mechanisms for metabolism of *N*-EtPFOSA to PFOS catalyzed by Compound I (**a** *N*-dealkylation of *N*-EtPFOSA to PFOSA and **b** deamination of PFOSA to PFOS)

decomposition of the ethanolamine in the non-enzymatic environment leads to acetaldehyde and PFOSA.

H-abstraction is the rate-limiting step in Cα-H hydroxylation process. Both the doublet and the quartet spin states acquire similar activation barriers, indicating that the Cα-H hydroxylation follows the two-state reactivity scenario. The decomposition of the ethanolamine initiates via shifting the H atom from the alcohol hydroxyl to the amine N atom, which induces the fission of the Cα-N bond and expels the ethyl group. Calculated activation barriers for decomposition of the ethanolamine bound with the porphyrin iron are significantly higher than those for the dissociative ethanolamine, meaning that the degradation of ethanolamine is more conceivable in the non-enzymatic aqueous medium. Calculations have indicated that water molecules can act as catalysts in the degradation, which assist the H-transfer through hydrogen bonding. Therefore, PreFOS substances can be subjected to *N*-dealkylation to form PFOSA once exposed in biota.

As corroborated by experimental results, PFOSA is identified as the primary intermediate product in biotransformation of almost all PreFOS, whereas the mechanisms pertinent to PFOS formation from PFOSA remain obscure. DFT calculations have elucidated a feasible pathway for PFOS evolution from PFOSA metabolism catalyzed by Compound I. *H*-abstraction from PFOSA was found unlikely due to the strong electron-withdrawing ability of the perfluorooctyl group, in contrast to the traditional *H*-abstraction and hydroxyl rebound scheme reported for amine oxidation. Metabolism of PFOSA starts from *O*-addition by Compound I (Fig. 17.13b), giving a *N*-oxide intermediate that either rearranges through H-shift to a hydroxylamine derivative or transforms to an epoxide analogue via O–S bond formation. The epoxide analogue rearranges afterward via N–S bond cleavage, followed by hydrolysis to yield PFOS and hydroxylamine.

Results revealed that the H-shift pathway leading to the hydroxylamine derivative has a higher barrier than the one generating the epoxide analogue. Therefore, the deamination of PFOSA to PFOS proceeds via an epoxidation mechanism, wherein the initial *O*-addition (*N*-oxidation) step is rate-determining. By contrasting the activation barrier for *N*-dealkylation of *N*-EtPFOSA to that for deamination of PFOSA, it can be concluded that transformation of PFOSA to PFOS is the rate-limiting step in the whole process of PreFOS biotransformation, which supports the in vitro observation [51].

17.3.2.4 Metabolism of Halogenated Alkanes and Alkenes Catalyzed by P450 Active Species

Halogenated alkyl moieties are frequently found in insecticides, pesticides, disinfection by-products, etc. By using DFT calculations, Ji et al. [52] investigated the metabolic scenario of two examples of halogenated alkanes, $CHCl_3$ and CCl_4 catalyzed by P450 enzymes. Results revealed two distinct pathways for the reaction of $CHCl_3$ with Compound I under both aerobic and anaerobic conditions. $CHCl_3$ was aerobically oxidized through C–H hydroxylation to $C(OH)Cl_3$, which further undergoes dehydrochlorination with the assistance of water molecules, leading to Cl_2O. Under anaerobic conditions, however, $CHCl_3$ preferably reacted with the divalent ferroporphyrin via reductive dehalogenation. For the perchlorinated CCl_4, aerobic oxidation via C–Cl abstraction was energetically infeasible, whereas a low O_2 concentration environment effectively obtained Cl_2O and $ClO\cdot$ as products. Aside from halogenated alkanes, alkenes with unsaturated C=C bonds are increasingly concerned due to their potential to be transformed into toxic epoxide intermediates. Zhang et al. [53] computed the activation barriers (ΔE, in kcal/mol) for epoxidation of 36 alkenes catalyzed by Compound I by using DFT calculations and found a strong linear correlation between ΔE and the ionization potential (IP) of these alkenes (Fig. 17.14): For alkene molecules with dipole moment <2.2 debyes (D), $\Delta E = 5.044IP - 31.315$, whereas $\Delta E = 2.666IP - 16.066$ for those with >2.2 D dipole moments. These correlations can serve as quick prediction tools for the epoxidation reactivity of alkenes oxidized by Compound I.

17.3.2.5 Metabolism of Substituted Phenolic Compounds Catalyzed by P450 Enzymes: A Novel Ipso-Substitution Pathway

Substituted phenolic compounds (SPCs), including bisphenol analogues, alkylphenols, and chlorophenols, are ubiquitous environmental contaminants that have drawn widespread attention due to their potential endocrine disrupting properties. Recent in vitro studies indicated that P450-catalyzed metabolism significantly enhances the endocrine disrupting activity of SPCs by arousing an alternative ipso-substitution pathway that converts them to hydroquinone with elimination of substituents at the ipso-position, the mechanism of which is far beyond elucidated. Ji et al. [54] investi-

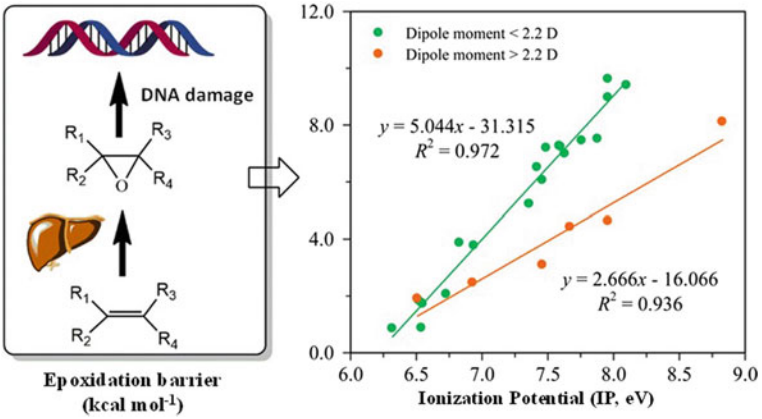

Fig. 17.14 Prediction of the metabolism of alkenes to epoxides catalyzed by P450 enzymes

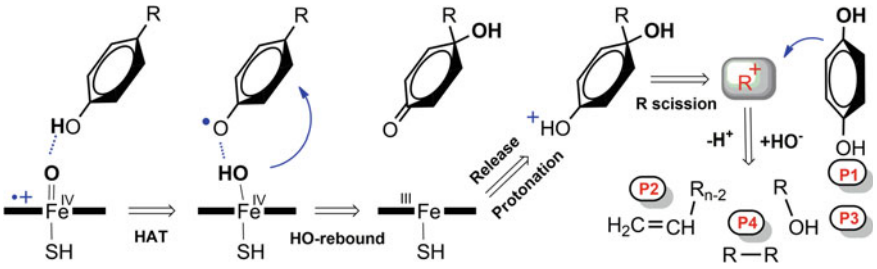

Fig. 17.15 Elucidated mechanism for ipso-substitution of substituted phenolic compounds catalyzed by Compound I

gated the P450-catalyzed ipso-substitution pathway by means of DFT calculations. Results uncovered that the ipso-substitution proceeded via precedent *H*-abstraction from SPCs by Compound I, followed by barrierless HO rebound to the ipso-position giving a quinol (Fig. 17.15), which can be spontaneously protonated and decomposed to carbocation (R^+) and hydroquinone (P1). This carbocation can further evolve into olefin (P2) and the highly estrogenic hydroxylated (P3) and dimer-type metabolites (P4). These identified mechanisms are beneficial for metabolite prediction from phenolic endocrine disruptors whose fate is affected by this alternative P450 reactivity, and thus enable screening of the metabolites for endocrine disrupting activity.

17.4 Conclusion and Future Directions

It remains a vital and difficult task to effectively predict the disposition and potential toxicology of environmental chemicals in chemical risk assessment. In this chapter,

molecular simulation studies with cluster models and DFT calculations are shown to provide a mechanism-based explicit prediction for the metabolism of xenobiotics catalyzed by P450 enzymes. Modeling the whole enzyme using hybrid quantum mechanics/molecular mechanics (QM/MM) methodology with accurate force field parameters and intensive samplings from molecular dynamic simulations may further resolve the difference of metabolic kinetics and product distribution among various P450 isoforms or other enzymes that participate in xenobiotic disposition in biota.

From a broader aspect, effective identification of the molecular initiating events (MIEs) that lead eventually to adverse outcomes requires accurate simulation of the interactions between xenobiotic chemicals and bio-macromolecules. Molecular simulations are undeniably applicable for this purpose and could further be adopted to develop quantitative predictive models for various parameters (binding constants, partition coefficients, even transformation rate constants, etc.) that determine the fate and toxicity of xenobiotic chemicals, provided that the computing capacity is greatly advanced and certain environmental/biological systems are properly treated. With a combination of these simulation techniques and quantitative models, it is possible to compile computational toxicology models/software that enable precise prediction of xenobiotic disposition and toxicology through a simple mouse clicking.

References

1. Rendic S, Guengerich FP (2012) Contributions of human enzymes in carcinogen metabolism. Chem Res Toxicol 25(7):1316–1383
2. Hamers T, Kamstra JH, Sonneveld E, Murk AJ, Visser TJ, Van Velzen MJM, Brouwer A, Bergman A (2008) Biotransformation of brominated flame retardants into potentially endocrine-disrupting metabolites, with special attention to 2,2',4,4'-tetrabromodiphenyl ether (BDE-47). Mol Nutr Food Res 52(2):284–298
3. Coon MJ (2005) Cytochrome P450: nature's most versatile biological catalyst. Annu rev pharmacol 45:1–25
4. Shaik S, Cohen S, Wang Y, Chen H, Kumar D, Thiel W (2010) P450 enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations. Chem Rev 110(2):949–1017
5. Wang B, Zhou SF (2009) Synthetic and natural compounds that interact with human cytochrome P450 1A2 and implications in drug development. Curr Med Chem 16(31):4066–4218
6. Gross MS, Butryn DM, McGarrigle BP, Aga DS, Olson JR (2015) Primary role of cytochrome P450 2B6 in the oxidative metabolism of 2,2′,4,4′,6-pentabromodiphenyl ether (BDE-100) to hydroxylated bdes. Chem Res Toxicol 28(4):672–681
7. Rendic S, Guengerich FP (2015) Survey of human oxidoreductases and cytochrome P450 enzymes involved in the metabolism of xenobiotic and natural chemicals. Chem Res Toxicol 28(1):38–42
8. Shaik S, Kumar D, de Visser SP, Altun A, Thiel W (2005) Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. Chem Rev 105(6):2279–2328
9. Wang BJ, Li CS, Dubey KD, Shaik S (2015) Quantum mechanical/molecular mechanical calculated reactivity networks reveal how cytochrome P450cam and its T252A mutant select their oxidation pathways. J Am Chem Soc 137(23):7379–7390
10. Rittle J, Green MT (2010) Cytochrome P450 Compound I: capture, characterization, and C-H bond activation kinetics. Science 330(6006):933–937
11. Blomberg MRA, Borowski T, Himo F, Liao RZ, Siegbahn PEM (2014) Quantum chemical studies of mechanisms for metalloenzymes. Chem Rev 114(7):3601–3658

12. Harris D, Loew G (1993) Determinants of the spin state of the resting state of cytochrome P450cam. J Am Chem Soc 115(19):8775–8779
13. Kumar D, de Visser SP, Sharma PK, Cohen S, Shaik S (2004) Radical clock substrates, their C–H hydroxylation mechanism by cytochrome P450, and other reactivity patterns: what does theory reveal about the clocks' behavior? J Am Chem Soc 126(6):1907–1920
14. Yoshizawa K, Kamachi T, Shiota Y (2001) A theoretical study of the dynamic behavior of alkane hydroxylation by a Compound I model of cytochrome P450. J Am Chem Soc 123(40):9806–9816
15. de Visser SP, Kumar D, Cohen S, Shacham R, Shaik S (2004) A predictive pattern of computed barriers for C–H hydroxylation by compound i of cytochrome P450. J Am Chem Soc 126(27):8362–8363
16. Shaik S, Kumar D, de Visser SP (2008) A valence bond modeling of trends in hydrogen abstraction barriers and transition states of hydroxylation reactions catalyzed by cytochrome P450 enzymes. J Am Chem Soc 130(31):10128–10140
17. de Visser SP, Ogliaro F, Harris N, Shaik S (2001) Multi-state epoxidation of ethene by cytochrome P450: a quantum chemical study. J Am Chem Soc 123(13):3037–3047
18. de Visser SP, Ogliaro F, Sharma PK, Shaik S (2002) What factors affect the regioselectivity of oxidation by cytochrome P450? A DFT study of allylic hydroxylation and double bond epoxidation in a model reaction. J Am Chem Soc 124(39):11809–11826
19. de Visser SP, Shaik S (2003) A proton-shuttle mechanism mediated by the porphyrin in benzene hydroxylation by cytochrome P450 enzymes. J Am Chem Soc 125(24):7413–7424
20. Sharma PK, de Visser SP, Shaik S (2003) Can a single oxidant with two spin states masquerade as two different oxidants? A study of the sulfoxidation mechanism by cytochrome P450. J Am Chem Soc 125(29):8698–8699
21. van der Kamp MW, Mulholland AJ (2013) Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. Biochemistry 52(16):2708–2728
22. Schöneboom JC, Lin H, Reuter N, Thiel W, Cohen S, Ogliaro F, Shaik S (2002) The elusive oxidant species of cytochrome P450 enzymes: characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. J Am Chem Soc 124(27):8142–8151
23. Guallar V, Friesner RA (2004) Cytochrome P450cam enzymatic catalysis cycle: a quantum mechanics/molecular mechanics study. J Am Chem Soc 126(27):8501–8508
24. Altun A, Guallar V, Friesner RA, Shaik S, Thiel W (2006) The effect of heme environment on the hydrogen abstraction reaction of camphor in P450cam catalysis: A QM/MM study. J Am Chem Soc 128(12):3924–3925
25. Ramanan R, Dubey KD, Wang B, Mandal D, Shaik S (2016) Emergence of function in P450-proteins: a combined quantum mechanical/molecular mechanical and molecular dynamics study of the reactive species in the H_2O_2-dependent cytochrome p450(sp alpha) and its regio- and enantioselective hydroxylation of fatty acids. J Am Chem Soc 138(21):6786–6797
26. Ogliaro F, Cohen S, de Visser SP, Shaik S (2000) Medium polarization and hydrogen bonding effects on Compound I of cytochrome P450: what kind of a radical is it really? J Am Chem Soc 122(51):12892–12893
27. Ogliaro F, Harris N, Cohen S, Filatov M, de Visser SP, Shaik S (2000) A model "rebound" mechanism of hydroxylation by cytochrome P450: stepwise and effectively concerted pathways, and their reactivity patterns. J Am Chem Soc 122(37):8977–8989
28. Kumar D, de Visser SP, Shaik S (2003) How does product isotope effect prove the operation of a two-state "rebound" mechanism in C–H hydroxylation by cytochrome P450? J Am Chem Soc 125(43):13024–13025
29. Wang Y, Kumar D, Yang C, Han K, Shaik S (2007) Theoretical study of N-demethylation of substituted N,N-dimethylanilines by cytochrome P450: the mechanistic significance of kinetic isotope effect profiles. J Phys Chem B 111(26):7700–7710
30. Li CS, Wu W, Kumar D, Shaik S (2006) Kinetic isotope effect is a sensitive probe of spin state reactivity in C–H hydroxylation of N,N-dimethylaniline by cytochrome P450. J Am Chem Soc 128(2):394–395

31. Ji L, Schueuermann G (2013) Model and mechanism: N-hydroxylation of primary aromatic amines by cytochrome P450. Angew Chem Int Edit 52(2):744–748
32. Li CS, Zhang LX, Zhang C, Hirao H, Wu W, Shaik S (2007) Which oxidant is really responsible for sulfur oxidation by cytochrome P450? Angew Chem Int Edit 46(43):8168–8170
33. Porro CS, Sutcliffe MJ, de Visser SP (2009) Quantum mechanics/molecular mechanics studies on the sulfoxidation of dimethyl sulfide by Compound I and Compound 0 of cytochrome P450: which is the better oxidant? J Phys Chem A 113(43):11635–11642
34. Bathelt CM, Mulholland AJ, Harvey JN (2008) QM/MM modeling of benzene hydroxylation in human cytochrome P450 2C9. J Phys Chem A 112(50):13149–13156
35. Hackett JC, Sanan TT, Hadad CM (2007) Oxidative dehalogenation of perhalogenated benzenes by cytochrome P450 Compound I. Biochemistry 46(20):5924–5940
36. Shaik S, Milko P, Schyman P, Usharani D, Chen H (2011) Trends in aromatic oxidation reactions catalyzed by cytochrome P450 enzymes: a valence bond modeling. J Chem Theory Comput 7(2):327–339
37. Asaka M, Fujii H (2016) Participation of electron transfer process in rate-limiting step of aromatic hydroxylation reactions by compound I models of heme enzymes. J Am Chem Soc 138(26):8048–8051
38. Zhai C, Peng SN, Yang LM, Wang QQ (2014) Evaluation of BDE-47 hydroxylation metabolic pathways based on a strong electron-withdrawing pentafluorobenzoyl derivatization gas chromatography/electron capture negative ionization quadrupole mass spectrometry. Environ Sci Technol 48(14):8117–8126
39. Luisa Feo M, Gross MS, McGarrigle BP, Eljarrat E, Barcelo D, Aga DS, Olson JR (2013) Biotransformation of BDE-47 to potentially toxic metabolites is predominantly mediated by human CYP2B6. Environ Health Persp 121(4):440–446
40. Wang XB, Wang Y, Chen JW, Ma YQ, Zhou J, Fu ZQ (2012) Computational toxicological investigation on the mechanism and pathways of xenobiotics metabolized by cytochrome P450: a case of BDE-47. Environ Sci Technol 46(9):5126–5133
41. Wang XB, Chen JW, Wang Y, Xie HB, Fu ZQ (2015) Transformation pathways of MeO-PBDEs catalyzed by active center of P450 enzymes: a DFT investigation employing 6-MeO-BDE-47 as a case. Chemosphere 120:631–636
42. Fu ZQ, Wang Y, Chen JW, Wang ZY, Wang XB (2016) How PBDEs are transformed into dihydroxylated and dioxin metabolites catalyzed by the active center of cytochrome P450s: a DFT study. Environ Sci Technol 50(15):8155–8163
43. Lupton SJ, McGarrigle BP, Olson JR, Wood TD, Aga DS (2009) Human liver microsome-mediated metabolism of brominated diphenyl ethers 47, 99, and 153 and identification of their major metabolites. Chem Res Toxicol 22(11):1802–1809
44. Lai YQ, Cai ZW (2012) In vitro metabolism of hydroxylated polybrominated diphenyl ethers and their inhibitory effects on 17β-estradiol metabolism in rat liver microsomes. Environ Sci Pollut R 19(8):3219–3227
45. Erratico CA, Szeitz A, Bandiera SM (2013) Biotransformation of 2,2′,4,4′-tetrabromodiphenyl ether (BDE-47) by human liver microsomes: Identification of cytochrome P450 2B6 as the major enzyme involved. Chem Res Toxicol 26(5):721–731
46. Agarwal V, Moore BS (2014) Enzymatic synthesis of polybrominated dioxins from the marine environment. ACS Chem Biol 9(9):1980–1984
47. Altarawneh M, Dlugogorski BZ (2013) A mechanistic and kinetic study on the formation of PBDD/Fs from PBDEs. Environ Sci Technol 47(10):5118–5127
48. Cao HJ, He MX, Han DD, Li J, Li MY, Wang WX, Yao SD (2013) OH-initiated oxidation mechanisms and kinetics of 2,4,4′-tribrominated diphenyl ether. Environ Sci Technol 47(15):8238–8247
49. Erickson PR, Grandbois M, Arnold WA, McNeill K (2012) Photochemical formation of brominated dioxins and other products of concern from hydroxylated polybrominated diphenyl ethers (HO-PBDEs). Environ Sci Technol 46(15):8174–8180
50. Fu ZQ, Wang Y, Wang ZY, Xie HB, Chen JW (2015) Transformation pathways of isomeric perfluorooctanesulfonate precursors catalyzed by the active species of P450 enzymes: in silicoinvestigation. Chem Res Toxicol 28(3):482–489

51. Benskin JP, Holt A, Martin JW (2009) Isomer-specific biotransformation rates of a perfluorooctane sulfonate (PFOS)-precursor by cytochrome P450 isozymes and human liver microsomes. Environ Sci Technol 43(22):8566–8572
52. Ji L, Zhang J, Liu WP, de Visser Sp (2014) Metabolism of halogenated alkanes by cytochrome P450 enzymes: aerobic oxidation versus anaerobic reduction. Chem –Asian J 9(4):1175–1182
53. Zhang J, Ji L, Liu WP (2015) In silico prediction of cytochrome P450-mediated biotransformations of xenobiotics: a case study of epoxidation. Chem Res Toxicol 28(8):1522–1531
54. Ji L, Ji SJ, Wang CC, Kepp KP (2018) Molecular mechanism of alternative P450-catalyzed metabolism of environmental phenolic endocrine-disrupting chemicals. Environ Sci Technol 52(7):4422–4431

Zhiqiang Fu obtained his Ph.D. in 2017 from Dalian University of Technology, China, where he currently holds a position of Lecturer, developing research in the "Environmental and Ecological Engineering" group led by Prof. Jingwen Chen. His research has focused on reaction kinetics and mechanisms via which xenobiotics are disposed by enzymes, and QSAR modeling toward predicting equilibrium partitioning behaviors of diverse chemicals.

Jingwen Chen is Professor (since 2001) of environmental chemistry and Dean for the School of Environmental Science and Technology at Dalian University of Technology. He graduated from the Northeast Normal University and received his B.S. in 1991 and M.S. in 1994. After getting Ph.D. from the Nanjing University in 1997, he performed postdoctoral studies at the German Research Center for Environmental Health as an Alexander von Humboldtian. His research interests concern ecological chemistry and computational toxicology on organic chemicals. He has published over 260 peer-reviewed papers, reviews, chapters, and 3 books (2 monographs and 1 textbook). He served as Member of editorial boards of several international (Ecotox. Environ. Safe., J. Enviorn. Sci.) and domestic (e.g., Environ. Chem., Asian J. Ecotox., Mar. Environ. Sci.) journals. In 2013, he won the National Science Fund for Distinguished Young Scholars and was appointed as Chair Professor of "Cheung Kong Scholars Program" by China Ministry of Education.

Chapter 18
Integrating QSAR, Read-Across, and Screening Tools: The VEGAHUB Platform as an Example

Emilio Benfenati, Alessandra Roncaglioni, Anna Lombardo and Alberto Manganaro

Abstract In silico models are evolving toward a more mature view, which integrates several perspectives. This integration proceeds on the application side toward a deeper exploitation of the data and information available, coping toward more challenging tasks. On a theoretical point of view, the QSAR models are nowadays most typically general models, at least in their ambition, while read-across is local. There are also general tools for prioritization. There are common aspects between these approaches, but also peculiar aspects. On the other side, users are interested in the application of these tools, for the evaluation of specific chemicals (which may relate to read-across and QSAR models), or for the assessment of populations of substances, also quite large (which may relate to QSAR and prioritization tools). In the development of VEGA, we tried to be as close as possible to the user's need, reducing the barriers between the different approaches, and providing a series of tools which may fit different purposes. We describe below the philosophy of VEGA, and how the user may take advantage of the complex tools for different purposes.

Keywords QSAR · Read-across · Weight-of-evidence · VEGA · ToxRead · Non-testing methods

E. Benfenati (✉) · A. Roncaglioni · A. Lombardo
Istituto di Ricerche Farmacologiche Mario Negri – IRCCS, Milano, Italy
e-mail: emilio.benfenati@marionegri.it

A. Roncaglioni
e-mail: Roncaglioni@marionegri.it

A. Lombardo
e-mail: anna.lombardo@marionegri.it

A. Manganaro
KODE, Pisa, Italy
e-mail: a.manganaro@kode-solutions.net

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_18

Abbreviations

AD	Applicability domain
ADI	Applicability domain index
BCF	Bioconcentration factor
CLP	Classification, labeling, and packaging
CMR	Carcinogenic, mutagenic, or reprotoxicants
ED	Endocrine disruptors
kNN	k-nearest neighbor
NTM	Non-testing methods
PBT	Persistent, bioaccumulative, and toxic
SA	Structural alerts
SOM	Self-organizing map
WoE	Weight-of-evidence

18.1 Introduction

In the last decades, dramatic changes occurred in the field of non-testing methods (NTM). NTM include QSAR models and read-across. These two areas moved from quite different perspectives and uses. Indeed, QSAR originally was a model developed for research purposes, to evaluate the reasons for the effects within specific chemical categories. The possible relationship between the chemical structure and the property value, in particular the toxicological and ecotoxicological properties, has been studied for decades. In the case of the ecotoxicological properties, most of the studies address certain chemical families and most commonly aquatic acute toxicity [1]. Generally, for these properties, regression models have been developed. In the case of toxicological properties, a number of studies addressed categorical methods to discriminate toxic versus non-toxic compounds. Indeed, many studies focused on mutagenicity via the Ames test [2].

For both toxicological and ecotoxicological studies, initially the assumption was that the model refers to a certain chemical family. In the case of ecotoxicity, the model was applicable to the family, and in the same way, for toxicological categorical models, the effect was associated with the fact that the chemical belongs to the family. Thus, initially, the attention started from the identification of a certain family of compounds, and the reasoning affected this local situation.

In a second phase, the studies explored the possibilities to extend the chemical domain of the model. One approach was simply to develop collections of sub-models. However, the availability of software to calculate a very large number of descriptors, of more sophisticated algorithms, and of more powerful computers at low cost resulted in a dramatic increase of models able to cover many families of compounds for many (eco)toxicological properties. At this point, the QSAR model was in principle a general model [3].

On the other hand, read-across has been used to address individual substances, for practical purposes, mainly related to the assessment of the chemicals subject to authorization. If data on the target compound were missing, the possibility to relate to similar compounds with data was explored and applied. This approach is by its very nature opportunistic because it is strictly applicable to the case where there is at least one similar chemical with experimental values of interest.

However, authorities request more and more elements to support the fact that the data on the source compound are sufficient and analyzing if possible reasons of concern can be excluded. This caused the shift toward a more systematic strategy, addressing theoretical aspects associated with the effect, and thus requiring the exploration of the factors related to the effect. Furthermore, the preference is for the use of not only one similar compound, again in order to further support the evidence [4].

Thus, in a certain way, both in the case of QSAR and read-across, the tendency has been toward more complex scenarios, able to introduce multiple factors which play a role in the toxicity phenomenon, or anyhow in the process under study.

This theoretical change occurred in parallel with the change of the application scenarios of the NTM and in particular of the QSAR models. Indeed, the NTM are more and more commonly used for regulatory purposes [5–8]. In Europe, in vivo methods are banned for cosmetics [9], but for industrial chemicals the presence of alternative to in vivo methods should be explored [10]. The REACH registration since its very first article promotes the use of alternative methods and addresses criteria for the use of QSAR and read-across [11]. Specific documents published by ECHA provide guidance on the use of QSAR and read-across [6, 12]. In addition, EFSA refers to NTM in its guidance documents on weight-of-evidence (WoE) and mixtures, for instance.

Similarly, the US EPA introduced several QSAR models to address human toxicity, ecotoxicity, and environmental properties (https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface). Furthermore, other regulators from other countries, such as Canada and Japan, are using QSAR models [2].

These cases demonstrate the interest in NTM for regulatory purposes. This interest promoted the change in the field of NTM toward a more structured and integrated approach. This approach is able to cope with requests coming from the regulatory bodies addressing the safety of the chemical substances and is thus a general target.

In this chapter, we describe the VEGA platform, which has been developed considering the regulators' point of view.

18.2 Method and Materials

18.2.1 The VEGAHUB Structure

Within the VEGAHUB platform, there are multiple tools, which are dedicated to the exploration and analysis of the properties of chemical substances. The main components of VEGAHAB are:

- VEGA
- ToxRead
- ToxWeight
- ToxDelta
- JANUS.

Furthermore, there are links to general tools, which can be used to develop new models: SARpy [13], VEGA-based tools, SOM (Self-Organizing Map) tool, and CORAL [14].

18.2.2 The QSAR Models in VEGA

VEGA can be freely downloaded from the website www.vegahub.eu, after registration. VEGA offers a collection of QSAR models. Indeed, there are tens of QSAR models available within VEGA, addressing physicochemical, environmental, ecotoxicological, and toxicological properties. Table 18.1 lists these models. We are continuing adding models, and thus, this list is changing rapidly.

There may be more than one model for the same endpoint. This increases the robustness of the overall prediction and the confidence of the final assessment. The different models usually have different origin, including training sets of compounds, chemical descriptors, and algorithms. The specific information on each model, including the chemicals in the training and test sets, is available from VEGA. Figure 18.1 shows where to find this information.

The user can make the prediction for one single chemical, or for a large set of compounds, and the software allow both methods of entry, for one or few chemicals (using the SMILES and an interactive page), or a collection of chemicals in a table.

VEGA automatically checks for the consistency of the SMILES and reports if there are errors. VEGA transforms the SMILES into its specific SMILES format, in order to have reproducible results. The results of each model are not simply the predicted value, but in addition, the most similar chemicals and the evaluation of the reliability of the prediction are shown.

The most similar compounds are identified using a specific program for similarity (the algorithm is described in [15]), which has been optimized to balance different approaches.

Table 18.1 List of the QSAR models within VEGA

Properties			
Toxicological	Ecotoxicological	Environmental	Physicochemical
Mutagenicity (Ames test) CONSENSUS model (version 1.0.2)	Fish acute (LC50) toxicity classification (SarPy/IRFMN) (version 1.0.2)	BCF model (CAESAR) (version 2.1.14)	LogP model (Meylan/Kowwin) (version 1.1.4)
Mutagenicity (Ames test) model (CAESAR) (version 2.1.13)	Fish acute (LC50) toxicity model (KNN/Read-Across) (version 1.0.0)	BCF model (Meylan) (version 1.0.3)	LogP model (MLogP) (version 1.0.0)
Mutagenicity (Ames test) model (SarPy/IRFMN) (version 1.0.7)	Fish acute (LC50) toxicity model (NIC) (version 1.0.0)	BCF model (KNN/read-across) (version 1.1.0)	LogP model (ALogP) (version 1.0.0)
Mutagenicity (Ames test) model (ISS) (version 1.0.2)	Fathead minnow LC50 96h (EPA) (version 1.0.7)	kM/Half-life model (Arnot/EpiSuite) (version 1.0.0)	
Mutagenicity (Ames test) model (KNN/read-across) (version 1.0.0)	Daphnia magna LC50 48h (EPA) (version 1.0.7)	Ready biodegradability model (IRFMN) (version 1.0.9)	
Carcinogenicity model (CAESAR) (version 2.1.9)	Daphnia magna LC50 48h (DEMETRA) (version 1.0.4)	Persistence (sediment) model (IRFMN) (version 1.0.0)	
Carcinogenicity model (ISS) (version 1.0.2)	Bee acute toxicity model (KNN/IRFMN) (version 1.0.0)	Persistence (soil) model (IRFMN) (version 1.0.0)	
Carcinogenicity model (IRFMN/Antares) (version 1.0.0)		Persistence (water) model (IRFMN) (version 1.0.0)	
Carcinogenicity model (IRFMN/ISSCAN-CGX) (version 1.0.0)			
Developmental Toxicity model (CAESAR) (version 2.1.7)			

(continued)

Table 18.1 (continued)

Properties			
Toxicological	Ecotoxicological	Environmental	Physicochemical
Developmental/reproductive toxicity library (PG) (version 1.1.0)			
Estrogen receptor relative binding affinity model (IRFMN) (version 1.0.1)			
Estrogen receptor-mediated effect (IRFMN/CERAPP) (version 1.0.0)			
Skin sensitization model (CAESAR) (version 2.1.6)			
Hepatotoxicity model (IRFMN) (version 1.0.0)			

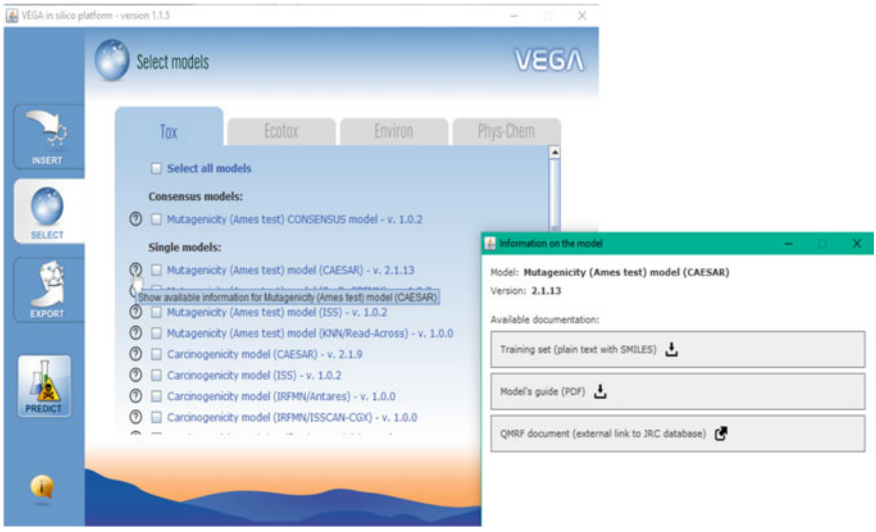

Fig. 18.1 Screenshot of VEGA. Clicking on the question mark, the user can download the pdf of the description of each model. A table with the list of the chemicals and their property value within the training and test set is also available

Using this similarity measurement, VEGA shows the six most similar compounds to the target chemical found in the training and test set together with their property values. Thus, VEGA immediately provides a simple way to make read-across.

The similarity search is also used to evaluate the reliability of the model prediction. VEGA applies a separate software for this, different from the QSAR model used to make the prediction. VEGA performs a sophisticated procedure to evaluate the reliability of the model, which refers to the applicability domain of the model. This assessment is quantitative and is provided through the applicability domain index (ADI). The ADI is a single overall measurement. ADI combined several evaluations depending on the model, since some models are quantitative (QSAR) and others are only categorical (SAR). Below we list the components of the overall assessment.

- Similarity of the most similar compounds. This parameter checks how similar are the most similar compounds.
- Chemometric check of the descriptor space. This verifies if the target compound is inside the range of the descriptors used within the model. In addition, the software checks if the molecular weight of the target compound is within the range of the molecular weights of the chemicals at the basis of the model.
- Check of the descriptor sensitivity. The software artificially modifies of up to 10% the values of the molecular descriptors and verifies if dramatic changes are recorded, indicating an unstable situation.
- Check for outliers based on specific fragments. For some models, outliers have been identified, which share a common fragment.
- Identification of the presence of rare fragments. This is done using an atom-centered fragment tool.
- Accuracy of the prediction. In this case, the software looks for the three most similar compounds and compares the predicted and experimental values for these compounds. This provides an evaluation on how well the model behaves for chemicals, which are close to the target compound.
- Concordance with the experimental values of the similar compounds and the predicted value of the target compound. This checks if there is agreement between the read-across and the QSAR prediction.

18.2.3 The ToxRead Software

In addition to the use of VEGA for read-across, as described above, there is a specific program in VEGAHUB, which is devoted to read-across: ToxRead [16]. ToxRead works by applying the same software for similarity as used in VEGA. Again, the most similar compounds are shown, as in VEGA. However, in this case the user can choose the number of the most similar compounds. Furthermore, the library of the similar compounds within ToxRead is not limited to one single model, but integrates the different collections of chemicals for the same endpoint. The real novelty of ToxRead is that, in addition to the most similar compounds, the software provides

the possible reasons of effects and the factors that may be associated with the property values, given as fragments or other parameters. The current version offers only two models: mutagenicity (Ames test) and bioconcentration factor (BCF).

The mutagenicity model includes a list of about 800 structural alerts (SA) associated with mutagenicity. These alerts derive from different collections:

- Benigni–Bossa, as from the ToxTree software (http://toxtree.sourceforge.net/);
- From the SARpy model for mutagenicity [14];
- From the CALEIDOS project, CRS4 tool (http://www.life-caleidos.eu/pages/project.php);
- From the list of fragments extracted by EB within the CALEIDOS project.

We notice that some fragments are associated with an increase in mutagenicity (toxic fragments), while other fragments are associated with a decrease in this effect. There are also fragments that are "neutral," i.e., their presence is not related to the effect (in the case of mutagenicity). ToxRead shows which SAs are present in the target compound. Figure 18.2 shows an example of the graphical output of the program. As from Fig. 18.2, the program indicates a number (up to the user) of chemicals similar to the target compound, sharing a certain SA.

Chemicals are represented by circles (red or green, depending if they are mutagenic or not), and SA are represented by triangles, with the same color code.

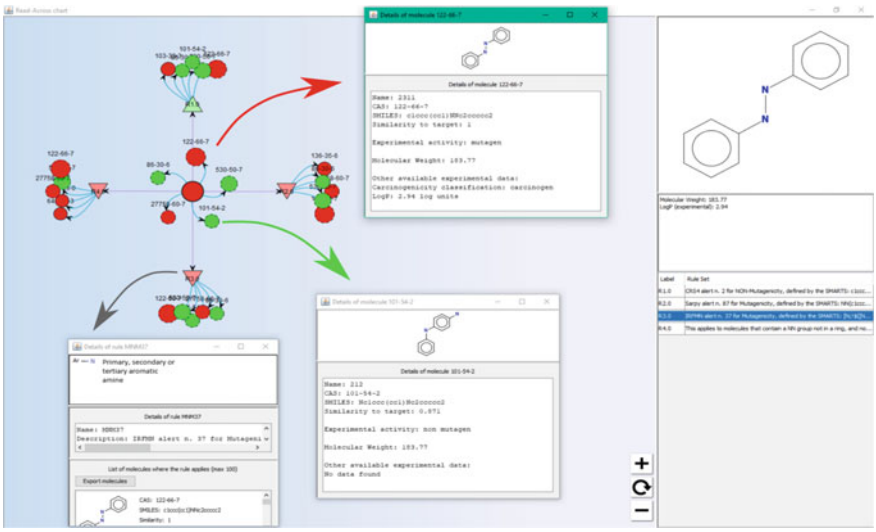

Fig. 18.2 Screenshot of ToxRead. The target chemical is in the center of the screen, surrounded by the most similar compound (circles) and the rules identified for the target chemical (triangles) with the similar compounds in which it is found. Clicking on both the similar compounds and the rules, a window with the details appear. Circles and rules are colored depending of the value: red for mutagenic substances/rules, green for the non-mutagenic ones

In the case of the BCF model, ToxRead also shows graphs with a plot reporting the BCF values versus the logP (the logarithm of the partition coefficient n-octanol/water) values of the most similar compounds. In this way, the user may apply a group approach to derive the BCF value of the target compound.

18.2.4 The ToxWeight Software

VEGAHUB integrates the results of the QSAR models and the read-across assessment, as in ToxRead. This program is called ToxWeight. It works automatically within the ToxRead model for mutagenicity. In this way, the software provides the summary of the different QSAR models, which are within VEGA, and in addition, it integrates the results obtained from the ToxRead read-across for mutagenicity.

Within VEGA, there are four separate QSAR models, and a fifth model integrates the results of these four models. The integration is done considering the call of the four models, mutagenic or not, and the reliability of each model as weight for the integration. The reliability is assessed through the ADI of each model. In the case of ToxRead, the prediction is done considering the mutagenicity values of the similar compounds and the SA.

In the present version, ToxWeight does not make a decision in the case of conflicting results coming from ToxRead and VEGA.

18.2.5 The ToxDelta Software

ToxDelta is a separate tool to further explore read-across [17]. ToxDelta identifies the maximum common substructure between two chemicals, in our case the target and the source compounds. Then, ToxDelta evaluates the parts in the molecule that are different in each chemical. These different fragments are compared with libraries of fragments, for instance mutagenicity.

18.2.6 The JANUS Software

JANUS is a program to prioritize chemical substances considering their environmental, ecotoxicological, and toxicological properties. It is intended to be used to address chemicals which are persistent, bioaccumulative and toxic (PBT), carcinogenic, mutagenic or reprotoxicants (CMR), and endocrine disruptors (ED).

The output of the program is an ordered table with substances including the reasons for concern. A score is assigned to each substance, ranging from 0 (not of concern) to 1. Uncertainty is also indicated. If uncertainty is high, the substance moves toward the 0.5 value. If multiple factors of concern exist, the substance moves toward 1.

Forty-eight different models run together in order to cover all these endpoints. As in the case of VEGA, it happens that for one endpoint more models exist. In this case, the multiple results are integrated into a unique value. The software generates microbial degradation products, which may occur in the environment, based on the program developed by the University of Minneapolis, and implemented within EnviPath (https://envipath.org/). It contains 235 separate degradation processes, identifying degradation products. JANUS evaluates the PBT, CMR, and ED properties of the parental compound and of the degradation products automatically.

The PBT and CMR properties refer to the European regulation, and thus, a chemical is considered PBT only if it is persistent, bioaccumulative, and toxic at the same time, according to the threshold values defined by REACH [11] and the regulation for the Classification, Labeling, and Packaging (CLP) [18]. Thus, for persistence, the software looks for persistence in water, sediment, and soil. A chemical is defined persistent if it is persistent in at least one of these three compartments. The evaluation for the ecotoxicity refers to aquatic toxicity only. It covers fish, daphnia, and algae, including both acute and chronic toxicity. A chemical is toxic if it is toxic for at least one of the three trophic levels. For fish acute toxicity, there are both general and specie-specific models. For bioaccumulation, the bioconcentration factor in fish is modeled.

In case of mutagenicity, the system covers the Ames test. For carcinogenicity, the system uses both regression and classifier models, in order to provide a quantitative value, related to the slope factor, both for oral and inhalation exposure. Also for reprotoxicity, there are regression and classifier models.

ED is covered considering general models for ED and specific models for estrogen and androgen receptors.

For each property addressed within the PBT, CMR, and ED, there are workflows integrating the results of multiple models, integrating both experimental (retrieved by the models) and predicted values. There are further models supporting the assessment. For instance, the workflow for BCF also verifies if the logP value is consistent with the BCF value. In case of aquatic toxicity, the water solubility is checked and compared with the toxicity value. For mutagenicity, the metabolism according to the S9 fraction is also considered.

The user can provide to the system experimental values, if available. Then, for P and B only, the workflow first refers to a series of families identified as priority pollutants. If the target compound belongs to one of these families, the score related to the family is assigned.

The components to assign the final score for the prioritization list refer to subscore related to the individual properties and endpoints assessed. As a general rule, experimental values have higher reliability than predicted values; multiple concordant values have higher reliability than single values; and the spread of the values in case of multiple values decreases the reliability.

18.3 Results

18.3.1 The VEGA Software

VEGA represents a sophisticated system to assist the user in the evaluation of the properties of chemical substances. VEGA provides three kinds of outputs, which conceptually are quite different: QSAR models, measurement of the reliability of the predictions, and read-across.

As the other systems with QSAR models, VEGA provides predictions. We try to offer models for a series of endpoints and if possible to give more than one model for the same endpoint. The models have been developed within a series of projects funded by the European Commission (EC): CAESAR, ANTARES, CALEIDOS, PROSIL, ORCHESTRA, ToxBank, EU-ToxRisk, COMBASE, VERMEER, and EDESIA. Further funds have been received by the German UBA and the German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety, within the projects PROMETHEUS, JANUS and toDIVINE, and by the Italian Ministero della Salute.

We also included some well-known models from other platforms. For instance, we have models replicated from EPISuite, or ToxTree. Some models have been developed in collaboration with US EPA, as in the case of the model for developmental toxicity. The model for reprotoxicity has been implemented based on the software developed by Procter and Gamble researchers [19], and we already mentioned the implementation of the EnviPath software within JANUS.

This is to say that we try to offer to the user good models, regardless of the theoretical assumptions and the sources, but looking at the quality of the model, in terms of quality of the data at its basis and statistical parameters. The statistical values of the models have been published in a series of papers [20–25].

VEGA provides an advanced measurement of the applicability domain (AD). It has been shown that the prediction is more reliable if we take into account the ADI. In this way, the user has a clear, automatic way to evaluate the reliability of the prediction, and the use of the ADI increases the confidence of the prediction. The software at the basis of the ADI is orthogonal compared to the software that makes the prediction.

Furthermore, VEGA can be used as a simple tool for read-across. Indeed, the user can even disregard the prediction and make his or her assessment based on the most similar compounds. This feature is probably not well-recognized by the user, but it is very simple. Also in this case, if there is more than one model for the same endpoint, we recommend use all of them. Indeed, the software for the ADI refers to the chemicals present in the different datasets at the basis of each model, and thus, each model may show different chemicals. It may also happen that different values are present in the different sets of chemicals used by the different models. In these cases, we left the original experimental values as used by the developers.

Behind the list of the most similar compounds with the experimental values, there is an additional tool for read-across which is present for the models that use SA. In this

case, VEGA, within the ADI program, indicates if there are fragments in common with the target compound, and/or fragments present only in the source compound. This is a very useful feature because it helps with reasoning about the real similarity between the source and target compounds, beyond chemical consideration. Indeed, the SA is endpoint specific, and if the target and the source compounds share the same SA, this increases the similarity, but if the source compound has a different SA, it can be responsible for the difference in the property values between the source and the target compound.

18.3.2 Other Programs for Read-Across

Above we mentioned all models available within VEGA can be used for read-across, simply by looking at similar compounds, and possibly at their SA. In addition, some models within VEGA are based on the k-nearest neighbor (kNN) algorithm. This approach is a kind of automatic read-across, which makes predictions based on similar compounds. Also in this case, we recommend looking at the results of the other models, in particular if there are models indicating SA. This because it may be that all the most similar compounds to the target chemical contain a SA which is responsible for the effect, but it is not present in the source compound.

The idea to carefully consider SA is at the basis of ToxRead. In this case, the software does the evaluation automatically. This increases the reproducibility of the results offered by the read-across tool. It has been shown that ToxRead is more reproducible and less affected by the subjectivity of the read-across evaluation, which is a major drawback of read-across [26]. This because ToxRead indicates a wide series of SA and not only those initially identified by the user.

ToxDelta may be used joined with ToxRead. Its basis is different from ToxRead, since ToxRead refers to the similarity (chemical and toxicological/environmental, through the SA and the other features it covers), while ToxDelta focuses on the dissimilarity. ToxRead looks for several similar compounds, while ToxDelta should be used for the substances to be compared.

18.3.3 The JANUS Software

JANUS further proceeds along the software PROMETHEUS for prioritization. It includes a high number of models, which will be also made available as individual tools within VEGA. The novelty of JANUS is that the prioritization covers a very large set of properties, related to a series of endpoints identified by regulation. Previous tools for prioritization, including PROMETHEUS, addressed mainly PBT. PROMETHEUS evaluated the T only through fish toxicity and did not give a numeric evaluation of each property but only a classification. Other software, like the PBT index [27], gives a PBT evaluation but not the three properties separately. Again,

other tools are thought of for the USA's regulations; therefore, the thresholds used are different compared to the European ones.

Furthermore, previous models applied one single model for each endpoint, typically. Within JANUS, 48 models run together and provide the uncertainty of the final score based on multiple results for the same endpoint.

18.4 Discussion

The VEGAHUB platform represents an example of a mature system, which goes beyond the predictions, and tries to extract from the individual tools all elements useful for the evaluation of the chemical substances. We moved our perspective from the point of view of the developers to the point of view of the user. This affected a number of our choices.

The software is freely downloadable, and this has been done in order to increase its use. The payment of a fee or a license is a barrier for some users, in particular public bodies, academia and small industries. Furthermore, industry and regulators do not like to disclose structures of the substances they are studying, and for this reason, we developed a tool that can be downloaded and used internally.

We are aware that multiple models exist or may be developed for the same endpoint. This increases the confidence of the user. Consequently, we tried to develop more than one model for the same endpoint. We also incorporated models from other systems and established collaborations with a series of developers. Indeed, VEGA is promoting links with other systems.

A peculiar feature provided by VEGA is the ADI. This is not available within other popular systems, such as ToxTree and EPISuite. Indeed, the evaluation of the AD within these systems is not possible, or is prohibitively time-consuming, since it has to be done manually on a set of parameters.

The ADI within VEGA is composed of several components, which relate to the three components of any QSAR model: chemical, computational, and property. Many models evaluate the AD considering the chemical component only and evaluate how similar the target compound is, compared to the population of substances within the training set of the model. However, this is only one part of the AD. Since at the basis of any QSAR model there are also aspects related to the algorithm and the biological/toxicological/environmental property, these aspects may surely affect the model and thus the AD. For this reason, we developed a sophisticated ADI.

Multiple factors are at the basis of the evaluation of substances, and we progressively integrate tools to assist the users facilitating this evaluation. ToxWeight is a clear example. The program automatically runs QSAR models, compares the results of the different models considering their individual reliability, and in parallel evaluates SA within a read-across perspective, which is merged with the QSAR results. This strategy has a clear reference with the Guidance on Weight-of-Evidence published by EFSA [8]. The evaluation of the reliability and consistency of the individual lines of evidence are indeed at the basis of this Guidance.

Similarly, these criteria (reliability and consistency) are adopted within JANUS.

The tools within VEGAHUB are transparent. The final decision is taken by the user, while VEGAHUB offers support representing the different lines of evidence. Thus, the user may opt for a conservative result, or prefer a realistic value.

We notice that different tools relate to different purposes. Users may be interested in the specific evaluation of a single chemical and apply a range of tools for this need. Alternatively, the user may be interested in screening, discriminating families of chemicals of concern from the safe one. In other cases, the user may need prioritization criteria. While groups of chemicals screened together may be treated together (PBT or not, CMR or not), for prioritization a list is necessary, because substances have to be ordered. Indeed, the user may have resources to separately assess a limited number of compounds, and it is necessary to provide a way to sort this number out of the group of CMR chemicals, for instance. For this reason, within JANUS we developed models that provide continuous values for carcinogenicity and reprotoxicity. For instance, carcinogenic compounds can be ordered according to their potency value.

18.5 Conclusion and Future Directions

Within VEGAHUB we provide a series of tools with the aim to better evaluate individual chemicals, or families of substances. The individual tools, QSAR, read-across, and grouping, are related to the methodological aspects. We tried to move toward results of the models, which provide details, interpretations, related values, occurrence of SA, graphics with associations and trend, etc. beyond the technicalities of the individual models. Within VEGAHUB multiple tools are connected and used simultaneously in order to assist the user.

We have already planned extensions of VEGAHUB. A major improvement will be achieved by joining VEGAHUB with MERLIN-Expo (https://merlin-expo.eu/), within the LIFE VERMEER project (https://www.life-vermeer.eu/) funded by the EC. In this way, the user will have combined tools for the hazard and for the exposure, and thus, the evaluation of risk assessment will be facilitated. Furthermore, VERMEER will identify safer substitutes of risky chemicals, and in this, it will be aligned with the principles of the REACH legislation.

The toDIVINE project funded by the German UBA will further integrate read-across (based on substances registered for REACH) with QSAR predictions.

The LIFE CONCERT REACH project, funded by the EC, will establish a network between VEGAHUB, the Danish QSAR database, and AMBIT (http://cefic-lri.org/toolbox/ambit/), to fully exploit data registered within REACH. All endpoints will be scrutinized, in an effort to develop QSAR models for the tens of endpoints addressed within REACH.

The Optitox project, funded by EFSA, will derive QSAR models from the EFSA OpenFoodTox database by leveraging information about the toxicokinetics.

All these initiatives will dramatically improve VEGAHUB and its capability to provide assistance to the evaluators of chemical substances, moving the field toward safer chemicals.

Acknowledgements We acknowledge the LIFE + Program for funding, through the LIFE VERMEER project (LIFE16/ENV/IT 000167).

References

1. Devillers J (2006) Application of QSARs in aquatic toxicology. In: Ekins S (ed) Computational toxicology. https://doi.org/10.1002/9780470145890.ch23
2. Honma M et al (2018) Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcome of the Ames/QSAR International Challenge Project. Mutagenesis (in press)
3. Benfenati E et al (2018) Computational toxicology and REACH. In: Ekins S (ed) Computational toxicology: risk assessment for chemicals. Wiley, NJ, USA
4. Patlewicz G et al (2017) Navigating through the minefield of read-across tools: a review of in silico tools for grouping. Comput Toxicol 3:1–18
5. European Chemicals Agency (2008) Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals
6. European Chemicals Agency (2017) Read-Across Assessment Framework (RAAF)
7. European Chemicals Agency (2016) Preparation of an inventory of substances suspected to meet REACH Annex III criteria. Technical documentation
8. EFSA Scientific Committee, Hardy A et al (2017) Scientific opinion on the guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 15(8):4971, 69 pp
9. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products (recast). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02009R1223-20160812&from=EN
10. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32010L0063
11. Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).
12. European Chemicals Agency (2016) How to use and reports (Q)SARs. Version 3.1
13. Ferrari T et al (2013) Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. SAR QSAR Environ Res 24:365–383
14. Toropova AP et al (2013) CORAL: Monte Carlo method as a tool for the prediction of the bioconcentration factor of industrial pollutants. Mol Inform 32:145–154
15. Floris M et al (2014) A generalizable definition of chemical similarity for read-across. J Cheminformatics 6:39
16. Gini G et al (2014) ToxRead: A tool to assist in read across and its use to assess mutagenicity of chemicals. SAR QSAR Environ Res 25:999–1011
17. Golbamaki A et al (2017) ToxDelta: a new program to assess how dissimilarity affects the effect of chemical substances. Drug Des 2017:6
18. Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures
19. Wu S et al (2013) A framework for identifying chemicals with structural features associated with potential to act as developmental or reproductive toxicants. Chem Res Toxicol 26:1840–1861
20. Como F et al (2017) Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model. Chemosphere 166:438–444

21. Golbamaki Bakhtyari A et al (2016) New clues on carcinogenicity-related substructures derived from mining two large datasets of chemical compounds. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 34:97–113
22. Manganaro A et al (2016) Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. Chemosphere 144:1624–1630
23. Gissi A et al (2015) Evaluation and comparison of benchmark QSAR models to predict a relevant REACH endpoint: the bioconcentration factor (BCF). Environ Res 137:398–409
24. Cassano A et al (2014) Evaluation of QSAR models for the prediction of Ames genotoxicity: a retrospective exercise on the chemical substances registered under the EU REACH regulation. J Environ Sci Health Part C: Environ Carcinogenesis Ecotox Rev 32:273–298
25. Golbamaki Bakhtyari N et al (2013) Comparison of in silico models for prediction of mutagenicity. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 31:45–66
26. Benfenati E et al (2016) Results of a round-robin exercise on read-across. SAR QSAR Environ Res 27:371–384
27. Papa E et al (2010) QSPR as a support for the EU REACH regulation and rational design of environmentally safer chemicals: PBT identification from molecular structure. Green Chem 12:836–843

Emilio Benfenati is head of the Laboratory of Environmental Chemistry and Toxicology at the Mario Negri Institute, Milan, Italy, since 1997. Previously he has been researcher at Stanford University, California, USA (in that period he also had a collaborative research at the Berkeley University, California, USA). He coordinates/coordinated 20 European projects and participated to about 50 projects in total, many of them on toxicity and environmental modeling. His research activities include: toxicity and environmental modelling, molecular descriptors, QSARs, toxicity prediction, characterization and assessment of contaminants, and risk assessment. He is the main developer of VEGA, ToxRead, and other in silico platforms. He is author or co-author of 350 papers in international journals and edited a few books. He organized some conferences, including SETAC 2011 and QSAR 2014.

Alessandra Roncaglioni is currently Head of the Research Unit on Computational Toxicology at the Mario Negri Institute for Pharmacological Research, Milan, Italy. She graduated in environmental science and got a Ph.D. in Life and Biomolecular Science. Her research interests are related to the use of computational chemistry methods applied to environmental and health issues, in particular in silico approaches (such as QSAR) and their use in the regulatory framework. She is co-author of nearly 50 papers in peer-reviewed journal and about 30 communications in international conferences.

Anna Lombardo with a degree in Environmental Science, has worked at the Istituto di Ricerche Farmacologiche Mario Negri IRFMN since March 2008. She specialized in biomedical research in 2011. She has worked in computational toxicology and in particular on the development and managing of datasets, developing and use of QSAR model especially for ecotoxicological and environmental endpoints (i.e., aquatic toxicity, bioconcentration, persistence). She has worked on the development of Integrated Testing Strategies, for aquatic toxicity and bioaccumulation, to satisfy REACH requirements. She is author of more than 30 papers published on international journals, posters, and public presentations.

Alberto Manganaro has a degree in computer science, with expertise in the theory of Molecular Descriptors, QSAR/QSPR model development and the software implementation of chemoinformatics tools. He has been part of the development group of the DRAGON software, and he is currently working on the development and update of the VEGA and ToxRead platforms. In 2012, he founded the consultancy company Kode srl.

Chapter 19
OpenTox Principles and Best Practices for Trusted Reproducible In Silico Methods Supporting Research and Regulatory Applications in Toxicological Science

Barry Hardy, Daniel Bachler, Joh Dokler, Thomas Exner, Connor Hardy, Weida Tong, Daniel Burgwinkel and Richard Bergström

Abstract Our aim in this work and initiative is to establish a practice and guidance for tracking and reporting modern in silico data analyses in a reproducible manner. The recommended reproducible principle supports the concept that data analyses, and more generally, scientific claims and regulatory evidence, are published with their raw data and software code so that others may verify the findings and build upon them. We discuss here how we are demonstrating implementations of trusted reproducible in silico evidence workflows and are enhancing their acceptance with an

B. Hardy (✉) · D. Bachler · J. Dokler · T. Exner · C. Hardy
Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, 4057 Basel, Switzerland
e-mail: Barry.Hardy@edelweissconnect.com

D. Bachler
e-mail: Daniel.Bachler@edelweissconnect.com

J. Dokler
e-mail: Joh.Dokler@edelweissconnect.com

T. Exner
e-mail: Thomas.Exner@edelweissconnect.com

C. Hardy
e-mail: Connor.Hardy@edelweissconnect.com

W. Tong
National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA
e-mail: Weida.Tong@fda.hhs.gov

D. Burgwinkel
Guardtime AS, Basel, Switzerland
e-mail: daniel.burgwinkel@guardtime.com

R. Bergström
Bergstroem Consulting GmbH, Zug, Switzerland
e-mail: richard.bergstrom.zug@gmail.com

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0_19

open knowledge community approach supported within OpenTox and OpenRiskNet. The general principle discussed in this article can be applied in regulatory settings.

Keywords In silico · Toxicology · Regulatory · Reproducibility · Data · Trust and provenance · Blockchain

Abbreviations

AI	Artificial intelligence
AOP	Adverse outcome pathway
API	Application programming interface
DILI	Drug-induced liver injury
EMA	European Medicines Agency
FAIR	Findable, accessible, interoperable, and reusable
FDA	Food and Drug Administration
GDPR	General Data Protection Regulation
GPR	Global Patient Registration
IATA	Integrated Approach to Testing and Assessment
IPFS	InterPlanetary File System
IST	In Silico toxicology
ITS	Integrated testing strategy
IUCLID	International Uniform ChemicaL Information Database
NDA	New Drug Application
OECD	Organisation for Economic Co-operation and Development
OHT	OECD Harmonised Template
OS	Operating System
QMRF	QSAR Model Reporting Format
QSAR	Quantitative Structure Activity Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
URL	Uniform Resource Locator
WoE	Weight of Evidence
XML	eXtensible Markup Language

19.1 Introduction

Reproducibility in science has raised significant concerns in recent years. For example, a recent study of 1500 scientists highlighted key issues with the finding: "*More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments.*" http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970.

In data science and computational modeling, this problem has reached new magnitudes of concern, due to the diversity of approaches, software tools, and hardware architectures involved. Many methods are emerging in the research field of alternative testing methods supporting safety assessment, but industry and regulators are finding the establishment of reproducible application, evaluation, and guidance challenging, and current regulatory acceptance is as a result moving slowly.

Our aim in this work and initiative is to establish a practice and guidance for tracking and reporting modern in silico data analyses in a reproducible manner. Reproducible research supports the concept that data analyses, and more generally, scientific claims, are published with their raw data and software code so that others may verify the findings and build upon them. Reproducibility allows people to focus on the actual content of a data analysis, rather than on superficial details reported in a written methods summary. In addition, reproducibility makes an analysis more useful to others because the data and code used to conduct the analysis are available. This paper focuses on literate statistical analysis tools which allow one to publish data analyses in a single repository that allows others to easily execute the same analysis to obtain the same results. Although many of the principles described here apply broadly to scientific research and associated data science and principles, we will focus on the domains of toxicology and risk assessment where data is generated and used and models are developed primarily for the purpose of ensuring the safe level of use of a chemical or drug in a human or the environment.

The work proposed here aims to exploit background developments related to the OpenTox community (http://www.opentox.net/), and its associated current infrastructure project OpenRiskNet (https://openrisknet.org/), so that information generated and processed by in silico methods are more suitable for purpose for industrial use and regulatory acceptance including:

(a) Establishment of application programming interfaces (APIs) for scientific data within the context of an open infrastructure providing reliable quality-controlled access to harmonized scientific data;
(b) Development of best practice in silico workflows, processing the data with the principle of reproducibility as established by the context of use;
(c) Engagement of the scientific community to develop and contribute best practice in silico workflows;
(d) Establishment within the OpenTox community of guidance to best practices for in silico workflows in predictive toxicology.

Our objective is to develop guidance, methods, and best practices supporting reproducible in silico computational toxicology and safety assessment. Implementations within OpenTox and OpenRiskNet will be established against use cases involving model building, validation and integrated testing. The approach will be extended to include additional contributions from the scientific and regulatory communities for elaboration and consensus building. This approach should support the independent verification of resources used in producing results as toxicological evidence.

Currently, basic research and outputs exist, but the approaches for regulatory use and acceptance are missing from practice. We need to develop reliable access to data,

specify the metadata needed, implement interoperability layers and workflows, and obtain community contributions and acceptance. Achievement of such goals will increase the reliability of predictions, increase the ability to reproduce a prediction and determine reasons for deviations, and support the independent verification of resources used.

19.2 Proposed Principle of Reproducibility for In Silico Modeling and Workflows and Implementation in Practice

We state the principle of reproducibility here as:

> The Principle of Reproducibility states that close agreement in scientific results can be obtained when a sufficiently well-described protocol is competently executed. The protocol can be experimental (in vitro or in vivo) or computational (in silico) or a combination of such protocols.

The main concern of the current paper addresses the achievement in practice of reproducible ***computational (in silico) protocols***. Bioinformatics and data science in general are fraught with challenge that hamper the achievement of the principle of reproducibility in practice. These include correctly identifying the dataset one is working with, identifying the transformations and manipulations that may have been done to it, validating these, as well as similarly recording new transformations that one applies. The ease of making, modifying, and distributing copies of digital data leads to the proliferation of multiple versions of datasets, which may have ambiguous origin and meaning. At the same time, technologies such as formal ontologies and blockchain provide opportunities to address this problem. At their core, the blockchain methods build on calculating hashes (checksums) of the data and the software used for data processing (Sect. 19.6 gives a more in-depth outline of this approach).

We suggest a solution structured around *verification and reproducibility annotations*, implemented according to the following practices (see Fig. 19.1):

- Formally identifying datasets, and versions of these, with a suitable hash function;
- Similarly identifying tools (computational steps that produce new or derived versions of data) and versions of these;
- Generating an "audit trail" that describes all transformations that have been applied to data that is retrieved, for example, through APIs, by referring to hashes of specific versions of necessary tools, inputs, and intermediate data. This would at a minimum be in a way that allows such transformations to be verified and ideally in a way that allows them to be reproduced. When verification fails, the supporting tools might suggest how to update or repair the workflow;
- Making it trivially easy to integrate or introduce new transformations (such as simple R or Python scripts) into this kind of auditing system, such as to not interfere with existing work habits;

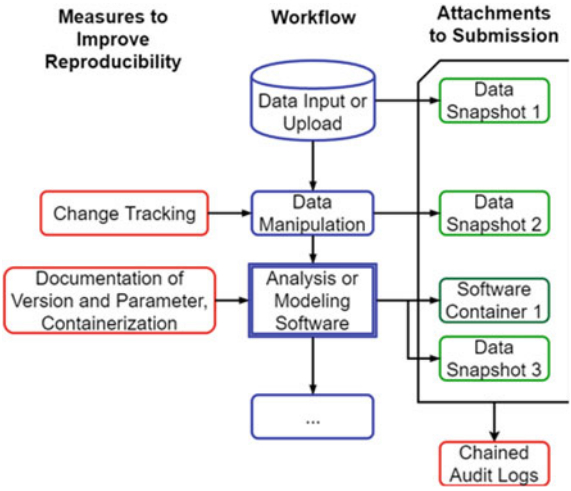

Fig. 19.1 Measures to improve reproducibility of in silico workflows

- At the same time developing and providing advanced methods for the full integration and documentation of the workflows with repositories like GitHub and finally by containerization of data including raw and intermediate data as well as the software tools for producing a complete snapshot of the knowledge generation process.

If these practices can be successfully realized (and the success of systems like the blockchain, GitHub, distributed file systems like IPFS, and containerization approaches like docker suggests that they can), then these reproducibility annotations may also form the basis for future applications that may, for example, actively generate and regenerate the latest versions of data, results, and conclusions using the latest tools and input data sources, according to well-defined recipes and in this way protocoling the scientific advances. Reproducibility annotations could also be a cornerstone of significant efficiency improvements in organizations that manipulate biological data.

19.3 Example 1: QSAR Model Building and Validation

The increasing complexity of molecular descriptors and machine learning algorithms presents a challenge for the reproducibility of QSAR models. The use of nonlinear algorithms is increasing, a practice that is assisted by the advancements in fields like neural networks and support vector machines among others. Not only does this require the explicit documentation of optimization methods, but also algorithm parameters, including random number seeds, must be noted down. Furthermore, the construction of metamodels, such as utilizing a bootstrap aggregation (bagging) and boosting protocols, adds yet another layer of complexity to the model building and

documentation making it harder to reproduce. Unless the authors contribute a saved copy of their model, it is almost impossible to reproduce the QSAR model based on a text description alone. A standard format (the QSAR model reporting format (QMRF)) has been suggested to report on the structure of a QSAR model. It takes the form of a harmonized template for summarizing and reporting key information on QSAR models, including the results of any validation studies. The information is structured according to the OECD QSAR validation principles. At the time of writing, the QMRF database published by the European Commission's Joint Research Centre (http://qsardb.jrc.it/qmrf/) includes over one hundred such reports.

Thousands of chemical descriptors have been documented [1]. While many descriptors are chemically intuitive, for instance "number of hydrogen-bond acceptors," their algorithmic implementation is often not. Many software packages have different interpretations of the same SMILES representation of a chemical structure resulting in a different count of what is to be considered "hydrogen-bond acceptor." This problem may often be aggravated by the poor description of the structure cleaning steps (such as whether the modelers renormalized aromaticity, re-optimized 3D structures, or neutralized charges). As a minimum requirement for reproducibility, the name of the software tool and its version have to be documented. A more practical approach is to internally store (and possibly publish) an exact copy of the software program and the preprocessing workflows that were used to calculate descriptors. Such a step is more easily attained for open software than proprietary programs due to the availability of the source code and the inviting license model.

The importance of validating QSAR models for regulatory acceptance cannot be overstated. A validated model is one that is able to produce consistent results when tested on external validation set(s) not involved in its training. For external researchers and regulators to validate a QSAR model, they must be able to reproduce it with sufficient accuracy. Many online platforms for QSAR model building and validation have been developed: OpenTox, OCHEM [2], Chembench [3], and AMBIT [4], among others. Many of these platforms can be managed through application programming interfaces (APIs) allowing the scripting control of complicated QSAR model building workflows by passing the parameters needed for machine learning algorithms, descriptor packages, variables selection, and preprocessing steps. Modelers can therefore document an entire workflow using such a script, while the online platforms store the underlying binaries and versions. OCHEM also offers its XML format for documenting such settings in a reproducible manner.

OpenTox proposed a best practice for building validated in silico QSAR models accompanied by specifications for APIs and supported by open standards and ontology (see Fig. 19.2) for harmonized knowledge descriptions and communications between components for data, algorithms, models, and validation. We propose to update and upgrade these specifications and best practices through making a proposal with accompanying case study examples to the community to include the goals of reproducibility, trust, and provenance discussed in this paper. Such a QSAR may also be deployed in an ITS as described in the example in the next section.

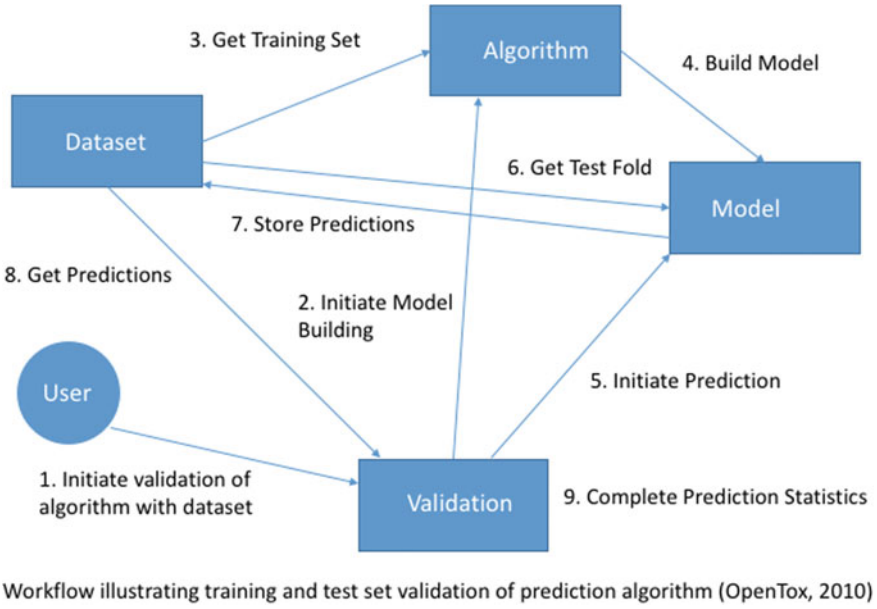

Fig. 19.2 OpenTox best practice for building a validated QSAR model

19.4 Example 2: Integrated Testing Strategies

Integrated testing strategies (ITSs) can be described as *combinations of test batteries covering relevant mechanistic steps and organized in a logical, hypothesis-driven decision scheme, which is required to make efficient use of generated data and to gain a comprehensive information basis for making decisions regarding hazard or risk* [5, 6]. They can be seen also as an algorithm to combine and establish links between (different) test result(s) and, potentially, non-test information (existing data, in silico extrapolations from existing data or modeling*)* and to come up with a combined test result. As such, ITS try to overcome the problems associated with the standard test batteries which are in most cases presented as a sequence of tests without formal integration of results. Consequently, the use of these standard batteries leads to lack of guidance on how to perform consistently and lack of transparent inference about the information target. ITS on the other hand are built upon the following conceptual requirements: (i) *transparency and consistency*, to ensure comprehensiveness and as a result credibility and acceptance; (ii) *rationality*, to ensure that all relevant information is fully exploited and optimally used; (iii) *flexibility* to ensure hypothesis-driven decisions and the possibility of adjustment of the initial hypothesis whenever new information is obtained or generated. Furthermore, they bundle different and possibly contradictory information and the respective uncertainties considered in a weight of evidence (WoE) approach. Additionally, in case of data gaps, the ITS would propose the most appropriate method to acquire the missing information.

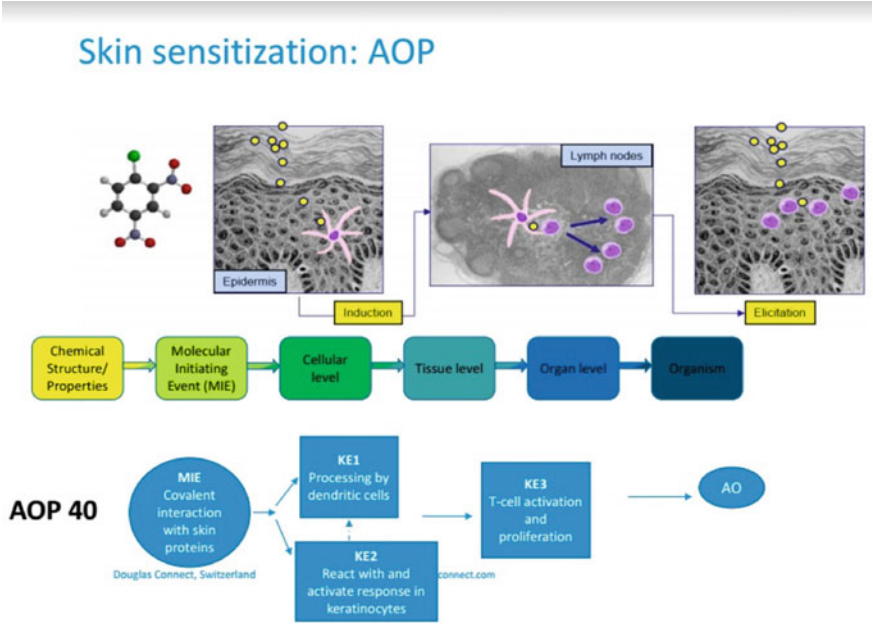

Fig. 19.3 Adverse outcome pathway for skin sensitization involving four key events (KEs)

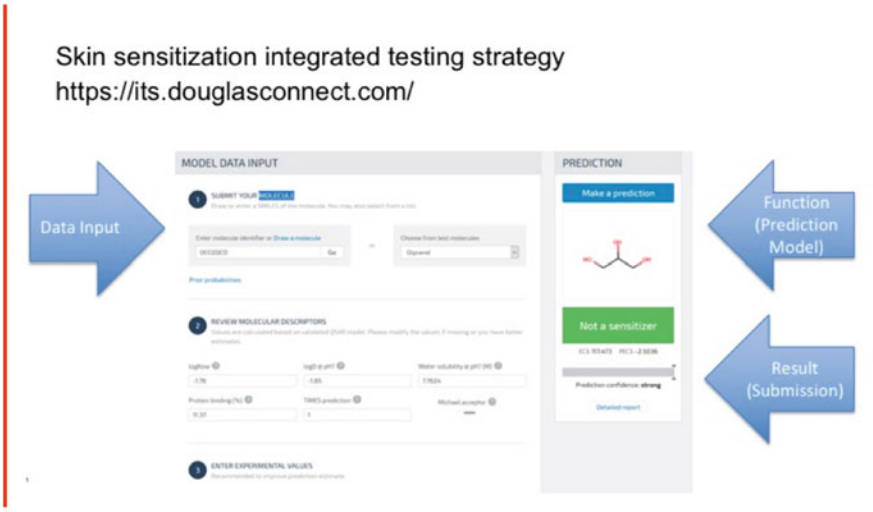

Fig. 19.4 Skin sensitization integrating testing application for skin sensitization

Reproducibility of a testing strategy would therefore encompass the reproducibility of the experimental data as well as the algorithmic or in silico model. Relying on a documented test protocol to conduct the experiment and keeping track of any deviations from the original protocol is a key. A knowledge management system where test protocols are tracked and versioned is a good practice. OECD publishes harmonized templates that describe the data formats for reporting information that is used for the risk assessment of chemicals. Its applications involve, not only "studies done on chemicals to determine their properties or effects on human health and the environment, but also for storing data on use and exposure." The harmonized templates are aimed at developers of database systems, as they prescribe the formats by which information can be entered into and maintained in a database. The test study summary information can be exchanged between governments, regulators, and industry electronically using these templates. The templates can also be used to report summary test results for different products. REACH registrants use the OHTs for submitting their results to the European Chemicals Agency (ECHA) using IUCLID. On the other hand, the theoretical model behind the ITS can vary from simple decision trees on the order by which risk assessors should perform their experiments [7] to more complicated machine learning algorithms [8] or Bayesian networks [9].

ITS skin sensitization example: The biology of skin sensitization is described by an adverse outcome pathway (AOP) involving a sequence of key events representing potential biological activity responses to chemical exposure of the skin (see Fig. 19.3). Validated assays against the key events may be used in combination with in silico methods for an ITS purpose. Within the OECD framework, a guideline exists for Integrated Approaches to Testing and Assessment (IATAs) which consist of an ITS where different strategies or algorithms may be used, as described by a "defined approach." We have recently developed an ITS application for skin sensitization (see Fig. 19.4) [10] reproducing the IATA and defined approach developed at Procter and Gamble and involving Bayesian networks [9]. The use of Bayesian networks allows tolerating missing information and conveys the probabilistic hypothesis of skin sensitization based on accumulative evidence from data. In building the ITS, open-source tools were favoured by Douglas Connect (now renamed Edelweiss Connect (EwC)). A public open-access reference application is made available for the scientific community to evaluate at https://its.douglasconnect.com.

19.5 Best Practices in Data Management: FAIR Principles, Data Completeness, Interoperability and Ontology

Comprehensive documentation of the scientific procedures, experimental as well as computational, is of uttermost importance to achieve reproducibility. Besides guidelines for running the experiments including good laboratory practice and cell culture practice guides (which are out of scope of this current paper), detailed test

method and protocol description, optimally also in a computer-readable version, data quality assurance measures and data management are at the heart of creditable scientific practice. This fact is now generally acknowledged resulting in the FAIR data principles (findable, accessible, interoperable, and reusable) endorsed by almost all major funding agencies and many high-impact journals as well as the Guidance on Good Data and Record Management Practices by the WHO [11]. Well-prepared data are valuable resources that can be used and reused with high confidence. In this way, data sharing facilitates new scientific inquiry, avoids duplication of experiments and data collection, provides rich real-life resources for method validation as well as education and training and, most importantly for the topic here, allows for the complete comparison of all steps from raw data to results when trying to reproduce the study. To allow for data usage without the need to refer to external sources like publications, final reports, working papers, or laboratory books, all important information should be included in the data submission. Good data documentation includes information on:

- author and affiliation contact, relevant dates, reference to test method and protocols as well as publications and reports;
- the context of data collection: project history, aim, objectives, and hypotheses;
- data collection methods, dataset structure of data files, study cases, relationships between files;
- data validation, checking, cleaning, and quality assurance procedures carried out;
- changes made to data over time since their original creation and identification of different versions of data files;
- information on access and use conditions or data confidentiality;
- names, labels, and descriptions for variables, records, and their values;
- explanation or definition of codes and classification schemes used;
- codes of, and reasons for, missing values;
- derived data created after collection, with code, algorithm, or command file.

It is clear that data completeness, quality and reproducibility is mainly influenced by procedures adopted during data collection and documentation of how data are collected provides evidence of such quality. The digitization and entering of data, the documentation of data manipulation, the processing, and most importantly complete in silico approaches like read-across and QSAR need to follow high-quality data standards. Errors during input and unintentional modification or disruption can be avoided or at least be minimized by standardized and consistent procedures with clear instructions. The procedures described below will provide guidance to follow these standards, provide tools for on-the-fly verification, as well as minimize human intervention by:

- setting up validation rules for data entry software as well as providing validation tools to be executed after each data modification;
- using automatic protocoling of data manipulations and software usage including hardware and software setup, needed data transformation procedures and program-specific run time parameters;

- supporting the usage of controlled vocabularies taken from and referencing standard ontologies, code lists, and choice lists to minimize manual data entry;
- eliminating manual file format transformations by data harmonization and enhancing the interoperability of the integrated analysis software.

19.6 Proposed Technology Solutions for Data Processing

In the domain of software engineering, the intersecting problems of versioning and collaboration were solved by the introduction of Distributed Version Control Systems (DVCS), most notably among them git [12]. These systems work by recording all changes to all source code files of a project over the entire history of the project, starting with an empty folder and arriving at the latest development version. When a changeset is created by the user, the DVCS compares the status of all files with the last recorded changeset and creates a list of differences across all files. This list of differences is then combined with metadata about the changeset (timestamp, author information, the hash of the previous changeset, etc.). The resulting data is hashed, meaning that a unique numeric identifier of fixed length is calculated and added to the inventory of changesets.

We envision a similar system for data and data transformations, to enable verification and reproduction of data and workflows operating on them. This would consist of a standardized description of metadata about the data and data transformation steps, linked together as a merkle tree [13] similar to the way git or bitcoin operates.

In a simple case, a downloaded raw data file would be hashed, metadata added (e.g., the URL the dataset was retrieved from, a timestamp of the retrieval), and the result recorded as a first step alongside the downloaded data. An analysis that would then be run on top of this as an R script would similarly be recorded into the system, but in this case, the metadata would include the path to the R script in a git repository and the hash of the version used when running the analysis. Crucially, the metadata of this second step would include the hash of the first step to link these two steps together.

We acknowledge that risk assessment is an environment where the use of such systems is not yet widespread, data manipulation is often done offline, and thus, no full chain of linked processing description steps exist from the final result to the raw experimental data (though such a full chain would certainly be possible and desirable). As such, we anticipate a verification system concentrating on the "tail end" of the analysis initially (e.g., starting with a downloaded dataset about which nothing more is known than the URL where it was downloaded from and then doing one or two analysis steps to reach a conclusion). As the usefulness of such an approach is recognized in the wider community, we hope that data managers either responsible for public or in-house data would themselves, in collaboration with the experimental scientists, start to record such information based on the approaches

developed for the verification system. Summary datasets prepared like this would contain recordings of all data transformations that were used to derive this data from the raw experimental data allowing for an in-depth validation and reproduction of the procedure during the review process of a regulatory submission. The technology primitives for such systems are well established and are in use in git, bitcoin and other blockchains, and IPFS (the InterPlanetary File System) and can be used to express nonlinear relationships between workflow steps. Another often volatile component of any data-based workflow is the system utilities and libraries installed as part of the operating system (OS). With the recent development of NixOS Linux [14], even these parts could be recorded into the merkle tree, because NixOS provides capability to hash all the tools available in an OS at a given point in time and to restore this exact state at a later point in time.

19.7 Locating the Source of Irreproducibility: Sub-tasks and Intermediate Datasets

The system we envision would contain the necessary information not just for a human to audit the data steps and be able to reference the precise files and software versions to reproduce a data analysis or transformation—it would indeed go further and describe the necessary execution steps in enough machine readable detail to reproduce the workflow automatically (probably by extending an existing workflow tool like NextFlow or similar). In such a scenario, the analysis could be replayed by the workflow tool one to one. A successful demonstration of such an approach is the literate programming environment R Markdown, which can regenerate an entire publication including all data downloads from original URLs, through all intermediate data cleaning and processing steps until the final generation of the report pdf. The default way of comparing any intermediate step would be to compare the hashes—this would allow a user reproducing the workflow to see if there is at any point a divergence between the original publication or analysis and their reproduction but would not give them more information. To make the system more powerful, we want to follow the example of git and make the actual diffing algorithm interchangeable. In this way, a data format aware diffing could be used to compare files with a deeper understanding of its content—for example, when two csv files (the first from the original study submitted with the report and the second from the validation study) differ, the divergence could be highlighted in the file instead of giving just a yes/no answer to the question “Are these files binary identical?”. Such a flexible diffing would allow users that reproduce the data to define a threshold of similarity, up until which they would still regard the result as being reproduced (e.g., if there is a newer original dataset that is used to run the same analysis).

19.8 Use of Software Containers

Web services are increasingly deployed using software container technologies like docker; they are already versioned with a cryptographic hash (since this is how docker containers are layered internally). This information should then be included in the audit log when a query is performed against such a system. If the container is publically available, a researcher reproducing a workflow could then recreate an analysis even if the Web service itself is no longer running or a newer version produces different results. Besides all used data (raw data of experiments specifically generated for the submission and all public sources), these containers could also become part of a regulatory submission guaranteeing that the analysis will always be exactly reproducible. Software needing a license to be run could be offered in a specific version only allowing this one calculation to be run. Additionally, data produced in intermediate steps (data snapshots) should also be stored in data containers. This would have the advantage that data integrity could be validated from snapshot to snapshot; e.g., copy errors could be identified by comparing the before and after state of the data, as long as such manual manipulations are still needed due to missing interoperability of the tools.

19.9 Regulatory Acceptance Practices

The regulatory acceptance for a test method is represented by its formal acceptance by regulatory authorities indicating that the test method can be used to provide information to meet a specific regulatory requirement. This includes a formal validation (i.e., reliability and relevance assessment of the method considering its reproducibility, transferability, predictive capacity, applicability domain, and performance standards) [15] and adoption by the International Organizations (i.e., OECD and EC) before implementation into specific regulations and related guidelines [16, 17]. Similarly, for the in silico methods, as they are playing an increasing role in predicting properties for hazard and risk assessment, the acceptance of computational approaches should be based on standards and criteria of reliability and relevance prior to be applied within a specific regulatory context. As an example, EU REACH Regulation explicitly includes the need to use QSAR models to reduce the extent of experimental testing, emphasizing the principle that information generated by QSARs may be used to indicate the presence or absence of a certain dangerous property instead of experimental data, provided that the following conditions are met: The results are derived from a QSAR model whose scientific validity has been established, the substance falls within the applicability domain of the QSAR model, the results are adequate for the purpose of classification and labelling and/or risk assessment, and adequate and reliable documentation of the applied method is provided [18]. Moreover, in order to facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with the following information: a defined endpoint, an unambiguous

algorithm, a defined domain of applicability, appropriate measures of goodness-of-fit, robustness and predictivity and a mechanistic interpretation [19].

Another strategy for including the in silico methods into decision making and accelerating the regulatory acceptance for toxicity testing is represented by the Integrated Approaches to Testing and Assessment (IATAs), which are currently proposed as practical solutions to integrate such alternative methods, with ITS being an example of such an approach. IATAs provide a means for combining the data from different methods, considering all available relevant information about a substance in a weight-of-evidence assessment, to inform regulatory hazard or risk decisions or the need for additional tests and rely on non-animal approaches to determine chemical hazard or risk [20]. In parallel, the development of the adverse outcome pathway (AOP) framework, which provides information on the adverse outcome of regulatory concern, offers the biological context to facilitate development of IATAs for regulatory decision making [21]. For example, in silico models like QSAR that are designed to predict key events in different pathways should be useful sources of information in IATA, whereas QSARs that are directly predictive of the adverse effect may be useful for increasing confidence in weight-of-evidence arguments [16]. However, a critical question for the risk assessor when applying a non-testing method for regulatory purposes is regarding the reproducibility of the model, thus on the reliability of the prediction.

We have selected OpenTox which is the leading global open knowledge community in new in silico toxicology methods as the ideal location for discussion of case studies and the above best practices of potential to benefit emerging regulatory frameworks. We aim to demonstrate with a broad community input on how new technologies are essential for regulatory science, more specifically by highlighting reproducible in silico practice via OpenTox, with a focus theme on the issues of reproducibility, predictive modeling, and other related enabling topics such as semantic interoperability of contributing resources. The process is not only about building predictive models, but also about placing observations on how predictive uses are constantly changing within a community evaluation context. In addition to building an application with a set of principles, other concerns shared by developers and practitioners are the implementation of best practices based on quality, reliability, robustness, interoperability, reproducibility, harmonization, completeness, openness, and confidence. These principles and their associated practices require both a dialogue and a consensus such as best practice protocols, context of use and fit for purpose issues. Another current related initiative is the in silico toxicology (IST) protocol consortium, organized by Glenn Myatt, Founder of Leadscope. This international consortium includes regulators, government agencies, industry, academics, model developers, and consultants across many different sectors, formed initially with the intention of creating the overall strategy for an in silico protocol development. Working subgroups will develop individual in silico toxicology protocols for major toxicological endpoints, including genetic toxicity, carcinogenicity, acute toxicity, reproductive toxicity, and developmental toxicity, and we plan to interact with this important initiative.

19.10 Enhancing Workflow Solutions Including Trust and Data Provenance Using Blockchain Technology Applied to Healthcare Applications

In addition to the above best practices in reproducible workflows, we are currently exploring the use of blockchain technology to add an independent verification of trust and data provenance. We are comparing the use of the major blockchain platforms in their utility for trust and provenance goals. In particular, we have commenced a collaboration with Guardtime where the technology has already been successfully used to secure electronic health records in Estonia. The goal is to take the above QSAR and ITS models and workflows and run them on the enterprise blockchain to provide signatures of execution history so as to secure the evidence that would be generated and submitted to a regulator. The case of preclinical evidence provides a suitable starting case study framework for healthcare applications, which we plan to subsequently extend to clinical data.

The vision of our work is to establish a trusted healthcare blockchain ecosystem supporting transparent and reliable data exchange, as well as provenance and completeness of data sharing, between all stakeholders so as to maximize patient benefits and outcomes. Our case study work draws upon the above preclinical flows executed by the Edelweiss Connect team in collaboration with the fully functioning Estonian Ministry of Health data sharing platform from Guardtime which harnesses blockchain to allow patients full, secure control of their data assets; the Guardtime technical team has built these systems for the Estonian Health Ministry. The provenance of data and management of consent is critically important. To deliver trust in the results (of machine learning and AI), there must be trust in the underlying data (authenticity and integrity). We see Blockchain as a trust layer to take us to the next level of knowledge about human biology. Blockchain can also serve as a trust layer in the sense of providing assurance of "how things are done." For example, did the supplier of an active substance produce it according to the set standards, or did the distributors handle the product (e.g., with temperature sensitivity) with sufficient care? Clinical trials may be sped up if source data verification can be done automatically.

In addition to providing a trusted interoperability layer for data sharing, we also believe this framework will additionally enable applications supporting goals such as regulatory acceptance of new methods and evidence, outcome-based contracts and moving toward patient-centric value-based healthcare.

In the healthcare and pharma sectors, regulatory oversight and patient privacy issues provide strong regulatory and legal requirements on the use of new Blockchain technology and related operation models such as decentralized networks that have to be fulfilled 100%. The implementation of GDPR in Europe has raised the awareness of data privacy, but operators are struggling to find workable models for "the right to be forgotten" and dynamic consent (where it is not only yes/no and when preferences may change over time). As healthcare regulation, including Global Patient Registration (GPR), implementation is often a national or even regional competency,

and our approach to legal review and framework will embrace the need for alignment with multiple legal contexts. We plan to interact closely with projects running under the Innovative Medicines Initiative's Big Data for Better Outcomes program, which aims to facilitate the use of diverse data sources to deliver results that reflect health outcomes of treatments that are meaningful for patients, clinicians, regulators, researchers, and healthcare decision makers.

19.11 Preclinical Case: Regulatory Acceptance of Workflows for Heterogeneous Knowledge Integration

Our initial case study work builds on preliminary work led by Edelweiss Connect in collaboration with the FDA and Guardtime to create best practices for trusted reproducible scientific workflows creating evidence to be used in future regulatory evaluation of new methods supporting drug, diagnostics, and medical development (see Figs. 19.5 and 19.6). Increasingly, regulators are considering the use of new scientific evidence generated by a broad toolset of new in silico and in vitro methods as providing value to the evaluation process of new medicines. We plan to engage with experts and regulatory groups from both EMA and the FDA in workshop activities on workflow evaluation with blockchains providing metadata management solutions linked to workflow sub-task processing of datasets ensuring desired properties for

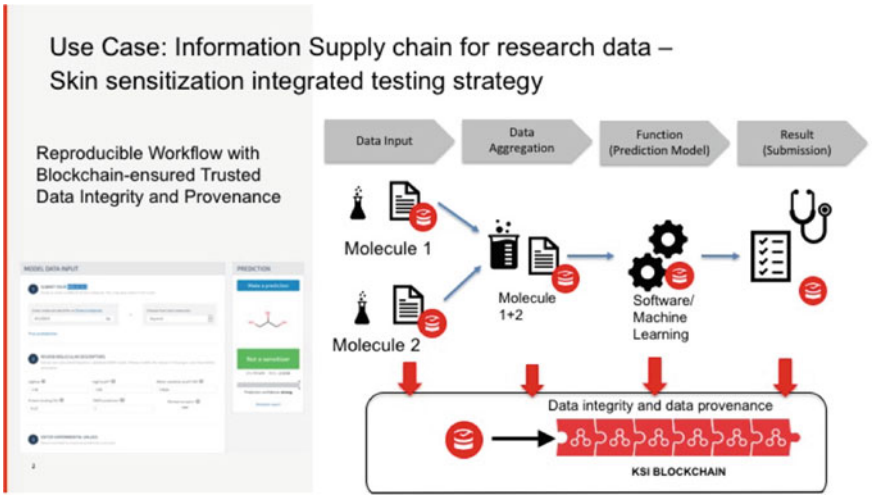

Fig. 19.5 Running the skin sensitization integrating testing application on the Guardtime blockchain

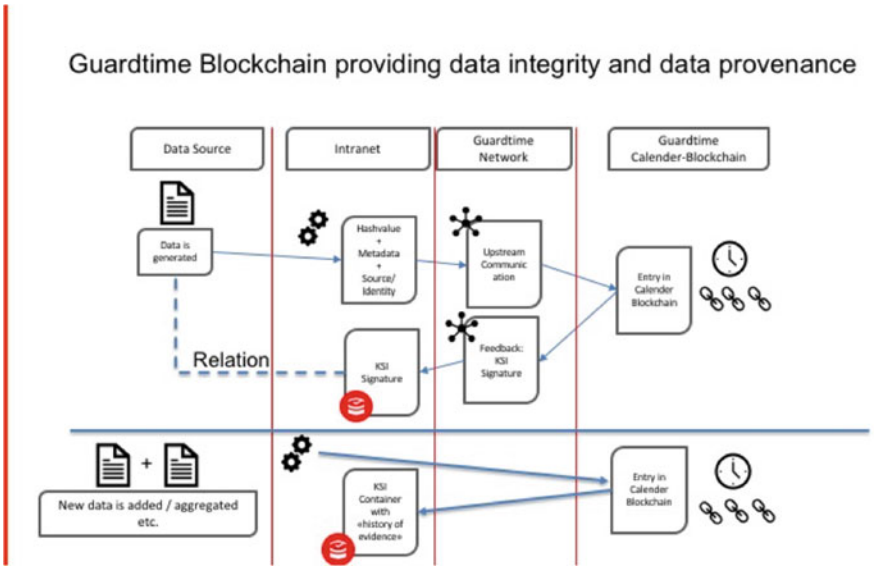

Fig. 19.6 Providing data integrity and provenance on the Guardtime blockchain

privacy, access, provenance, integrity, and completeness. We propose to initially consider the following three cases:

(a) Building and validation of a predictive machine learning model (for property evaluation purposes, e.g., in the qualification of a characterization method of a new nanomedicine under consideration for an NDA);
(b) Evaluation of an integrated testing and assessment defined method for evaluating a safety assessment endpoint, e.g., DILI liability management for an anti-cancer drug;
(c) Use of processed data from bioassays and personalized omics analysis combined with other healthcare data supporting stratification and outcome enhancement in clinical trials (will interact with clinical trial cases).

19.12 Approach to Community Assessment of Best Practice Developments

We will share our experience within the OpenTox community for developing trusted reproducible safety assessment solutions including databases, algorithms, modeling, protocol management, ontology, APIs, Web services, and integrating infrastructure. Following the philosophy of an open knowledge community and sustainability we have pioneered within OpenTox, we will make reference solutions, practices, and guidance created by us available to all the community in open-access mode according

to creative commons licenses, enabling maximum use and update of developments. Both QSAR and ITS applications will be made available providing cases and practice examples for review and discussion both online and during the OpenTox conferences and workshops planned for Europe, USA, and Asian regions. The proposed practices will include:

- Principle of reproducibility—review propositional principle;
- Practices and specifications for application interoperability layers for data—providing APIs to important public databases in predictive toxicology and supporting the principle of reproducibility;
- Reproducible Q(SAR) demonstration and evaluation—reproducible workflow for Q(SAR) model building and applied to published reference predictive models;
- Reproducible integrated testing strategy (ITS) demonstration—reproducible workflow for ITS and risk assessment of skin sensitization.

Based on the results of the discussions and exercises at the OpenTox conferences and workshop, we will draft an OpenTox guidance to provide draft best practice guidance for implementation of the Reproducibility Principles.

The need for reproducibility is increasing dramatically as data analyses becomes more complex, involving larger datasets and more sophisticated computations. By building case studies for in silico reproducibility, we aim to encourage scientists to follow demonstrated examples and publishers to adopt the developed guidelines as recommendations for future publications. Within OpenTox and related initiatives such as OpenRiskNet and NanoCommons, we propose to establish propositional "practices for reproducible in silico computational science" and specific recommendations on applying them to computational toxicology. We will also establish a guideline for data collection (by performing systematic literature review), data curation (both experimental and structural data and involving outlier detection techniques), data management (including long-term storage and publication), tracking (workflow documentation and version control), and licensing of computational tools (giving preference for open and/or free tools). The two areas initially addressed will be: (a) the field of (Quantitative) structure activity relationships, aiming to establish reproducibility, complementing existing OECD guidelines on best practices for QSAR modeling; (b) integrating testing strategies (ITSs) involving a Bayesian approach to weight of evidence from a combination of in silico models and in vitro testing results.

In both cases (QSAR and ITS), we are implementing demonstration implementations as workflows which are presented online to the OpenTox community for evaluation within a community challenge proposed to enhance and enrich best practices while simultaneously reproducing published studies. We will identify key studies published in the last decade; our selection criteria includes: the publications' impact, number of citations and its influence in shaping the field of new alternative methods in safety assessment (voted by a panel of domain experts) as well as the availability and support of the original authors to collaborate, if possible. We will evaluate and select case studies for the community challenge, where we will work with community members and the study authors to rehabilitate their published outcomes for reproducible in silico computations, which will be discussed at an international sci-

entific conference. Finally, we will distill the learnings from the above activities into updated principles, best practice guidance, and improved implementations shared with the community.

Disclaimer The views presented in this article do not necessarily reflect current or future opinion or policy of the US Food and Drug Administration or National Institutes of Health. Any mention of commercial products is for clarification and not intended as endorsement.

References

1. Todeschini R, Consonni V, Mannhold R et al (2009) Molecular descriptors for chemoinformatics, vol 41 (2 volume set), 2nd, revised and enlarged edition. Retrieved from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-3527318526.html. Accessed on 31 Mar 2017
2. Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, Teetz W et al (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J Comput Aided Mol Des 25:533–554
3. Walker T, Grulke CM, Pozefsky D, Tropsha A (2010) Chembench: a cheminformatics workbench. Bioinformatics 26:3000–3001. Oxford University Press
4. Cefic LRI - Toolbox: AMBIT. Retrieved from http://cefic-lri.org/lri_toolbox/ambit/. Accessed on 31 Mar 2017
5. Jaworska J, Hoffmann S. Integrated testing strategy (ITS)—opportunities to better use existing data and guide future testing in toxicology. Available: http://www.altex.ch/resources/ALTEX_4_10_Jaworska_Hoffmann.pdf. Accessed on 30 Mar 2017
6. Hartung T, Luechtefeld T, Maertens A, Kleensang A (2013) Food for thought … integrated testing strategies for safety assessments. ALTEX. NIH Public Access 30:3
7. Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A et al (2012) Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. Regul Toxicol Pharmacol 63:489–504
8. Strickland J, Zang Q, Kleinstreuer N, Paris M, Lehmann DM, Choksi N et al (2016) Integrated decision strategies for skin sensitization hazard. J Appl Toxicol 36:1150–1162
9. Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M (2015) Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Arch Toxicol 89:2355–2383
10. SaferSkin. Retrieved from https://its.douglasconnect.com/. Accessed on 31 Mar 2017
11. Guidance on good data and record management practices. WHO technical report series, No. 996, 2016, Annex 5. Retrieved from http://apps.who.int/medicinedocs/en/m/abstract/Js22402en/. Accessed on 31 Mar 2017
12. Git - Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Git. Accessed on 31 Mar 2017
13. Merkle tree - Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Merkle_tree. Accessed on 31 Mar 2017
14. NixOS Linux. Retrieved from https://nixos.org/. Accessed on 31 Mar 2017
15. Hartung T et al A modular approach to the ECVAM principles on test validity. PubMed—NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15656771. Accessed on 31 Mar 2017
16. EC JRC Report (2014) Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/JRC91361/echa_jrc_sla_report_public_05-09-14_withcover%20ipo.pdf. Accessed on 30 Mar 2017
17. Eskes C, Whelan M (eds) (2016) Validation of alternative methods for toxicity testing. Springer, Cham

18. EUR-Lex—02006R1907-20140410—EN—EUR-Lex. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006R1907-20140410. Accessed on 30 Mar 2017
19. OECD series on testing and assessment, books/OECD series on testing and assessment/ N°69—guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. Retrieved from http://www.oecd-ilibrary.org/docserver/download/9750691e.pdf?expires=1490892571&id=id&accname=guest&checksum=23D26D174819D41691EDC490EAA82326
20. OECD (2016) Retrieved from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)67&doclanguage=en. Accessed on 30 Mar 2017
21. Tollefsen KE et al. Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). PubMed—NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25261300. Accessed on 30 Mar 2017

Barry Hardy is leading Douglas Connect and its team supports the development of new integrating solutions in industrial safety assessment. An example application is that for integrated testing for skin sensitization available at https://its.douglasconnect.com/. He has coordinated the OpenTox project in predictive toxicology and the ToxBank infrastructure development project. He is currently President of the OpenTox Association, founded in 2015 as an international non-profit organization promoting an open knowledge community approach to new methods in predictive toxicology. He recently led the infrastructure development for the IMI EBiSC stem cell banking project and the eNanoMapper project developing OpenTox solutions supporting nanotechnology safety assessment. New projects include leading OpenRiskNet, knowledge infrastructure development for ACEnano and Eu-ToxRisk and translation of research methods to industrial practice within ToxHQ. He has led the development of research and best practice activities in drug design and toxicology through founding the eCheminfo Community of Practice, InnovationWell and leading the Scientists Against Malaria project. He obtained his Ph.D. in 1990 from Syracuse University working in computational science. He was National Research Fellow at the FDA Center for Biologics and Evaluation, Hitchings-Elion Fellow at Oxford University, and CEO of Virtual Environments International.

Daniel Bachler is Senior Software Engineer at Douglas Connect and is specialized in data integration and software container technologies with a focus on reproducibility of scientific software methods. He has worked in different fields from high-performance statistics applications to sensitive data processing pipelines and is highly skilled in many database technologies as well as moderns cloud deployment solutions. He has a strong interest in designing and developing highly maintainable and reliable software solutions by employing purely functional programming languages.

Joh Dokler is Chief Technology Officer (CTO) at Douglas Connect, architecting and developing Web-based solutions. He met the Web in the early 1990 at the Josef Stefan Institute in Ljubljana, Slovenia. Since then he successfully completed more than 100 commercial, academic, and government Web projects. He combines his software development skills with his expertise as an information architect to develop sound, solid, and data-driven Web solutions. He initially studied physics at University of Ljubljana and later obtained his MSc in computing and software engineering at Open University, UK.

Thomas Exner is Chief Scientific Officer (CSO) at Douglas Connect. He studied chemistry and did his Ph.D. at the Technische Universität Darmstadt followed by a postdoc at the University of Saskatchewan, Canada. After 12 years of independent academic research (University of Konstanz and University of Tübingen), he joined the Douglas Connect team in 2015. In April 2016, he

was appointed Chief Scientific Officer (CSO). Main research areas he is coordinating are data and knowledge management, integrating testing strategies combining alternative in vitro and in silico methods for predictive toxicology and risk assessment as well as approaches to foster collaborative research and community activities. He is Author and Co-Author of more than 50 articles in peer-reviewed international journals and several other scientific contributions.

Connor Hardy is Web Publisher and Technical Assistant at Douglas Connect. He is an experienced Web Designer with a demonstrated history. He has experience in Search Engine Optimization (SEO), Corporate Website Management, Client/Product Service, Publication Strategy, Video Editing/Production, and Community Web Project Management. He graduated from Mercyhurst University.

Weida Tong is Director of Division of Bioinformatics and Biostatistics at FDA's National Center for Toxicological Research (NCTR/FDA). He has served as Science Advisory Board Member for several large projects involving multiple institutes in Europe and the USA. He also holds several adjunct positions at universities in the USA and China. His division at FDA works to develop bioinformatic methodologies and standards to support FDA research and regulation and to advance regulatory science and personalized medicine. The most visible projects from his group are (1) leading the microarray quality control (MAQC) consortium to develop standard analysis protocols and quality control metrics for emerging technologies to support regulatory science and precision medicine; (2) development of liver toxicity knowledge base (LTKB) for drug safety; (3) in silico drug repositioning for the enhanced treatment of rare diseases; and (4) development of the FDA bioinformatics system, ArrayTrackTM suite, to support FDA review and research on pharmacogenomics. In addition, his group also specializes in molecular modeling and QSARs with a specific interest in estrogen, androgen, and endocrine disruptor. He has published more than 230 papers and chapters.

Daniel Burgwinkel is an expert in blockchain technology located in Basel, Switzerland, and currently working for Guardtime. He is an experienced Advisor and Program Manager with a demonstrated history of working in the pharmaceuticals industry, financial services, and e-government. He is skilled in blockchain technology, information governance, record management, and business process management. He is a strong entrepreneurship professional with a Doctor of Philosophy (Ph.D.) from University of St. Gallen and a Masters in Business Engineering.

Richard Bergström is External Pharma Lead, SICPA SA. He previously was Director General of the European Federation of Pharmaceutical Industries and Associations (EFPIA) from April 2011 to 2016. He served for nine years as Director General of LIF, the Swedish Association of the Pharmaceutical Industry, following positions in Switzerland in regulatory affairs at the pharmaceutical companies Roche and Novartis. He was also appointed by the Swedish Government to the Board of the Karolinska Institutet. He is Pharmacist by training, receiving his MScPharm from the University of Uppsala, Sweden in 1988.

Index

H. Hong (ed.), *Advances in Computational Toxicology*, Challenges and Advances in Computational Chemistry and Physics 30,
https://doi.org/10.1007/978-3-030-16443-0