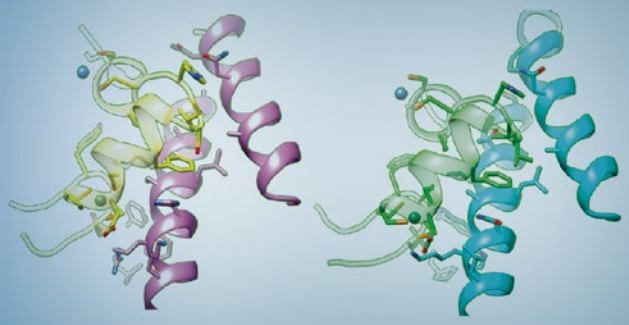
Methods in Molecular Biology 2247

Springer Protocols



Arnaud Poterszman *Editor*

Multiprotein Complexes

Methods and Protocols

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences,
University of Hertfordshire, Hatfield,
Hertfordshire, UK

For further volumes:
http://www.springer.com/series/7651

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

Multiprotein Complexes

Methods and Protocols

Edited by

Arnaud Poterszman

Institut de Génétique et de Biologie Moléculaire et Cellulaire, Center for Integrated Biology, Integrated Structural Biology Department, Equipe labellisée Ligue Contre le Cancer, CNRS UMR 7104 - Inserm U 1258, University of Strasbourg, Illkirch, France

Humana Press

Editor
Arnaud Poterszman
Institut de Génétique et de Biologie
Moléculaire et Cellulaire, Center for
Integrated Biology, Integrated
Structural Biology Department, Equipe
labellisée Ligue Contre le Cancer
CNRS UMR 7104 - Inserm U 1258
University of Strasbourg
Illkirch, France

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-0716-1125-8 ISBN 978-1-0716-1126-5 (eBook)
https://doi.org/10.1007/978-1-0716-1126-5

Preface

Molecular complexes of interacting proteins govern virtually all biological processes such as metabolism, cell signaling, DNA repair, or gene expression. Macromolecular assemblies are also of great biomedical relevance as factors that perturb biomolecular interaction networks underlie a number of diseases, and deliberate inhibition of protein–protein interactions is an increasingly common strategy in drug discovery initiatives. Unraveling their functions and mechanisms of action is often only potentially accessible through a detailed structural description and the integration of dynamic information. This volume of the *Methods in Molecular Biology* series aims to provide the scientific community with strategies and detailed protocols for the preparation of macromolecular complexes and their characterization in view of structural analysis.

Protein engineering and production are essential tools for structural, biophysical, and functional studies as well as for biotechnology and medical applications. Strategies to prepare proteins and protein complexes have tremendously been improved in part thanks to recent structural genomic programs. Yet, no universal solution has been implemented, and the production and/or reconstitution of protein complexes remains a major bottleneck. This is in particular the case for complexes composed of many subunits which are often incompletely characterized. Additional difficulties result from their low natural abundance or their versatile nature, in part because regulation often involves the formation of transient complexes with low binding constants and in part because their composition varies with the physiological context.

The first section of this book focuses on sample preparation. While Chapters 1 and 2 concentrate on strategies for recombinant expression of multiprotein complexes in prokaryotic and eukaryotic hosts, Chapter 3 illustrates how genome editing with the CRISPR-Cas9 system can be used to precisely modify protein coding genes in mammalian cells; this offers the possibility to replace any coding gene by a reshaped version fused to an affinity tag protein or to a fluorescent reporter, enabling the characterization of endogenous macromolecular complexes expressed under near physiological conditions. In the first section, the production of recombinant antibodies and artificial binding proteins which emerge as key reagents in multiprotein complex research, as inhibitors of protein–protein interactions to modulate activity, as probes for cellular imaging or to facilitate structural determination is also discussed. Chapter 4 details the production of recombinant antibodies in different formats, Chapter 5 the intervening removable affinity tag (iRAT) system for the production of recombinant antibody fragments, and Chapter 6 the isolation of artificial binding proteins (Affimer reagents) based on a non-antibody scaffold.

The second section of the book details a set of biophysical methods that can provide useful indicators for sample optimization and often complement structural information obtained with core technologies for structure determination (X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy) by quantitative solution data, helping to understand how biological systems function. Three chapters focus on techniques for the biophysical characterization of biological macromolecules and their complexes. Chapter 7 details interaction measurements of protein–DNA complexes by isothermal titration calorimetry (ITC) and microscale thermophoresis (MST), Chapter 8 the use of the switchSENSE technology for the analysis of enzyme kinetics, and Chapter 9 the sedimentation

velocity methods for the characterization of protein heterogeneity and protein affinity interactions. Another set of articles describe mass spectrometry (MS)-based approaches for macromolecular complex analysis. Chapter 10 focuses on the applications of native mass spectrometry for the characterization of multiprotein complexes ranging from 16 to 801 kDa, Chapter 11 on hydrogen/deuterium exchange mass spectrometry for analyzing protein–DNA interactions, and Chapter 12 on integrative mass spectrometry-based approaches for modeling macromolecular assemblies.

Although high-resolution structure determination using X-ray crystallography or single-particle cryo-electron microscopy (cryo-EM) is now producing a rapid stream of breakthroughs in structural biology, the preparation of suitable crystals or hydrated frozen samples on EM grids is often quite challenging. Purified samples, intact and structurally homogeneous in the test tube, may not crystallize or survive the standard methods of preparing thin aqueous films on grids. In the case of cryo-EM, optimization of sample stability and extensive screening of parameters for grid preparation are often required to collect high-quality datasets. The two last chapters of this section address sample optimization. Chapter 13 provides detailed protocols for preparing negatively stained and hydrated frozen EM-grid while Chapter 14 addresses solubilization screening using membrane proteins as model system.

Finally, the third section of this book addresses the characterization of multiprotein complexes in a cellular environment using state-of-the-art imaging technologies and in vivo approaches. Chapter 15 describes practical aspects of super-resolution imaging and Chapter 16 multi-color FRET-FLIM microscopy in live cells. Chapters 17 and 18 present applications of directed evolution systems and of context-specific and proximity-dependent labeling using the BioID technology.

This book is expected to be used not only by structural/molecular biologists who need to prepare multi-components complexes for their own applications but also by scientists from other fields who are working on macromolecular assemblies from other standpoints and need an overview of state-of-art approaches. I am especially thankful to all the authors for their great contributions, devoting their valuable time to the preparation of the manuscripts. I am also indebted to the Series Editor John M. Walker, to the editorial staff members of Springer and to Marie Christine Poterszman for their kind support in making this book publishable. I hope that this volume provides a useful overview preparation and structural analysis of macromolecular complexes and fills a need for well-described hands-on protocols.

Illkirch, France *Arnaud Poterszman*

Contents

Contributors

Abdullah Ahmed • *Central Laser Facility, Science and Technology Facilities Council (STFC) Rutherford Appleton Laboratory, Research Complex at Harwell, Oxford, UK; Plant Cell Biology, Biological and Medical Sciences, Oxford Brookes University, Oxford, UK*

Leonid Andronov • *Centre for Integrative Biology (CBI), Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*

Magali Aumont-Nicaise • *Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-s-Yvette, France*

Guillaume Bec • *Institut de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France*

Julien Béthune • *Hamburg University of Applied Sciences, Hamburg, Germany; Cluster of Excellence CellNetworks, Heidelberg University, Heidelberg, Germany; Heidelberg University Biochemistry Center (BZH), Heidelberg, Germany*

Catherine Birck • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*

Stanley W. Botchway • *Central Laser Facility, Science and Technology Facilities Council (STFC) Rutherford Appleton Laboratory, Research Complex at Harwell, Oxford, UK; Plant Cell Biology, Biological and Medical Sciences, Oxford Brookes University, Oxford, UK*

Alice Brion • *Laboratoire Structure et Instabilité des Génomes, Muséum National d'Histoire Naturelle (MNHN), Institut National de la Santé et de la Recherche Médicale (INSERM), U1154, Centre National de la Recherche Scientifique (CNRS), UMR7196, Paris, France*

Jean-Baptiste Charbonnier • *Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-s-Yvette, France*

Sarah Cianférani • *Laboratoire de Spectrométrie de Masse BioOrganique, Institut Pluridisciplinaire Hubbert Curien (IPHC), Centre National de la Recherche Scientifique (CNRS), UMR 7178, Université de Strasbourg, Strasbourg, France*

Jean-Paul Concordet • *Laboratoire Structure et Instabilité des Génomes, Muséum National d'Histoire Naturelle (MNHN), Institut National de la Santé et de la Recherche Médicale (INSERM), U1154, Centre National de la Recherche Scientifique (CNRS), UMR7196, Paris, France*

Emily Cooke • *Central Laser Facility, Science and Technology Facilities Council (STFC) Rutherford Appleton Laboratory, Research Complex at Harwell, Oxford, UK*

Vincent Corvest • *CALIXAR, Lyon, France*

Michel Desmadril • *Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-s-Yvette, France*

Christine Ebel • *Institut de Biologie Structurale (IBS), Univ Grenoble Alpes, Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Centre National de la Recherche Scientifique (CNRS), Grenoble, France*

Maria Ecke • *Department of Structural Biochemistry, Max-Planck-Institute of Molecular Physiology, Dortmund, Germany*
Stefanie Egetemaier • *Cluster of Excellence CellNetworks, Heidelberg University, Heidelberg, Germany; Heidelberg University Biochemistry Center (BZH), Heidelberg, Germany*
Eric Ennifar • *Institut de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France*
Stéphane Erb • *Laboratoire de Spectroméérie de Masse BioOrganique, Universitéde Strasbourg, CNRS, IPHC UMR7178, Strasbourg, France*
Karim Essabri • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*
Ruzena Filandrova • *Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic; Faculty of Sciences, Charles University, Prague, Czech Republic*
Alexandre Frechard • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Integrated Structural Biology, Equipe labellisée Ligue Contre le Cancer, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*
Sylvain Geny • *Laboratoire Structure et Instabilité des Génomes, Muséum National d'Histoire Naturelle (MNHN), Institut National de la Santé et de la Recherche Médicale (INSERM), U1154, Centre National de la Recherche Scientifique (CNRS), UMR7196, Paris, France*
Amandine Gontier • *Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-s-Yvette, France; Structural Biology and Biophysics (SBB), Bio Structure and Biophysics (BSB), Sanofi, Vitry-sur-Seine, France*
Michael Hust • *Technische Unversität Braunschweig, Institut für Biochemie, Biotechnologie und Bioinformatik, Abteilung Biotechnologie, Braunschweig, Germany; YUMAB GmbH, Braunschweig, Germany*
So Iwata • *Department of Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan*
Sophie Jacquemin • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Equipe labellisée Ligue Contre le Cancer, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*
Anass Jawhari • *Independent Advisor, Lyon, France*
Qiyang Jiang • *Viva Biotech (Shanghai) Limited, Shanghai, China*
Burak V. Kabasakal • *School of Biochemistry, University of Bristol, Bristol, UK; Institute of Accelerator Technologies, Ankara University, Gölbaçı, Ankara, Turkey*
Alan Kadek • *Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic; Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany*
Daniel Kavan • *Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic*
Bruno P. Klaholz • *Centre for Integrative Biology (CBI), Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*

Olga Kolesnikova • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Département of Integrated Structural Biology, Equipe Labellisée Ligue, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France; Structural and Computational Biology Unit, EMBL Heidelberg, Heidelberg, Germany*

Jens König • *Department of Structure and Function of Proteins, Helmholtz Zentrum für Infektionsforschung GmbH, Braunschweig, Germany*

Andy M. Lau • *Department of Chemistry, King's College London, London, UK*

Petr Man • *Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic*

Julien Marcoux • *Institut de Pharmacologie et de Biologie Structurale (IPBS), Université de Toulouse, CNRS, UPS, Toulouse, France*

Michael J. McPherson • *Astbury Centre for Structural and Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds, UK*

Clément Nemoz • *Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-s-Yvette, France*

Heinz Neumann • *Department of Structural Biochemistry, Max-Planck-Institute of Molecular Physiology, Dortmund, Germany; Department of Chemical Engineering and Biotechnology, University of Applied Sciences, Darmstadt, Germany*

Norimichi Nomura • *Department of Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan*

Yayoi Nomura • *Department of Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan*

Petr Novak • *Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic*

Simon Pichard • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Equipe labellisée Ligue Contre le Cancer, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*

Argyris Politis • *Department of Chemistry, King's College London, London, UK*

Arnaud Poterszman • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, Center for Integrated Biology, Integrated Structural Biology Department, Equipe labellisée Ligue Contre le Cancer, CNRS UMR 7104 - Inserm U 1258, University of Strasbourg, Illkirch, France*

Cinthia Amaya Ramirez • *Cluster of Excellence CellNetworks, Heidelberg University, Heidelberg, Germany; Heidelberg University Biochemistry Center (BZH), Heidelberg, Germany*

Jean-Baptiste Renaud • *Laboratoire Structure et Instabilité des Génomes, Muséum National d'Histoire Naturelle (MNHN), Institut National de la Santé et de la Recherche Médicale (INSERM), U1154, Centre National de la Recherche Scientifique (CNRS), UMR7196, Paris, France*

Virginie Ropars • *Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-s-Yvette, France*

Paola Rossolillo • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*

Yumi Sato • *Department of Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan*
Christiane Schaffitzel • *School of Biochemistry, University of Bristol, Bristol, UK*
Damian Schiller • *Department of Structural Biochemistry, Max-Planck-Institute of Molecular Physiology, Dortmund, Germany*
Jennifer Schoberer • *Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, BOKU Vienna, Vienna, Austria*
Patrick Schultz • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Integrated Structural Biology, Equipe labellisée Ligue Contre le Cancer, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*
Grigory Sharov • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Integrated Structural Biology, Equipe labellisée Ligue Contre le Cancer, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France; EM Facility, MRC Laboratory of Molecular Biology, Cambridge, UK*
Martin Spinck • *Department of Structural Biochemistry, Max-Planck-Institute of Molecular Physiology, Dortmund, Germany*
Anna A. S. Tang • *Astbury Centre for Structural and Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds, UK*
Christian Tiede • *Astbury Centre for Structural and Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds, UK*
Darren C. Tomlinson • *Astbury Centre for Structural and Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds, UK*
Joop van den Heuvel • *Department of Structure and Function of Proteins, Helmholtz Zentrum für Infektionsforschung GmbH, Braunschweig, Germany*
Paloma F. Varela • *Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-s-Yvette, France*
Jean-Luc Vonesch • *Centre for Integrative Biology (CBI), Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*
Maximilien Werderer • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Integrated Structural Biology, Equipe labellisée Ligue Contre le Cancer, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France*
Gala Ramon Zamorano • *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Département of Integrated Structural Biology, Equipe Labellisée Ligue, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg, Illkirch, France; Department of Life and Medical Sciences, University of Hertfordshire, Hatfield, UK*

Part I

Recombinant Expression and Isolation from Endogenous Source

Chapter 1

Production of Multi-subunit Membrane Protein Complexes

Burak V. Kabasakal, Qiyang Jiang, and Christiane Schaffitzel

Abstract

Membrane proteins constitute an important class of proteins for medical, pharmaceutical, and biotechnological reasons. Understanding the structure and function of membrane proteins and their complexes is of key importance, but the progress in this area is slow because of the difficulties to produce them in sufficient quality and quantity. Overexpression of membrane proteins is often restricted by the limited capability of translocation systems to integrate proteins into the membrane and to fold them properly. Purification of membrane proteins requires their isolation from the membrane, which is a further challenge. The choice of expression system, detergents, and purification tags is therefore an important decision. Here, we present a protocol for expression in bacteria and isolation of a seven-subunit membrane protein complex, the bacterial holo-translocon, which can serve as a starting point for the production of other membrane protein complexes for structural and functional studies.

Key words Membrane protein complex, Expression in bacteria, Endogenous host, Complex purification, Translocon

1 Introduction

1.1 Aim of the Study

Membrane proteins represent more than 25% of the proteome of all cells. They mediate the cell's interaction with its environment, i.e., transmission of signals, cell adhesion, and transport across membranes. Membrane proteins are of prime pharmaceutical interest as they constitute ~50% of known potential drug targets. In addition to this, membrane proteins are the natural entry and/or anchoring points for infectious agents. Dysfunctional membrane proteins are the basis of many disorders such as cystic fibrosis and Alzheimer's. Understanding the function and molecular structure of this class of proteins is of key interest.

As in the cytosol, the majority of proteins in the membrane occur in and function in complexes. For instance in budding yeast, membrane proteins were found to interact on average with two to three interaction partners [1]. Due to their association with cellular membranes, expression and purification of membrane proteins and their complexes are often a challenging task. Progress is often

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_1,

slowed down by the requirement to optimize protein expression and to subsequently identify conditions to solubilize the membrane proteins without compromising integrity and activity. While the optimal expression and purification conditions needs to be optimized for each new protein and complex, standard protocols as presented here offer a good starting point for subsequent optimization by variation of conditions.

1.2 Recombinant Expression in Escherichia coli

E. coli has traditionally been, and still is, the most popular expression system for proteins. This is because of low costs, low expenditure of time, and the fact that it is less labor-intense than other expression systems.

1.2.1 Choice of Promoter and Strain

When expressing membrane proteins, saturation of the *E. coli* membrane protein translocation and folding machinery should be avoided. This would lead to protein misfolding, aggregation, and formation of inclusion bodies. Very strong promoters such as the T7 promoter frequently lead to these problems. Weaker promoters synthesize membrane proteins at more moderate levels. Examples include T7/*lac* hybrid promoters or the tightly regulated arabinose promoter, both of which are often used for membrane protein expression [2].

Addressing the same issues, tailor-made strains have been developed for membrane protein production in *E. coli*. The T7 RNA polymerase-based expression strain BL21λ(DE3) and its mutant strains C41λ(DE3) and C43λ(DE3) [3] are the most popular strains for expressing membrane proteins in *E. coli* [2]. C41 and C43 contain mutations in the *lac*UV5 promoter region of the gene encoding for T7 RNA polymerase leading to lower expression of the T7 RNA polymerase [4] and consequently slower transcription and translation rates of the proteins under the control of T7 promoter. The same principle applies to the *E. coli* BL21-AI strain (Invitrogen) which expresses T7 RNA polymerase under the control of a tightly regulated arabinose promoter.

1.2.2 Other Considerations

Auto-induction media are commonly used for improved membrane protein production [5]. They can be used for all IPTG-inducible expression systems. Protein production is induced in auto-induction media at high cell density when the glucose in the media is depleted, which is during the mid/late log phase of growth. Lactose uptake then leads to the production of allolactose which causes the release of the lac repressor and induction of T7 RNA polymerase from the *lac*UV5 promoter.

For membrane protein complexes, a common strategy is the co-expression of the proteins from their natural or artificial operons [6] (*see* example in Fig. 1a). Co-expression of all subunits of a complex allows complex assembly in vivo in the cell. This is particularly important for membrane protein complexes, where

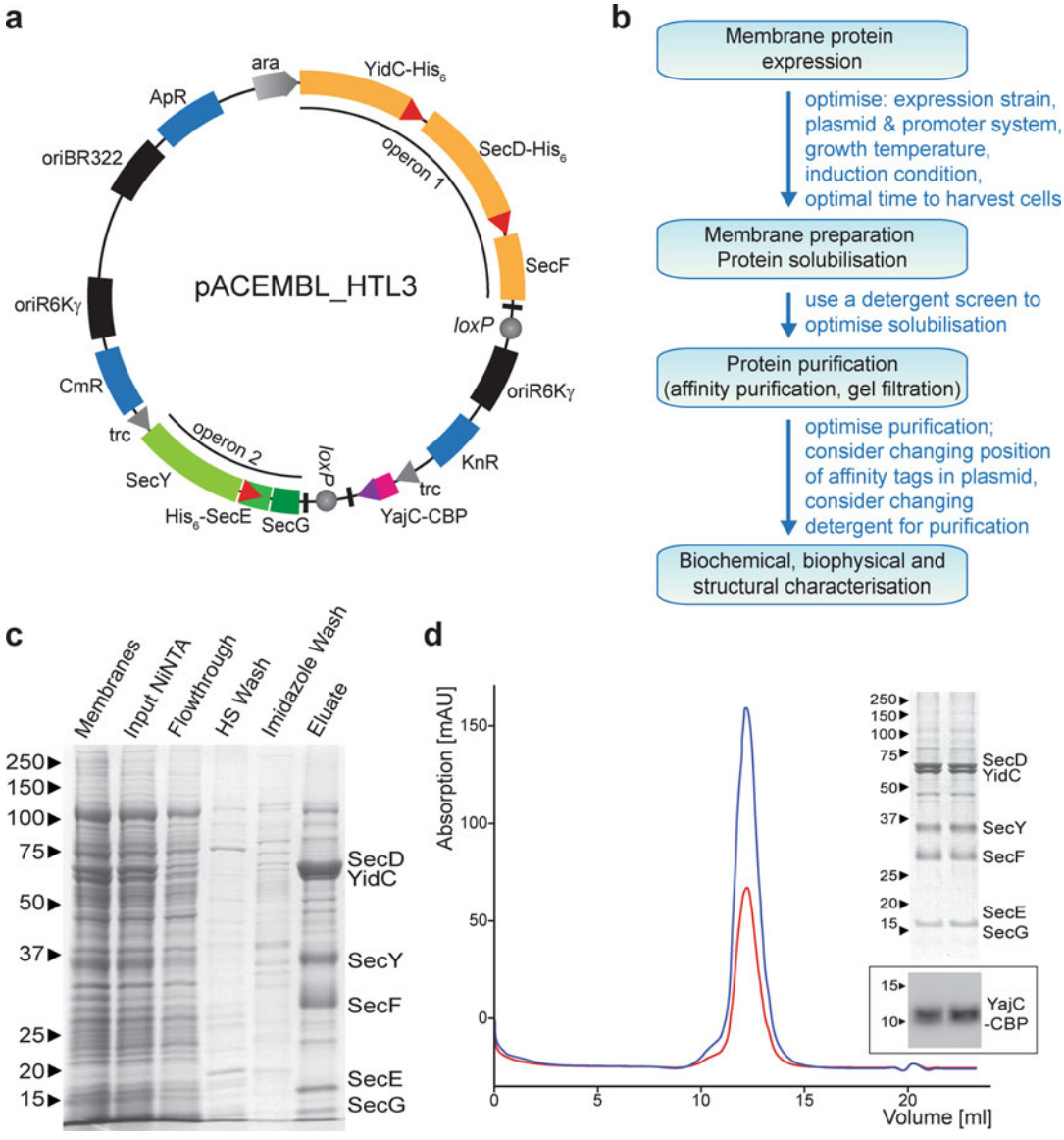


Fig. 1 The production of the holo-translocon complex, comprising seven membrane proteins. (**a**) The expression vector for the holo-translocon was generated using the ACEMBL system [13]. It comprises two artificial operons encoding for YidC, SecD, and SecF (operon 1) and for SecY, SecE, and SecG (operon 2), as well as the gene encoding for YajC with a C-terminal CBP-tag. The hexa-histidine tags on SecD, YidC, and SecE are indicated as red triangles. The position of the CBP-tag is indicated with a purple triangle. Origins of replication of donor (oriR6Kγ) and acceptor (oriBR322) vectors are indicated in black. Arabinose (ara) and trc promoters are shown as grey arrows and terminators as black rectangles. Antibiotic resistance genes are colored blue (*Ap* ampicillin, *Kn* kanamycin, *Cm* chloramphenicol). (**b**) Scheme showing the general workflow for holo-translocon production and parameters that require optimization when a new membrane protein complex is produced. (**c**) Coomassie-stained SDS PAGE gel of holo-translocon, analyzing the membrane fraction, the detergent-solubilized protein fraction (Ni-NTA input), the flow-through from Ni-NTA, wash fractions, and the eluate peak fraction (*HS* high salt). The bands for holo-translocon subunits are marked. (**d**) Size exclusion chromatogram of holo-translocon and SDS PAGE of the peak fractions. The blue and red curves show the absorption at 280 nm and 260 nm, respectively. Upper inset: Coomassie-stained SDS PAGE showing the peak fractions; below: Western blot using an antibody directed against the CBP-tag to detect YajC-CBP

reconstitution of the complexes from the purified, detergent-solubilized subunits is often not possible. Co-expression of the subunits of a complex from one polycistronic messenger RNA supports a balanced expression. Unexpectedly, *E. coli* operons were shown to be translated with the proper stoichiometry even when the subunit stoichiometry significantly deviates from one copy each per complex [7]. Accordingly, the order of genes in a synthetic operon needs to be experimentally optimized if the protein production is not balanced [8].

1.2.3 Recombinant Membrane Protein Expression

When expressing eukaryotic proteins in *E. coli*, codon optimization of the corresponding gene is recommended in order to account for the *E. coli* codon usage bias and thus avoid low or unreliable expression. Overexpression of recombinant membrane proteins can be optimized by fusing a green fluorescent protein (GFP) to the membrane protein. Using fluorescence detection, GFP serves as a marker for membrane protein localization, quantity of folded protein, and the efficiency of solubilization from the membrane by detergents [9, 10].

In this chapter, we describe the production of the *E. coli* holo-translocon, a seven-subunit membrane protein complex comprising SecY, SecE, SecG, SecD, SecF, YajC, and YidC. It is one of the largest recombinant membrane protein complexes produced to date. The holo-translocon complex comprises all proteins known to be important for protein translocation into and across the *E. coli* plasma membrane [11, 12]. For the balanced expression of the holo-translocon, the ACEMBL system was used [13, 14]. ACEMBL comprises acceptor and donor vectors which have multi-integration elements for individual genes or polycistrons. Donor vectors differ from acceptor vectors and other commonly used plasmids in that they comprise a conditional origin of replication (oriR6Kγ) and therefore need to be propagated in strains with the phage R6Kγ *pir* gene. All vectors have a LoxP site, but different antibiotic resistances. Donor and acceptor vectors can be fused by Cre recombinase and the donor–acceptor fusion plasmid can be selected in a conventional *pir*$^-$ *E. coli* strain using the appropriate combination of antibiotics [13].

For the expression of the holo-translocon, the conserved SecY, SecE, and SecG subunits were encoded by a synthetic polycistronic operon on a donor vector (Fig. 1a). A second donor vector carried the gene for YajC, which is the only nonessential subunit of the holo-translocon [15]. These two donor vectors were fused to an acceptor vector with a polycistron encoding for YidC, SecD, and SecF. Hexa-histidine tags were fused to YidC, SecD, and SecG and a calmodulin-binding peptide to YajC (Fig. 1a). Double-affinity purification via these tags results in pure holo-translocon complex, comprising one copy of each subunit [15, 16].

1.3 Recombinant Expression in Eukaryotic Expression Systems

Expression of recombinant eukaryotic membrane protein in *E. coli* frequently results in protein misfolding and aggregation. This can be due to differences in the lipid composition, a different chaperone repertoire, or the lack of post-translational modifications required to obtain folded, active membrane protein. In these cases, eukaryotic expression systems that are more closely related to the natural source of the protein have to be used. Unfortunately, membrane protein expression in mammalian cells often results in moderate or low protein yields. This is attributed to a limited capability of mammalian transport systems to import additional proteins or to replace its own proteins within the membranes of a particular subcellular compartment. GFP fusion to the membrane protein of interest is less useful for quality control and for optimizing expression in mammalian cells because in contrast to *E. coli*, the GFP will fold independently from the membrane protein fused to it. Thus, the GFP fusion method detects folded proteins as well as misfolded proteins which are retained in the endoplasmic reticulum membrane [17].

The majority of membrane proteins produced for structural and functional analysis to date are from baculovirus insect cells [18]. Baculovirus insect cell expression has been particularly successful for production of G protein-coupled receptors (GPCRs), leading to the first crystal structures of this class of membrane proteins [19, 20]. The MultiBac system [21] has the same modular architecture of donor/acceptor vectors as described for the ACEMBL system for *E. coli* expression above. This supports the production of membrane protein complexes in insect cells, such as the active γ-secretase complex comprised of four membrane protein subunits [22].

1.4 Overexpression in the Endogenous Host

Isolation of membrane protein complexes from the natural host used to be restricted to membrane protein machines which are highly abundant in the cell, such as complexes for photosynthesis, respiration, and oxidative phosphorylation [23]. However, with the advent of miniaturization and automatization of biochemical, biophysical, and structural analyses, less abundant membrane protein complexes can now be studied.

Purification of endogenous complexes is significantly facilitated if a purification tag is fused to one of the subunits, which is possible if the source can be genetically modified. The tagged subunit is ideally an integral subunit of the complex. The chosen affinity purification tag should be compatible with the presence of detergents in the buffer (frequent choices include His-tags, 3xFLAG-tag, and CBP-tag).

The *E. coli* holo-translocon production described below uses the endogenous host, thus providing the required chaperone repertoire and the physiological lipid composition. The purified complex has been shown to be active in co- and post-translational

translocation and successfully used for biophysical and structural analyses by electron cryo-microscopy and small-angle neutron scattering [14–16]. A scheme for the general workflow of holo-translocon production and parameters to be optimized is presented in Fig. 1b.

2 Materials

Prepare all solutions using double-distilled water and analytical grade reagents. Store all stock solutions at 4 °C.

2.1 Expression of HTL

1. pACEMBL HTL3 plasmid [15] (Fig.1a).
2. Chemically competent or electrocompetent *E. coli* BL21 Star (DE3) cells (Thermo Fisher) or *E. coli* C43λ(DE3) cells [3] are prepared according to published protocols [24, 25].
3. Prepare antibiotic stock solutions: 100 mg/mL ampicillin, 25 mg/mL chloramphenicol (dissolved in EtOH), and 50 mg/mL kanamycin.
4. Prepare LB medium: 10 g/L Bacto tryptone, 5 g/L yeast extract, 10 g/L NaCl (verify pH 7.5, or adjust with NaOH), and autoclave.
5. Prepare LB agar: Add 15 g/L Bacto agar to 1 L LB medium and autoclave.
6. Incubator 37 °C.
7. Shaker 37 °C.
8. Spectrophotometer.
9. Baffled 5 L Erlenmeyer flasks.
10. Prepare *Terrific Broth:* 24 g/L yeast extract, 20 g/L Bacto tryptone, 4 mL/L glycerol, 0.017 M KH_2PO_4, 0.072 M K_2HPO_4, and autoclave.
11. Prepare stock solutions of 10% L-arabinose by dissolving 1 g L-arabinose in 10 mL of *water*. Filter with 0.22 μm syringe filter and store aliquots at 4 °C. 1 M IPTG (isopropyl β-D-1-thiogalactopyranoside) is prepared by dissolving 2.38 g IPTG in 10 mL of water. Filter with a 0.22 μm syringe filter.
12. Centrifuge, e.g., Sorvall Lynx 6000 with F9 6x1000 Lex rotor (Thermo Fisher).

2.2 Membrane Preparation

1. French Press (Sim Aminco, Spectronic Instruments) or Cell Disrupter/Microfluidizer (Microfluidics).
2. Centrifuge, e.g., Sorvall RC6 and SS-34 rotor (Sorvall).
3. Prepare 1 L HSGM buffer: 20 mM HEPES-KOH pH 8.0 (20 mL of 1 M stock solution), 130 mM NaCl (26 mL of

5 M stock solution), 5 mM $MgOAc_2$ (5 mL of 1 M stock solution), 10% glycerol.

4. cOmplete protease inhibitor cocktail tablets, EDTA-free (Roche).
5. Prepare 100 mM PMSF (phenylmethylsulfonyl fluoride) stock solution: add 17.4 mg of *PMSF* per 1 mL of isopropanol and freeze 20 μL aliquots and store at −20 °C.
6. Ultracentrifuge, Ti70 rotor, and centrifugation tubes (Beckman Coulter).
7. Dounce homogenizer.

2.3 Detergent Solubilization

1. Prepare 10% w/v DDM (n-dodecyl-β-D-maltoside, Anatrace) stock solution in water. Dissolve by gently agitating; do not vortex.
2. Head-over-tail rotator or equivalent.
3. Ultracentrifuge, Ti70 rotor, and centrifugation tubes (Beckman Coulter).

2.4 NiNTA Affinity Chromatography

1. Ni-NTA agarose resin (Qiagen).
2. Gravity flow column (e.g., MoBiTec, #10141).
3. Prepare 100 mL Ni-NTA buffer: 20 mM HEPES-KOH pH 8.0 (2 mL of 1 M stock solution), 130 mM NaCl (2.6 mL of 5 M stock solution), 5 mM $MgOAc_2$ (0.5 mL of 1 M stock solution), 10% glycerol, 10 mM imidazole (1 mL of 1 M stock solution, adjusted to pH 8.0), 0.1% DDM (1 mL of 10% stock solution).
4. Prepare 50 mL high salt buffer: 20 mM HEPES-KOH pH 8.0 (1 mL of 1 M stock solution), 500 mM NaCl (5 mL of 5 M stock solution), 5 mM $MgOAc_2$ (250 μL of 1 M stock solution), 10% glycerol, 10 mM imidazole (0.5 mL of 1 M stock solution, adjusted to pH 8.0), 0.1% DDM (0.5 mL of 10% stock solution).
5. Prepare 100 mL Ni-NTA Wash buffer: 20 mM HEPES-KOH pH 8.0 (2 mL of 1 M stock solution), 130 mM NaCl (2.6 mL of 5 M stock solution), 5 mM $MgOAc_2$ (0.5 mL of 1 M stock solution), 10% glycerol, 50 mM imidazole (5 mL of 1 M stock solution, adjusted to pH 8.0), 0.1% DDM (1 mL of 10% stock solution).
6. Prepare 100 mL Ni-NTA elution buffer: 20 mM HEPES-KOH pH 8.0 (2 mL of 1 M stock solution), 130 mM NaCl (2.6 mL of 5 M stock solution), 5 mM $MgOAc_2$ (0.5 mL of 1 M stock solution), 10% glycerol, 300 mM imidazole (30 mL of 1 M stock solution, adjusted to pH 8.0), 0.1% DDM (1 mL of 10% stock solution).
7. Nanodrop spectrophotometer (Thermo Scientific).

2.5 Desalting Step

1. 5 mL HiTrap Desalting column prepacked with Sephadex G-25 resin (GE Healthcare).
2. AKTA protein purification system or equivalent.
3. Prepare 100 mL CBP-binding buffer: 50 mM HEPES-KOH pH 8.0 (5 mL of 1 M stock solution), 130 mM NaCl (2.6 mL of 5 M stock solution), 10% glycerol, 2 mM $CaCl_2$ (200 μL of 1 M stock solution), 0.03% DDM (0.3 mL of 10% stock solution).

2.6 Calmodulin Affinity Chromatography

1. Calmodulin affinity resin (Agilent Technology).
2. Gravity flow column (e.g., MoBiTec, #10141).
3. Prepare 100 mL CBP-binding buffer: 50 mM HEPES-KOH pH 8.0 (5 mL of 1 M stock solution), 130 mM NaCl (2.6 mL of 5 M stock solution), 10% glycerol, 2 mM $CaCl_2$ (200 μL of 1 M stock solution), 0.03% DDM (0.3 mL of 10% stock solution).
4. Prepare 100 mL CBP washing buffer: 50 mM HEPES-KOH pH 8.0 (5 mL of 1 M stock solution), 130 mM NaCl (2.6 mL of 5 M stock solution), 10% glycerol, 0.2 mM $CaCl_2$ (20 μL of 1 M stock solution), 0.03% DDM (0.3 mL of 10% stock solution).
5. Prepare 25 mL CBP elution buffer: 50 mM HEPES-KOH pH 8.0 (1.25 mL of 1 M stock solution), 400 mM NaCl (2 mL of 5 M stock solution), 3% glycerol, 2 mM EGTA (100 μL of 0.5 M stock solution), 0.03% DDM (0.3 mL of 10% stock solution).
6. NanoDrop spectrophotometer (Thermo Scientific).

2.7 Sample Concentration

1. Amicon Ultra centrifugal filter units, molecular weight cut-off 100 kDa, Ultra-15 (Amicon).
2. Centrifuge, e.g., Heraeus Megafuge 16R.
3. Prepare 10 mL S6 buffer: 20 mM HEPES-KOH pH 8.0 (0.2 mL of 1 M stock solution), 130 mM NaCl (0.26 mL of 5 M stock solution), 5 mM $MgOAc_2$ (50 μL of 1 M stock solution), 0.03% DDM (30 μL of 10% stock solution).

2.8 Size Exclusion Chromatography

1. Superose 6 Increase 10/300 GL column (GE Healthcare).
2. AKTA protein purification system or equivalent.
3. Prepare 250 mL S6 buffer: 20 mM HEPES-KOH pH 8.0 (5 mL of 1 M stock solution), 130 mM NaCl (6.5 mL of 5 M stock solution), 5 mM $MgOAc_2$ (1.25 mL of 1 M stock solution), 0.03% DDM (0.75 mL of 10% stock solution).

3 Methods

Once the proteins are solubilized from the isolated membranes, it is important to work at 4 °C and proceed quickly with the protein purification to avoid protein aggregation and degradation.

3.1 Expression of HTL (SecY-SecE-SecG + YidC-SecD-SecF + YajC)

1. Transform the pACEMBL HTL3 plasmid (Fig. 1a) into chemically competent or electrocompetent *E. coli* BL21 Star (DE3) or *E. coli* C43λ(DE3) cells [3] and select for growth on LB agar plates with ampicillin, chloramphenicol, and kanamycin at 37 °C overnight.
2. Inoculate 100 mL preculture in LB medium supplemented with 50 μg/mL ampicillin, 20 μg/mL chloramphenicol, and 30 μg/mL kanamycin with one colony from the LB agar plate and grow at 37 °C overnight by shaking at 180 rpm.
3. Inoculate 1 L of *Terrific Broth* with 50 μg/mL ampicillin, 20 μg/mL chloramphenicol, and 30 μg/mL kanamycin in a 5 L baffled Erlenmeyer flask with approximately 20 mL preculture to reach an initial optical density (OD_{600nm}) of 0.05 at 600 nm. Shake the *E. coli* cultures at 37 °C and 150 rpm. Typical culture volumes for holo-translocon preparations are 6–12 L.
4. When the cell cultures reach an OD_{600} of 1.2–1.5 (mid-log phase), induce protein expression by the addition of 21 mL of 10% L-arabinose (0.2% final concentration) and 0.5 mL of 1 M IPTG.
5. After 3 h of protein expression (37 °C, 150 rpm), the cells are harvested by centrifugation for 20 min at 5000 × *g* (4 °C) (*see* **Note 1**).
6. Discard the supernatant and determine the weight of the cell pellet (*see* **Note 2**).

3.2 Membrane Preparation

1. Thaw ~10 g *E. coli* HTL3 cell pellet on ice.
2. Add 30 mL of HSGM buffer and one tablet EDTA-free protease inhibitor cocktail and 0.5 mM PMSF.
3. Resuspend the cells by pipetting.
4. Open the cells by two passages through a French press cell at 1000 psi (*see* **Note 3**).
5. Centrifuge the cell lysate at 16,000 × *g* for 30 min (at 4 °C).
6. Transfer the supernatant into an ultracentrifugation tube and discard the pellet.
7. Centrifuge the cleared lysate for 2 h in a Ti70 rotor at 45,000 rpm (149,000 × *g*) at 4 °C.

8. Discard the supernatant and resuspend the brown-yellow pellet (membranes) in ~5 mL HSGM buffer using a Dounce homogenizer. Add cOmplete EDTA-free protease inhibitor cocktail and 0.5 mM PMSF (*see* **Note 4**).
9. Analyze the membrane fraction using a 15% acrylamide gel to confirm the presence of SecD (67.5 kDa), YidC (63 kDa), SecY (50 kDa), SecF (35 kDa), SecE (15 kDa), SecG (14 kDa). The presence of YajC-CBP (14.7 kDa) needs to be confirmed by Western blotting using an anti-CBP antibody (*see* Fig. 1c, lane 1). If the Coomassie stained gel is difficult to interpret, the presence of SecD, YidC, and SecE can be confirmed by Western blotting using an anti-His antibody. SecD and SecE are encoded by polycistrons and part of SecYEG and SecDF protein complexes. Therefore, the other complex subunits are likely to be also present in the membrane.

3.3 Detergent Solubilization

1. Add DDM to a final concentration of 1.5% w/v to the membranes (*see* **Note 5**).
2. Gently agitate the sample at 4 °C (in the cold room) for 2 h by *head-over-tail rotation*.
3. Centrifuge solubilized membranes for 1 h in a Ti70 rotor at 45,000 rpm (149,000 × *g*) at 4 °C (*see* **Note 6**).

3.4 NiNTA Affinity Chromatography (All Steps at 4 °C)

1. Equilibrate 5 mL Ni-NTA resin with ten column volumes of Ni-NTA buffer (*see* **Note 7** for using a prepacked column).
2. Add the Ni-NTA resin to the solubilized membrane supernatant (from **step 3** of Subheading 3.3).
3. Gently agitate for 1 h by *head-over-tail rotation*.
4. Transfer the sample and beads into a gravity flow column. Collect the flow-through.
5. Wash with at least 20 column volumes of Ni-NTA buffer (*see* **Note 8**, to change the detergent to LMNG at this step).
6. Wash with three to five column volumes of high salt buffer.
7. Wash with at least five column volumes of Ni-NTA buffer.
8. Wash with 20 column volumes of Ni-NTA wash buffer.
9. Verify that there is no A_{280nm} signal in the wash fraction using a NanoDrop spectrophotometer before proceeding.
10. Elute with Ni-NTA elution buffer and collect the eluate in 1.5–2 mL tubes.
11. Determine the protein concentration of each fraction by A_{280nm} measurements using a NanoDrop spectrophotometer (*see* **Note 9**).
12. Add 0.1 mM PMSF and cOmplete protease inhibitor cocktail to the protein-containing fractions.

13. Analyze the input, flow-through, wash, and eluate fractions by SDS PAGE using a 15% acrylamide gel (*see* Fig. 1c for a typical result).
14. Pool the protein-containing fractions.

3.5 Desalting Step (at 4 °C)

1. Equilibrate a 5 mL desalting column with CBP-binding buffer using the AKTA system (*see* **Notes 10** and **11**).
2. Inject 3 mL of Ni-NTA eluate (use combined protein-containing fractions from **step 14** of Subheading 3.4) onto the desalting column (*see* **Note 12**).
3. Collect 1 mL fractions and pool protein-containing fractions. Repeat this procedure until all holo-translocon-containing eluate fractions from **step 14** of Subheading 3.4 are in CBP-binding buffer.

3.6 Calmodulin Affinity Chromatography (at 4 °C)

1. Equilibrate 5 mL calmodulin affinity resin with CBP-binding buffer.
2. Add the pooled fractions from **step 3** of Subheading 3.5 to the resin.
3. Incubate overnight at 4 °C and gently agitate.
4. Transfer the sample into a gravity flow column.
5. Wash the column with at least ten column volumes of CBP washing buffer.
6. Verify by NanoDrop measurements that there is no A_{280nm} signal in the last wash fraction before proceeding to the next step.
7. Elute with five column volumes of CBP elution buffer and collect fractions.
8. Determine the concentration of each fraction by NanoDrop A_{280nm} measurements (*see* **Note 9**).
9. Analyze the input, flow-through, wash, and eluate fractions by SDS PAGE using a 15% acrylamide gel.
10. Pool holo-translocon-containing fractions.

3.7 Sample Concentration

1. Equilibrate a 100 kDa molecular weight cut-off concentrator with S6 Buffer.
2. Load the holo-translocon-containing fractions from **step 10** of Subheading 3.6 into the concentrator.
3. Centrifuge at 3500 × *g*, 4 °C until the sample volume is reduced to 500 μL (monitor volume regularly).

3.8 Size Exclusion Chromatography (Optional)

1. Equilibrate a Superose 6 column with water and then with two column volumes of S6 Buffer (*see* **Note 13**).
2. Load 500 µL of sample from **step 3** of Subheading 3.7 onto the column.
3. Collect 0.5 mL fractions.
4. Analyze the protein-containing fractions on a 15% SDS PAGE gel.

Pool the peak fractions (Fig. 1d), if required concentrate to ~1 mg/mL using a concentrator with 100 kDa molecular weight cut-off. Freeze aliquots in liquid nitrogen and store the purified holo-translocon at −80 °C.

4 Notes

1. For each membrane protein complex, the growth temperature and the optimal time point to harvest the cells have to be determined experimentally. It is recommended to also test different strains (Fig. 1b).
2. It is possible to freeze the cells in liquid nitrogen and store the cells at −80 °C at this step.
3. Alternatively, two passages through a microfluidizer or cell disrupter at 25 kPsi can be used to open the cells.
4. It is possible to freeze the dissolved membranes in liquid nitrogen and store at −80 °C after this step.
5. For each new membrane protein complex, detergent screens have to be tested to identify a suitable mild, non-denaturing, solubilizing detergent [26], ideally followed by an activity assay (Fig. 1b). Different screens have been reported [27, 28], various commercial detergent screens are available (e.g., Qiagen, Jena Bioscience).
6. If the centrifuge tube is not completely filled, it is recommended to either fill the Ti70 centrifugation tubes to the top with buffer or to reduce the centrifugation speed.
7. Alternatively, use a HisTrap FF column (GE Healthcare) coupled to a peristaltic pump for loading and washing the sample. Prior to loading, the supernatant from ultracentrifugation is passed through a 0.22 µM filter to further remove any large particles. During the washing step, the color of the column will change from brown to almost light blue over the course of about 1 h. During this time, detergent exchange (to LMNG for example) or supplementation of any additive (e.g., cholesteryl hemisuccinate) can be performed. For elution, the column is connected to an AKTA chromatography system.

8. It is possible to exchange the detergent to Lauryl Maltose Neopentyl Glycol (LMNG, Anatrace) at this step. To do this, replace DDM in the Ni-NTA Wash buffer with 0.1% w/v LMNG and in all other buffers with 0.03% w/v LMNG. In the final size exclusion chromatography step, the concentration of LMNG in the S6 buffer can be further lowered to 0.002% w/v LMNG. The critical micelle concentration of LMNG is 0.001% w/v in water.
9. The molecular weight of holo-translocon is 240 kDa and the extinction coefficient is ~500,000 M^{-1} cm^{-1} at 280 nm.
10. Equilibrate the desalting column during the Ni-NTA wash steps.
11. The maximum pressure of the desalting column is 0.3 MPa.
12. Very important: Quickly proceed from **step 14** of Subheading 3.4 to the desalting step to avoid aggregation of the holo-translocon sample in the Ni-NTA elution buffer due to high imidazole concentration.
13. Superose 6 maximum pressure is 1.5 MPa; recommended flow rate is 0.4 mL/min.

Acknowledgments

This work is supported through a BBSRC Responsive Mode Grant (BB/P000940/1) and a Wellcome Trust Investigator Grant (210701/Z/18/Z). We would like to thank Kyle Powers, Flint Stevenson-Jones, and Josh Bufton for critically reading the manuscript.

References

1. Babu M, Vlasblom J, Pu S et al (2012) Interaction landscape of membrane-protein complexes *in Saccharomyces cerevisiae*. Nature 489:585–589
2. Hattab G, Warschawski DE, Moncoq K et al (2015) *Escherichia coli* as host for membrane protein structure determination: a global analysis. Sci Rep 5:12097
3. Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli:* mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol 260:289–298
4. Wagner S, Klepsch MM, Schlegel S et al (2008) Tuning *Escherichia coli* for membrane protein overexpression. Proc Natl Acad Sci U S A 105:14371–14376
5. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207–234
6. Zorman S, Botte M, Jiang Q et al (2015) Advances and challenges of membrane-protein complex production. Curr Opin Struct Biol 32:123–130
7. Li GW, Burkhardt D, Gross C et al (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157:624–635
8. Smolke CD, Keasling JD (2002) Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. Biotechnol Bioeng 80:762–776

9. Drew D, Lerch M, Kunji E et al (2006) Optimization of membrane protein overexpression and purification using GFP fusions. Nat Methods 3:303–313
10. Kawate T, Gouaux E (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14:673–681
11. Duong F, Wickner W (1997) Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. EMBO J 16:2756–2768
12. Scotti PA, Urbanus ML, Brunner J et al (2000) YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. EMBO J 19:542–549
13. Bieniossek C, Nie Y, Frey D et al (2009) Automated unrestricted multigene recombineering for multiprotein complex production. Nat Methods 6:447–450
14. Komar J, Botte M, Collinson I et al (2015) ACEMBLing a multiprotein transmembrane complex: the functional SecYEG-SecDF-YajC-YidC Holotranslocon protein secretase/insertase. Methods Enzymol 556:23–49
15. Schulze RJ, Komar J, Botte M et al (2014) Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. Proc Natl Acad Sci U S A 111:4844–4849
16. Botte M, Zaccai NR, Lycklama JA et al (2016) A central cavity within the holo-translocon suggests a mechanism for membrane protein insertion. Sci Rep 6:38399
17. Thomas JA, Tate CG (2014) Quality control in eukaryotic membrane protein overproduction. J Mol Biol 426:4139–4154
18. Kost TA, Condreay JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nat Biotechnol 23:567–575
19. Akermoun M, Koglin M, Zvalova-Iooss D et al (2005) Characterization of 16 human G protein-coupled receptors expressed in baculovirus-infected insect cells. Protein Expr Purif 44:65–74
20. Rasmussen SG, Choi HJ, Rosenbaum DM et al (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature 450:383–387
21. Fitzgerald DJ, Berger P, Schaffitzel C et al (2006) Protein complex expression by using multigene baculoviral vectors. Nat Methods 3:1021–1032
22. Khan I, Krishnaswamy S, Sabale M et al (2018) Efficient production of a mature and functional gamma secretase protease. Sci Rep 8:12834
23. Mesa P, Deniaud A, Montoya G et al (2013) Directly from the source: endogenous preparations of molecular machines. Curr Opin Struct Biol 23:319–325
24. Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene 96:23–28
25. Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16:6127–6145
26. Linke D (2014) Explanatory chapter: choosing the right detergent. Methods Enzymol 541:141–148
27. Kubicek J, Block H, Maertens B et al (2014) Expression and purification of membrane proteins. Methods Enzymol 541:117–140
28. Lantez V, Nikolaidis I, Rechenmann M et al (2015) Rapid automated detergent screening for the solubilization and purification of membrane proteins and complexes. Eng Life Sci 15:39–50

Chapter 2

Production of Multiprotein Complexes Using the Baculovirus Expression System: Homology-Based and Restriction-Free Cloning Strategies for Construct Design

Paola Rossolillo, Olga Kolesnikova, Karim Essabri, Gala Ramon Zamorano, and Arnaud Poterszman

Abstract

Most cellular processes are mediated by multi-subunit protein complexes which have attracted major interest in both academia and industry. Recombinant production of such entities in quantity and quality sufficient for functional and structural investigations may be extremely challenging and necessitate specific technologies. The baculovirus expression vector system is widely used for the production of eukaryotic multiprotein complexes, and a variety of strategies are available to assemble transfer vectors for the generation of recombinant baculoviruses. Here we detail applications of homology-based cloning techniques for one-step construction of dual promoter baculovirus transfer plasmids and of restriction-free (RF) cloning for the modification of existing constructs.

Key words Baculovirus insect cell expression, Multiprotein complex, Homology-based cloning, Sequence and ligation–independent cloning (SLIC), restriction-free (RF) cloning

1 Introduction

Most processes in living systems are mediated by multi-subunit protein complexes which have attracted major interest in both academia and industry. Recombinant production of such entities in quantity and quality sufficient for functional and structural investigations requires specific technologies. In favorable cases, components of a complex can be expressed separately and assembled in vitro. This approach is unfortunately not always applicable as isolated components of a complex are often poorly soluble or misfolded in the absence of interacting partners. Complex samples, including self-assembling multi-protein complexes, proteins requiring specific interacting partners, or post-translational modifications

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_2,

frequently require the simultaneous expression of several proteins foreign to the host cell.

With the baculovirus insect cell expression system, protein targets can be expressed from multiple viruses carrying a single foreign gene each, from a single baculovirus carrying multiple foreign genes, or from a combination of the two approaches. For routine production of complex systems, it is now widely accepted that the use of multigene expression constructs is extremely efficient. Progress in our understanding of the baculovirus biology together with the development of new tools for DNA manipulation have allowed to streamline protocols for the generation of viruses, expression screening, and large-scale productions (*see* refs. 1, 2 for recent reviews). A major breakthrough was accomplished through the development of the Multibac technology, which directs co-expression of multiple genes from a single virus under the control of multiple copies of the late p10 promoter and the very late polyhedron (PH) promoter [3, 4].

Preparation of multigene expression constructs is a multistep process with a first round consisting in the assembly of single or dual expression cassettes. In the absence of prior knowledge about the proteins of interest, experiments with viruses containing the single or dual expression cassettes will provide valuable information on expression level/solubility of individual subunits; moreover, different affinity tags can be tested. At this stage, coinfection experiments deliver first insights into the protein–protein interaction network and the architecture of the complex.

In the second step, expression cassettes are combined to generate multigene transfer vectors and produce the corresponding baculoviruses (Fig 1a). This can be achieved by using a combination of restriction-ligation and in vitro Cre-mediated plasmid fusion [5–7] as well as with restriction-free technologies such as In-Fusion [8] or USER (Uracil-Specific Excision Reagent) cloning [9] (Fig. 1b).

We have previously described optimized protocols to generate, evaluate, and amplify recombinant viruses, as well as to perform expression screening and large-scale production of difficult-to-express proteins and complexes [10, 11]. In this chapter, we focus on the assembly of multigene expression constructs and detail applications of homology-based cloning techniques (sequence and ligation–independent cloning (SLIC) and similar commercially available technologies such as In-Fusion, Gibson, or NEBuilder) to assemble pairs of promoters to create dual expression plasmids. We also provide a simple protocol for the modification of existing constructs based on restriction-free (RF) cloning approach. These protocols were implemented in our laboratory to generate constructs for the reconstitution and characterization of human multiprotein complexes that include the CAK and pTefb kinase complexes as well as the ten subunit transcription/DNA repair factor TFIIH (*see* for example [12–14]).

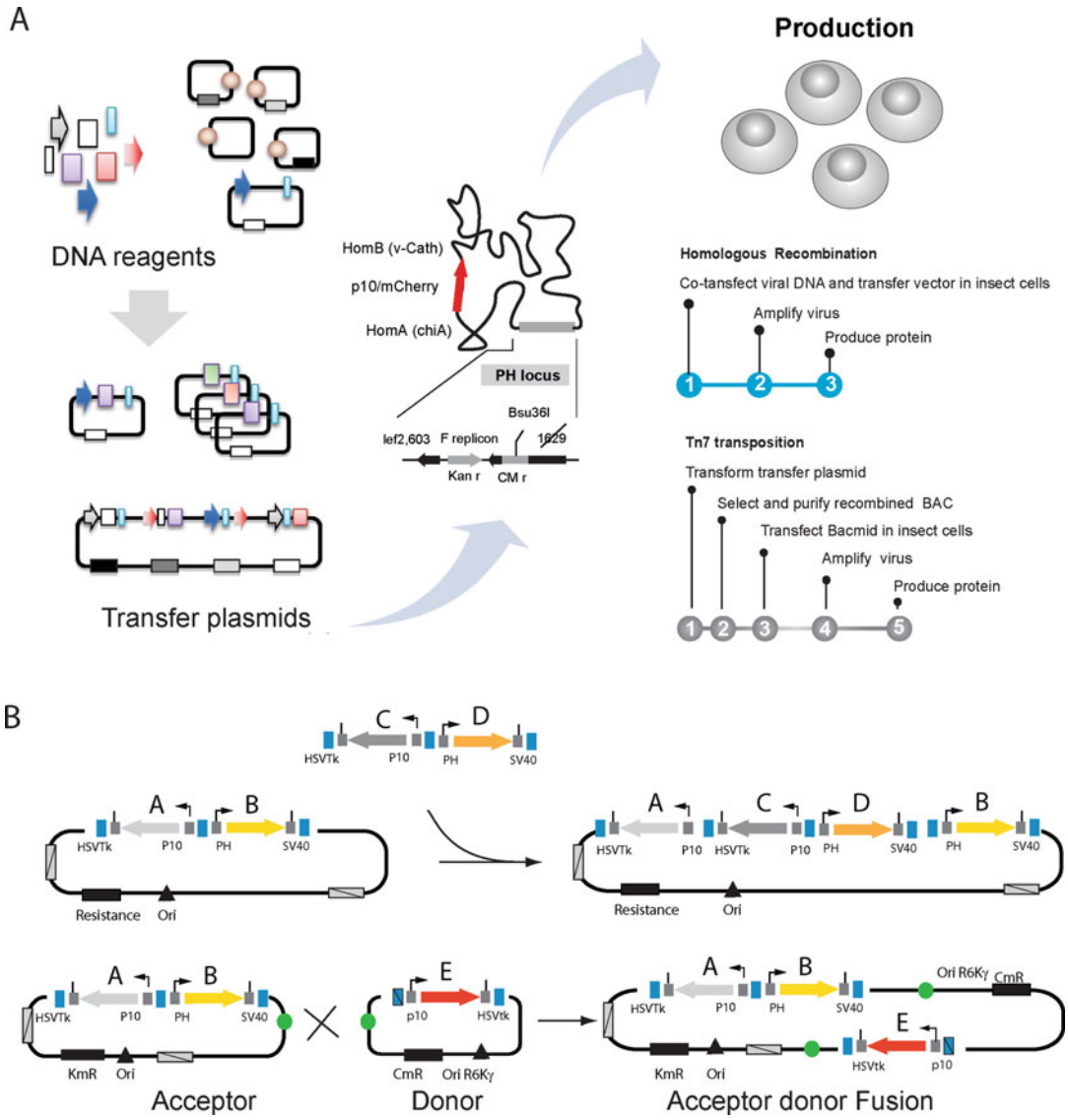


Fig. 1 Construction of multigene transfer vectors for the expression of multiprotein complexes. (**a**) DNA elements comprising genes of interest, promoters, and terminators are combined with plasmid backbone modules into multigene expression transfer vectors (left). These plasmids are transferred into the viral genome through Tn7 transposition in bacterial BH10Bac cells or by homologous recombination after co-transfection in insect cells (right). Recombinant viruses produce proteins of interest (here under the control of the PH promoter) together with a reporter gene (here mCherry under the P10 promoter). (**b**) To assemble multigene constructs, individual or dual expression cassettes can be excised by digestion with a pair of restriction endonucleases or amplified by PCR and inserted via compatible restriction sites or homology-based cloning into the multiplication module of a progenitor plasmid. The top panel shows an example in which an expression cassette containing genes C and D is cloned into a plasmid containing genes A and B. Alternatively, multigene constructs can be assembled using Cre-mediated fusion via LoxP sites (green circles). Acceptors have a regular replication origin (ori ColE1), whereas donors have a conditional origin derived from phage (ori R6Kγ); the two plasmids also have different antibiotic resistance. The bottom panel illustrates the Cre-mediated fusion of an acceptor plasmid containing a dual expression cassette bearing genes A and B with a donor vector bearing gene E

2 Materials

2.1 Working Environment and Instruments

You need access to standard equipment for molecular biology including PCR machines, agarose gel electrophoresis apparatus, UV trans-illuminator, refrigerated benchtop centrifuge, nanodrop for DNA quantification, fridge, and freezers (−20 °C and −80 °C).

Prepare all solutions using ultrapure water (prepared by purifying deionized water, to attain a sensitivity of 18 MΩ-cm at 25 °C) and analytical grade reagents. Unless indicated otherwise, prepare and store all reagents at room temperature. Diligently follow all waste disposal regulations when disposing waste materials.

2.2 Bacterial Strains and Plasmids

1. Luria–Bertani (LB) medium with antibiotics added at the following concentrations: ampicillin (100 μg/mL), kanamycin (30 μg/mL), spectinomycin (90 μg/mL), and chloramphenicol (15 μg/mL).
2. *E. coli* DH5α, TOP10, XL-1 blue, or Pir1 (*see* **Note 1**) competent cells (chemically or electrocompetent cells).
3. Dual expression transfer vectors (*see* Table 1). These plasmids are available from Addgene.
4. Templates for amplification of the target genes. When a starting plasmid containing the cDNA of your gene of interest is not present, chemical synthesis is often an advantageous option for generating cDNAs. In any case, the removal of restriction sites that might be used for the transfer of the expression cassettes is

Table 1
Dual promoter plasmids for baculovirus expression systems

Plasmid	Method	Resistance	Size	Origin	References
pFastBac Dual	Tn7 transposition	Ampicillin, gentamicin	5.2 kbp	pUC ori	None
pFL	Tn7 transposition	Ampicillin, gentamicin	5.8 kbp	ColE1	[6]
pKL	Tn7 transposition	Kanamycin, gentamicin	4.9 kbp	pBR322	[5]
pAC8_MF	Homology recombination	Ampicillin	6.9 kbp	pUC ori	Kolesnikova et al. (in prep)
pUCDM[a]	Cre-mediated fusion	Chloramphenicol	3.0 kbp	R6Kγ	[5]
pSPL[a]	Cre-mediated fusion	Spectinomycin	2.9 kbp	R6Kγ	[6]

These plasmids contain the expression cassette detailed in Fig. 2a and with the exception of pFastBac Dual all possess a LoxP site. Vectors pFastBac Dual, pFL, and pKL use Tn7 transposition while pAC8_MF relies on homology recombination to integrate the expression cassette in the viral DNA

[a]Vectors pUCDM and pSPL cannot be used directly to generate recombinant viruses but can be fused to pFL, pKL, and pAC8_MF through using Cre-mediated reactions

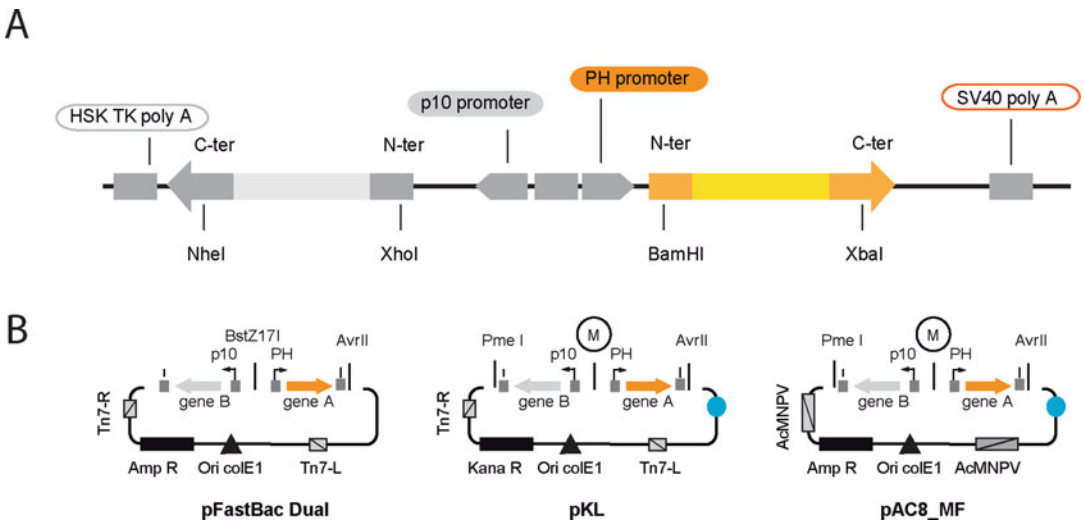


Fig. 2 Dual expression plasmids. (**a**) Expression cassette from the transfer vector pFastBac Dual, where promoters (p10 and PH), cloning sites to insert the genes of interest (A and B) and terminators (HSV Tk and SV40 polyA) are indicated. **(b)** Representation of the transfer vectors pFastBac Dual, pKL, and pAC8_MF. The expression cassettes of pKL and pAC8_MF are flanked by a pair of unique restriction sites (AvrII and PmeI) and contain a multiplication module (M) composed of the BstZ17I, SpeI, or NruI recognition sequences. These can be used to assemble constructs for the co-expression of three and more cDNAs. Plasmids pKL and pAC8_MF also contain a LoxP site (blue circle). Elements for the transfer into the viral genome (Tn7-R and Tn7-L or AcMNPV sequences), a replication unit (ColE1) and antibiotic resistance genes (Amp^R and Km^R) are also indicated

strongly suggested. Try in particular to eliminate PmeI, AvrII, BstZ17I, NruI, and SpeI recognition sequences from your starting plasmids or cDNAs as these sites are essentials for the creation of multigene vectors (Fig. 2).

2.3 Reagents for the Preparation of the Linearized Vector and DNA Fragments

1. Plasmid containing the target cDNAs and dual expression transfer vectors (*see* Subheading 2.2).
2. Synthetic oligonucleotides for PCR amplification designed as detailed in Subheading 3.2 (*see* Table 2). Dry oligonucleotides are usually resuspended at 100 μM and further diluted at 10 μM.
3. Phusion Polymerase (ThermoScientific™) with 5xHF or GC buffer and DMSO or another high-fidelity polymerase.
4. dNTPs mix. Prepare and aliquot 10 mM dNTP stock solutions in ultrapure water.
5. DpnI (10 u/μL) restriction enzyme and corresponding buffer.
6. If the plasmid backbone is prepared by digestion: BamHI, XbaI, XhoI, NheI (10 u/μL) restriction enzymes, and the corresponding buffers.

Table 2
Oligonucleotides for the amplification of the dual promoter region, the plasmid backbone, and the two target cDNAs and their cloning

Region to amplify	Primer name	Primer sequence
Native constructs (pFastBac Dual, pKL. or pAC8_MF)		
Dual promoter module	Prom-F (PH)	GGTGGATCCGCGCCCGATGG
	Prom-R (p10)	GGTGGCTCGAGATCCCGGGTG
Plasmid backbone	Backbone-R (PH)	TCTAGAGCCTGCAGTCTCG
	Backbone-F (p10)	GCTAGCAGCTGATGCATAG
Gene under pH promoter	pH-Gene-START-F	*CATCGGGCGCGGATCCACC/*ATG. . .first 15–21 nt of the gene (BamHI+Kozak)
	pH-Gene-STOP-R	*CGAGACTGCAGGCTCTAGA/*STOP. . .RC(last 15–21 nt of the gene) (XbaI)
Gene under p10 promoter	p10-Gene-START-F	*CACCCGGGATCTCGAGCCACC/*ATG. . .first 15–21 nt of the gene (XhoI+Kozak)
	p10-Gene-STOP-R	*CTATGCATCAGCTGCTAGC/*STOP. . .RC(last 15–21 nt of the gene) (NheI)
C-terminal histidine-tagged construct (pKL-PH—3C-10His-Cter)		
Plasmid backbone	Backbone-3C-R (PH)	*TCTAGACTGGAAGTTCTGTTTC*
Gene under pH	pH-Gene-10His-3C R	*GAAACAGAACTTCCAGTCTAGA*. . .RC(last 15–21 nt of the gene) (XbaI)

The NheI, XhoI, BamHI, and XbaI restriction sites are underlined. Note that oligonucleotides Prom-F and Prom-R contain a kozak consensus sequence (**CCACC** in bold). pKL-PH-3C-10His-Cter is a pKL derivative designed to co-express a cleavable C-terminally histidine-tagged protein under the control of the PH promoter with an untagged protein under the control of p10. For this vector, the gene under PH promoter is amplified using the pH-Gene-START-F/pH-Gene-10His-3C-R primer pair and the plasmid backbone with the Backbone-3C-R (PH)/Backbone-F (p10) pair. The gene under p10 promoter is amplified with the p10-Gene-START-F/p10-Gene-STOP-R pair. RC stands for reverse and complement (*see* Figs. 4 and 5)

2.4 Reagents for a SLIC Reaction

1. T4 DNA Polymerase (New England Biolabs).
2. 10× NEB 2.1 buffer (Tris-HCl 100 mM, $MgCl_2$ 100 mM, NaCl 500 mM, BSA 1 mg/mL, pH 7.9).
3. 10 mM dCTP.
4. 10× Ligation buffer (e.g., Tris-HCl 500 mM, $MgCl_2$ 100 mM, DDT 100 mM, pH 7.5) or any DNA ligase buffer for annealing.
5. 1 μg of linearized vector and of each amplified DNA fragment to be assembled prepared as described in Subheading 3.3.

2.5 Reagents for an In-Fusion Reaction

1. In-Fusion or Gibson assembly kit.
2. 1 μg of linearized vector and of each amplified DNA fragment to be assembled prepared as described in Subheading 3.3.

2.6 Solution and Media for Restriction-Free Cloning

1. Phusion Polymerase (ThermoScientific™) with 5xHF or GC buffer and DMSO or another high-fidelity polymerase.
2. dNTPs mix. Prepare and aliquot 10 mM dNTP stock solutions in ultrapure water.

   DpnI (10 u/μL) restriction enzyme and corresponding buffer.

3 Methods

Here we detail the preparation of dual expression plasmids using homology-based cloning and the modification of existing constructs. We focus on an expression cassette composed of two divergent baculovirus late promoters (PH and p10), each followed by a multiple cloning site (MSC1 and MSC2) and a transcription termination signal (SV40 pA and HSV-TK pA) (Fig. 2a). This cassette organization is found in the transfer vector FastBac Dual and in its Multibac derivatives which include the acceptor vectors pFL and pKL and the donor plasmids pUCDM and pSPL (Table 1). It was also inserted into the backbone of the transfer vector pBacPack8 (plasmid pAC8_MF) to enable the generation of recombinant baculoviruses by homologous recombination directly in insect cells (Fig. 2b). In a typical experiment, the plasmid backbone is combined with the promoter module (comprising the PH and p10 promoters) and cDNAs encoding the two target genes, in a four-fragment assembly reaction (Fig. 3). The first cDNA is typically inserted between the BamHI and XbaI sites of MCS1 and the second between the XhoI and NheI sites of MSC2 (Fig. 2a).

Following a general outline of homology-based technologies for fragment assembly in Subheading 3.1, we detail primer selection in Subheading 3.2 and provide protocols for the preparation

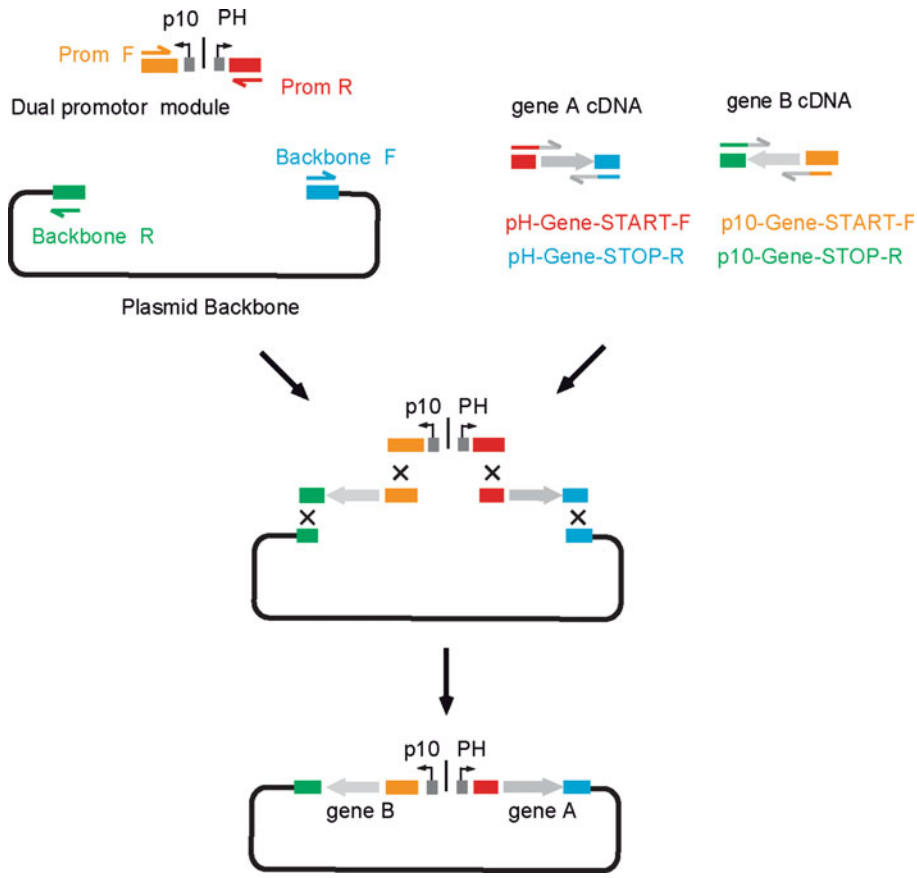


Fig. 3 Homology-based assembly of dual expression vectors. The plasmid backbone, the promoter module (comprising the PH and p10 promoters), and the cDNAs encoding the two target genes are amplified by PCR using primers designed to generate end-terminal homology and combined in a single-step four-fragment assembly reaction

and assembly of the linearized vector and DNA fragments in Subheadings 3.3–3.5. In Subheading 3.6, the use of restriction-free (RF) cloning for modification of existing constructs is illustrated.

3.1 Principles of Homology-Based Technologies for Fragment Assembly

Several seamless cloning technologies are now available to assemble multiple DNA fragments into a destination plasmid without introducing undesired nucleotides or scar sequences between vector and insert. Homology-based techniques—Gibson Assembly [15], In-Fusion Cloning (Clontech™, [16]), and SLIC [17]—rely on PCR to add homologous sequence overhangs to the DNA fragments to be assembled adjacently and on an exonuclease activity to expose single-stranded overhangs that will anneal with complementary overlapping fragments.

The Gibson assembly technique [15] relies on an enzyme mix composed of the T5 exonuclease, which chews back the 5′ ends of DNA fragments, generating long overhangs and allowing the

single-stranded regions with homology to anneal; a DNA polymerase, to fill in the gaps; and a DNA ligase, to seal the nicks between the annealed and filled-in sequences.

SLIC (sequence and ligation–independent cloning) [17] and In-Fusion Cloning (Clontech™, [16]) are based on the ligation-free cloning of overlapping PCR products treated by a DNA polymerase (the T4 and the vaccinia virus DNA polymerases, respectively), which possesses a 3′ to 5′ proofreading exonuclease activity. The polymerase chews back one strand of each fragment, creating single-stranded compatible ends. These overhangs are annealed in vitro, and the recombinant circular construct is rescued in *E. coli*.

3.2 Primer Selection for Homology-Based Cloning

The assembly of a dual expression transfer plasmid involves the combination of four DNA fragments with overlapping ends which are in general obtained by PCR amplification: the vector backbone, the dual promoter module, and the two cDNAs (Fig. 3). Four pairs of oligonucleotides are required.

(a) The region of the plasmid including the divergent promoters spanning from XhoI to BamHI sites is amplified with primers named *prom-F* and *prom-R*, both including the restriction site in their sequence and containing a Kozak consensus in their 5′end (*see* Table 2 for the primer sequences).

(b) The two genes are amplified by a PCR performed with the primer pairs *pH-gene-START-F/pH-gene-STOP-R* and *p10-gene-START-F/p10-gene-STOP-R*. They are composed of a 5′ part corresponding to the cloning junction followed by 3′ part complementary to the insert target (*see* **step 4** below).

(c) Finally, the vector backbone, from XbaI to NheI sites, is amplified with primers *Backbone-F* and *Backbone-R*. Note that, alternatively to PCR amplification, the linearized vector can be obtained by XbaI/NheI restriction digest followed by gel purification (*see* Subheading 3.3).

Construct design and selection of primers for PCR amplification are detailed using an example where a dual expression cassette for the co-expression of eGFP and mCherry is assembled (Fig. 4 and Table 2). We strongly advise using a DNA cloning software (i.e., Serial cloner, ApE, SnapGene) to define the desired constructs, select a set of specific primers, and carefully simulate the cloning procedure in silico before starting any experiment.

1. Identify plasmids/cDNAs from which the sequences encoding the proteins of interest will be amplified and carefully verify their nucleotide sequence. Do not hesitate to re-sequence if cDNAs were obtained from external sources.

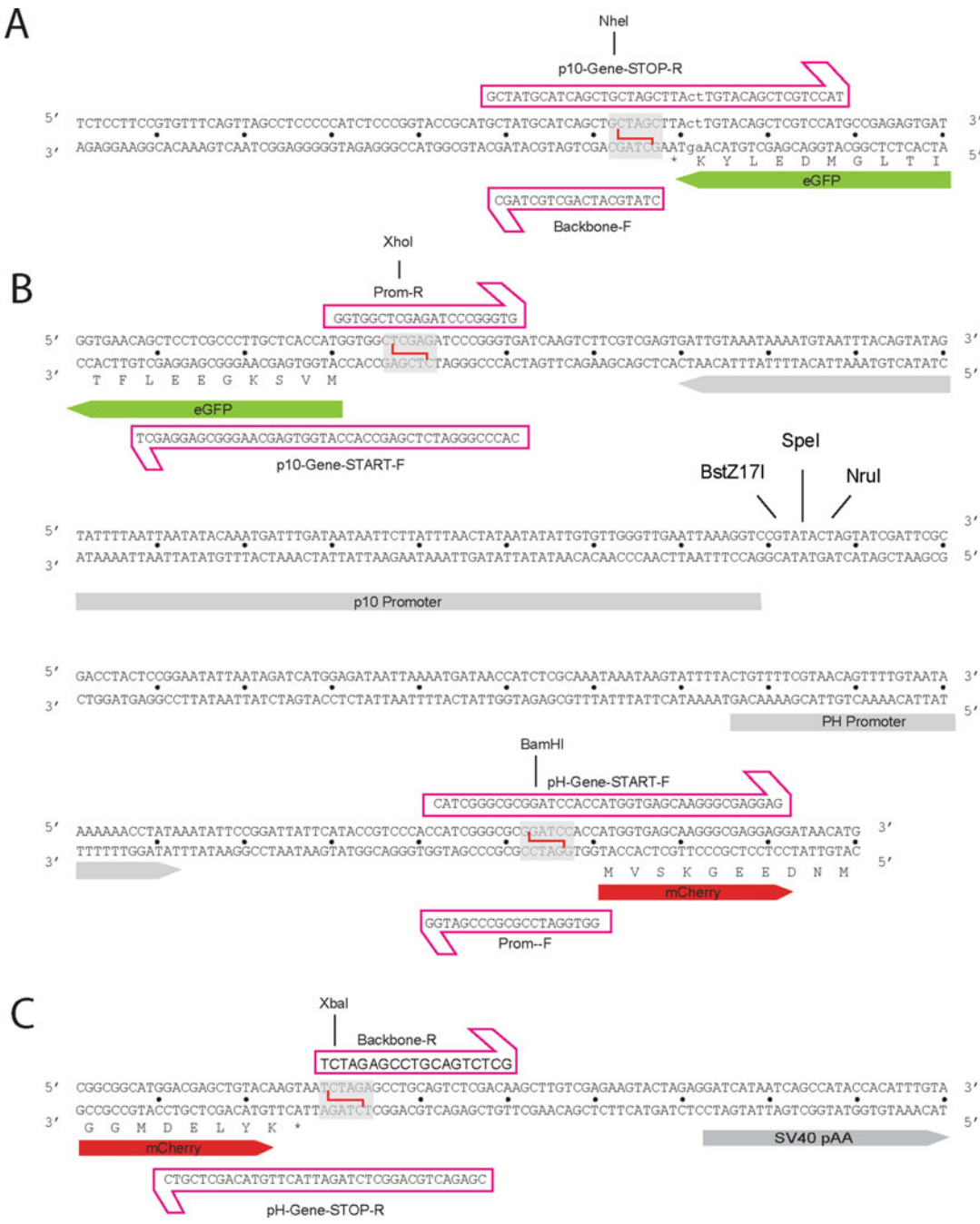


Fig. 4 Plasmid and primer sequences at junctions of a dual expression cassette for co-expression of eGFP and mCherry. The eGFP and mCherry cDNAs were cloned into pKL (plasmid available from Addgene # Plasmid #110741). PCR primers were designed to generate DNA fragments with 19- to 21-bp-long overhangs and maintain the NheI, XhoI, BamHI, and XbaI restriction sites. The 3′ region of the eGFP is detailed in panel (**a**), the dual promoter region with the 5′ region of the eGFP, the p10 and PH promoters, and the 5′ region of the mCherry cDNA in panel (**b**) and the 3′ region of the mCherry cDNA in panel (**c**). The plasmid backbone is amplified with the oligonucleotide pair *Backbone-R*/*Backbone-F* and the dual-promoter module with the

2. Decide on the constructs to be generated and select the transfer vector accordingly. Consider, for example, the presence and position of affinity tags.
3. Design a pair of generic primers to amplify the vector backbone (*Backbone-F* and *Backbone-R*) and another one for the dual promoter module (*Prom-F* and *Prom-R*). Whenever possible, oligonucleotides *Prom-F* and *Prom-R* contain a kozak consensus sequence (CCACCATGG, where ATG encodes the initiator Methionine) and are perfectly complementary to their vector template.
4. Design pairs of specific primers to amplify the target cDNAs (*pH-gene-START-F*, *pH-gene-STOP-R*, *p10-gene-START-F* and *p10-gene-STOP-R*). They should be composed of a 5′ part corresponding to the cloning junction (which provides a homology to the vector sequence for the homology-based cloning reaction) followed by a 3′ part complementary to the insert target (which provides a priming region for the PCR). The START primers include the first 15–21 nt of the cDNA, preceded by a sequence complementary to that of the oligonucleotide used to amplify the promoter region. The STOP primers include the last 15–21 nt of the gene, followed by a STOP codon and the reverse complement of the primers used for the backbone amplification.
5. Simulate the assembly of the expected transfer plasmids in silico and carefully check that the tags, if any, are inserted in phase with the genes of interest.
6. Check the propensity of the primers to form secondary structures or dimerize. Structures with a ΔG greater than −6 kcal/mole suggest that redesign may be required, particularly for those at the 3′ end, for example, by increasing the length of the 5′-tail.
7. Order the oligonucleotides.

By default, oligonucleotides are chosen to generate PCR fragments with 20 nucleotide-overlapping ends. They are usually designed taking into account the length and base composition of each primer pair, so that the Tm of the primers/templates are well matched and close to 60 °C (*see* **Note 2**). To minimize the formation of secondary structures, instead of using long oligonucleotides (e.g., to append an affinity or detection tag), do not hesitate to use nested PCRs.

Fig. 4 (continued) Prom-R/*Prom-F* pair. The eGFP and mCherry cDNAs are amplified with the *p10-gene-START-F*/*p10-gene-STOP-R* and the *pH-gene-START-F*/*pH-gene-STOP-R*. Note that oligonucleotides for the amplification of the "promoter" cassette and of the 5′ end of the cDNAs are designed to contain a Kozak consensus CCACC**ATG**G sequence where ATG (in bold) encodes the initiator codon (*see* Table 2)

If an affinity or detection tag needs to be fused to one of the target genes, the tag coding sequence can be added to the gene with nested PCRs, or a modified transfer plasmid whose multiple cloning site already contains the tag sequence can be used. The plasmid and primer sequences at junctions of a dual expression cassette for co-expression of eGFP and mCherry fused to a C-terminal histidine tag are shown Fig. 5.

3.3 Preparation of the Linearized Vector and of the DNA Fragments

The linearized vector and the DNA fragments to be assembled are prepared by PCR, except when the expression cassette is to be cloned into large transfer vectors such as pAC8_pMF derivatives.

1. Set up 50 μL PCR to amplify the promoter and backbone regions of the plasmid and the two genes (Table 3). Use 1–50 ng of template DNA if you amplify your fragments from a plasmid. Use 100–200 ng of template if you amplify your genes from cDNA.
2. Run the PCR in a thermocycler by using the conditions suggested by the polymerase manufacturer. We typically denature the template DNA at 98 °C for 30 s, perform 25 cycles of amplification (denaturation at 98 °C for 10 s, annealing at the calculated Tm for 15 s, and elongation at 72 °C for 30 s in the case of a 1 kb gene) followed by a final incubation at 72 °C for 5 min.
3. Load 1/10 of the PCR volume on a 1% agarose gel in parallel with a molecular weight marker to verify that the PCR amplification resulted in a single band of the expected size.
4. If the fragments are amplified from a plasmid, add ten units of the DpnI restriction enzyme to the PCR mix and incubate at 37 °C for 1 h. DpnI will digest the (methylated) plasmid template.
5. Purify the PCR on PCR clean-up columns following the instruction manual, elute bound DNA in 30 μL, and determine the concentration of the extracted DNA spectrophotometrically.

We usually prepare between 1 and 2 μg of each fragment. Do not hesitate to optimize PCR conditions (polymerase buffer, annealing temperature, addition of DMSO,) or redesign PCR primers if yield or PCR specificity are not sufficient (*see* **Note 3**).

If DNA fragments are to be cloned into large transfer vectors such as pAC8_pMF derivatives, the backbone DNA is prepared by XbaI/NheI restriction and gel purification:

1. Digest 3–5 μg of plasmid in a total volume of 30 μL for 3 h at 37 °C with ten units of NheI-HF and XbaI in CutSmart Buffer (New England Biolabs).

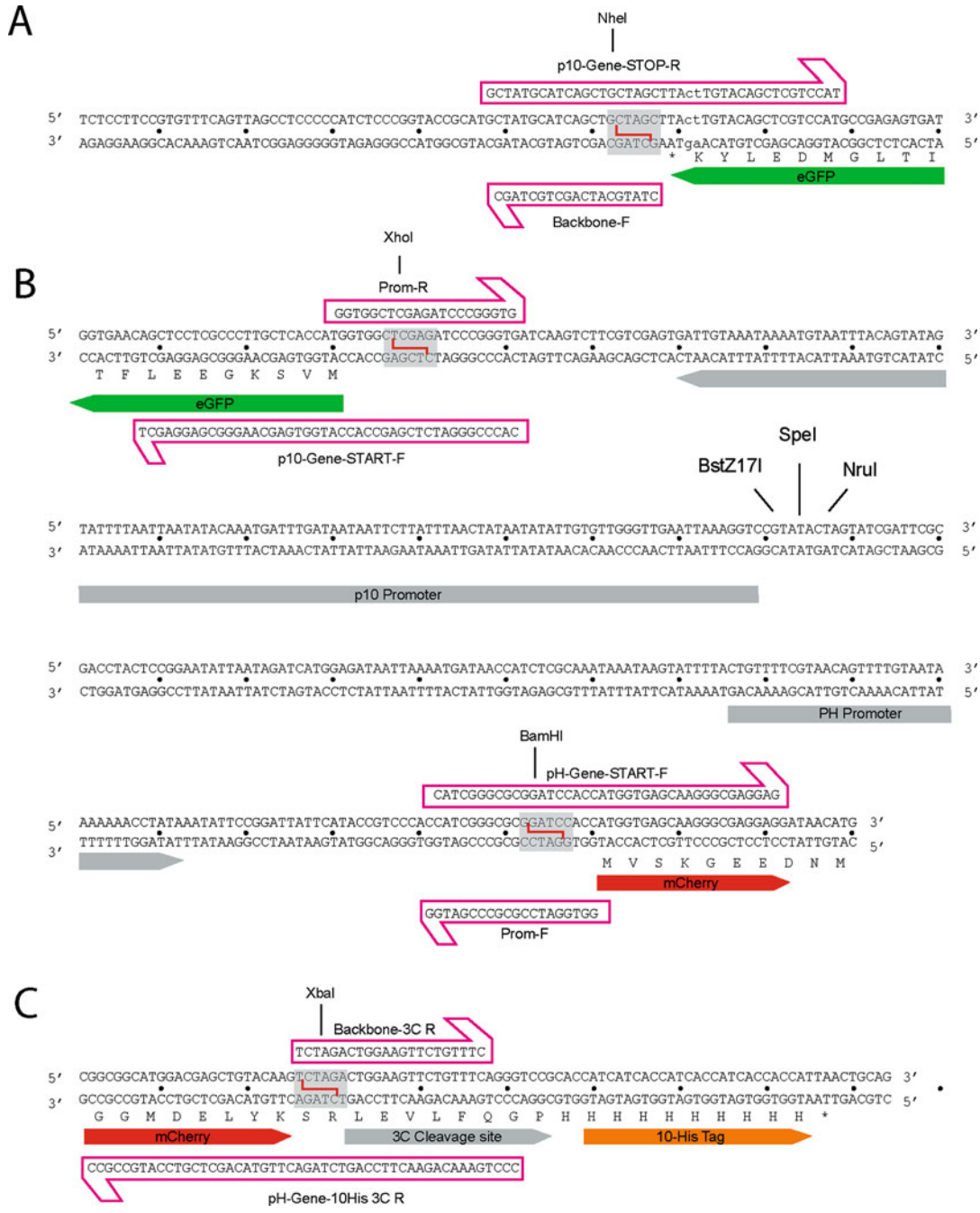


Fig. 5 Plasmid and primer sequences at junctions of a dual expression cassette for co-expression of eGFP and mCherry fused to a C-terminal histidine tag. The eGFP and mCherry cDNAs were cloned into the pKL derivative pKL-PH-3C-10His-Cter. PCR primers were designed to generate DNA fragments with 19- to 21-bp-long overhangs and maintain the NheI, XhoI, BamHI, and XbaI restriction sites. The 3′ region of the eGFP cDNA is detailed in panel (**a**), the dual promoter region with the 5′ region of the eGFP cDNA, the p10 and PH promoters, and the 5′ region of the mCherry cDNA in panel (**b**), and the 3′ region of the mCherry cDNA in panel (**c**). The plasmid backbone is amplified with the oligonucleotide pair *Backbone-3C-R*/*Backbone-F* and the dual-promoter module with the *Prom-R*/*Prom-F* pair. The eGFP and mCherry cDNAs are amplified with the *p10-gene-START-F*/*p10-gene-STOP-R* and the *pH-gene-START-F*/*pH-gene-10His-3C-R*. The amino acid sequence recognized by the prescission protease (3C) is LEVLFQ!GP (*see* Table 2)

Table 3
Composition of the PCR mixes for amplification of the plasmid backbone, of the dual-promoter module, and of the two cDNAs

	Plasmid backbone	Promoter module	cDNA 1 under pH promoter	cDNA 2 under p10 promoter
PCR pre-mix	25 μL	25 μL	25 μL	25 μL
Plasmid template	10–50 ng	10–50 ng	10–50 ng	10–50 ng
cDNA 1			10–50 ng of plasmid or 100–200 ng of total cDNA	
cDNA 2				10–50 ng of plasmid or 100–200 ng of total cDNA
Backbone-F Backbone-R	20 pmoles 20 pmoles			
Prom-F Prom-R		20 pmoles 20 pmoles		
pH-gene-START-F pH-gene-STOP-R			20 pmoles 20 pmoles	
p10-gene-START-F p10-gene-STOP-R				20 pmoles 20 pmoles
H_2O	qsp 50 μL	qsp 50 μL	qsp 50 μL	qsp 50 μL

For a 50 μL reaction, we typically mix 25 μL of PCR pre-mix, 3 μL of plasmid template at 10 ng/μL, and 2 μL of forward and 2 μL of reverse primer at 10 μM each. The PCR pre-mix is prepared by mixing 10 μL of 5× polymerase buffer, 1 μL of 10 mM dNTP mix, one unit of Phusion DNA polymerase, and water to a 25 μL volume

2. Load 1/10 of the reaction volume on a 1% agarose gel with a molecular weight marker to verify that the digestion is complete and resulted in the expected products.
3. Treat with Calf Alkaline Phosphatase: add one unit of Calf Alkaline Phosphatase (New England Biolabs) per microgram of DNA in the restriction reaction. Increase the reaction volume to 50 μl and add 1/10 of the volume of CIAP buffer. Incubate at 37 °C for 1 h. The Alkaline Phosphatase will eliminate phosphate groups from DNA extremities and will prevent the re-circularization of the plasmid.
4. Purify the linearized plasmid using clean-up columns following the instruction manual, elute bound DNA in 30 μL, and determine the concentration of the extracted DNA spectrophotometrically.

3.4 SLIC Reaction

As mentioned above, different approaches can be used to assemble DNA fragments into a linearized vector containing overlapping ends. Sequence and ligation–independent cloning (SLIC), in contrast to most other homology-based cloning strategies, does not necessitate specific kits and only requires T4 DNA polymerase to be tested first.

1. Set up the T4 DNA polymerase treatment for the four PCR products in separate tubes. In each tube, mix 1 μg of DNA, 3 μL of 10× NEB 2.1 buffer, and 0.5 U of T4 DNA Polymerase in a final volume of 30 μL. Incubate the reaction for 30 min at room temperature (*see* **Note 4**).
2. Stop the exonuclease reaction by adding 3 μL of a 10 mM dCTP solution (1/10 of the reaction volume) and leave on ice. In the absence of dNTPs, the T4 DNA polymerase has a 3′ to 5′ exonuclease activity and "chews" the DNA leaving a single-stranded region. Addition of dCTP stops the exonuclease activity and shifts it to a polymerase activity at the first occurrence of this nucleotide.
3. Prepare an annealing mix containing the polymerase-treated vector backbone, the dual promoter unit, and the two cDNAs in a 1:2:2:2 ratio using 250 ng (0.15 pmoles) of a 3 kb vector and appropriate amount of inserts (*see* **Note 5**), in a total volume of 20 μL. Incubate the annealing mix at 37 °C for 30 min and leave on ice until transformation or store at −20 °C. A negative control where one of the insert is omitted can be included to evaluate background.
4. Use 10 μL of the annealing reaction to transform 100 μL of *E. coli* competent cells (do not forget to use Pir1 cells for plasmids with an R6Kγ replication origin). Plate the transformation mixture on LB agar plates containing the appropriate antibiotic and incubate overnight at 37 °C. If a control experiment lacking one of the fragments was included, significantly more colonies should be obtained in the presence of all inserts than in the negative control.
5. Purify plasmid DNA from 6 to 12 colonies for restriction analysis and sequencing. We typically first digest 500 ng of purified plasmid with BamHI/XbaI and XhoI/NheI (both in CutSmart Buffer) to control the presence of the desired inserts (*see* **Note 6**). In the second step, two positive clones are selected, and their expression cassettes are sequenced (see sequencing oligos in Table 4).

3.5 Gibson Assembly®, NEBuilder® HiFi DNA Assembly, or In-Fusion® HD Cloning

Since SLIC is a simple and cost-effective technology, it is usually tested by default in our laboratory. However, the assembly of some constructs may prove more challenging than others, and it does not always yield successful constructs through SLIC. In these cases, we try commercial technologies such as NEBuilder HiFi (NEB) or

Table 4
List of primers for colony PCR analysis and sequencing

Region to sequence	Primer name	Primer sequence
5′ end of cDNA 1 (pH)	PH-F	TAAAATGATAACCATCTCGC
3′ end of cDNA 1 (pH)	SV40 p(A)-R	TTTAAAGCAAGTAAAACCTC
5′ end of cDNA 2 (p10)	p10-F	CCCAACACAATATATTATAG
3′ end of cDNA 2 (p10)	HSV TK p(A)-R	CACCCGTGCGTTTTATTCTGTC

In-fusion HD (Takara) cloning. NEBuilder HiFi cloning kit uses the same technology as Gibson Assembly, although with higher efficiency and better performance especially on short (<250 bp) fragments.

These technologies are available as user-friendly cloning kits, which provide the reagents necessary for the assembly of two or more DNA fragments in a ligation-free reaction. However, preliminary work is still needed, as for the SLIC reaction, to produce DNA fragments with homologous sequence overhangs. A guide for primer design is described in Subheading 3.2, and the preparation of all DNA fragments, including linearization of the vector backbone, is described in Subheading 3.3.

Similar to the SLIC reaction, the purity of the fragments is critical for their successful assembly. PCR and restriction digestion products should be cleaned up or gel purified, if necessary, to remove undesired by-products (*see* **Note 7**) before proceeding to cloning.

1. On ice, set up a 20 μL reaction containing 1× master mix and all DNA fragments in the selected molar ratio. We typically work with 100 ng of vector and a molar ratio of 1:2:2:2, vector to inserts (*see* **Note 5**), when assembling the dual promoter unit and two cDNA fragments. A negative control reaction, in which one of the fragments is omitted, may be set up in parallel.
2. Incubate the reaction at 50 °C for 60 min. The samples can be kept on ice until transformation or stored at −20 °C.
3. Use 2–5 μL of the assembled product to transform 50 μL of *E. coli* competent cells (Pir1 cells for plasmids with an R6Kγ replication origin). Plate the transformation mixture on LB agar plates containing the appropriate antibiotic and incubate overnight at 37 °C. If a control experiment was included, significantly more colonies should be obtained in the presence of all inserts than in the negative control.

4. Purify plasmid DNA from 6 to 12 colonies for restriction analysis and sequencing. We typically first digest 500 ng of purified plasmid with BamHI/XbaI and XhoI/NheI (both in CutSmart Buffer) to control the presence of the desired inserts (*see* **Note 6**). In the second step, two positive clones are selected, and their expression cassettes are sequenced (Table 4).

3.6 Restriction-Free (RF) Cloning for the Modification of the Existing Constructs

Preparation of multiprotein complexes has its own challenges and specificities. Typically, constructs might need to be modified by mutation, truncation, or deletion of low complexity regions and the position or nature of affinity tags might have to be modified. In the example shown in Fig. 6, RF cloning is used to introduce a cleavable affinity tag at the 3′ extremity of a cDNA in a complex expression construct.

Restriction-free (RF) cloning, developed for the introduction of a foreign DNA into a plasmid at any predetermined position, is based on the overlap extension site-directed mutagenesis and popularized under the name QuickChange mutagenesis (Stratagene™) [18–20]. The sequence of interest is PCR-amplified using two hybrid primers designed with complementarity to the desired insert and the destination plasmid (Fig. 7a). The double-stranded PCR product is used as a mega-primer for the second amplification reaction (Fig. 7b). In this step, each of the DNA strands anneals to the destination vector at the predetermined position and is extended in a linear amplification reaction that leads to the formation of a double-stranded nicked plasmid. The parental DNA is then removed by DpnI treatment (Fig. 7c), and the newly synthesized plasmid, containing the DNA insert, is then transformed into *E. coli* cells, in which the nicked DNA is ligated by endogenous enzymatic activity.

1. Collect the plasmid/cDNA required for the first PCR and the construct that you wish to modify, carefully verify their sequences, and design the desired construct.
2. Design a pair of hybrid primers with a 5′ part complementary to the destination plasmid followed by a 3′ part complementary to the insert and compatible melting temperatures (*see* **Note 2**). Oligonucleotides typically have a 18–25 bp overlap (Tm# 50 °C) with the destination vector and 18–25 bp overlap (Tm# 50 °C) with the insert of interest.
3. Set up the first PCR (50 μL) to amplify the fragment to be introduced using high-fidelity polymerase according to the manufacturer's instructions (*see* **Note 8**). Use 1–50 ng of template DNA containing the sequence of interest if you amplify your fragment from a plasmid. Use 100–200 ng of template if you amplify your fragment from cDNA.

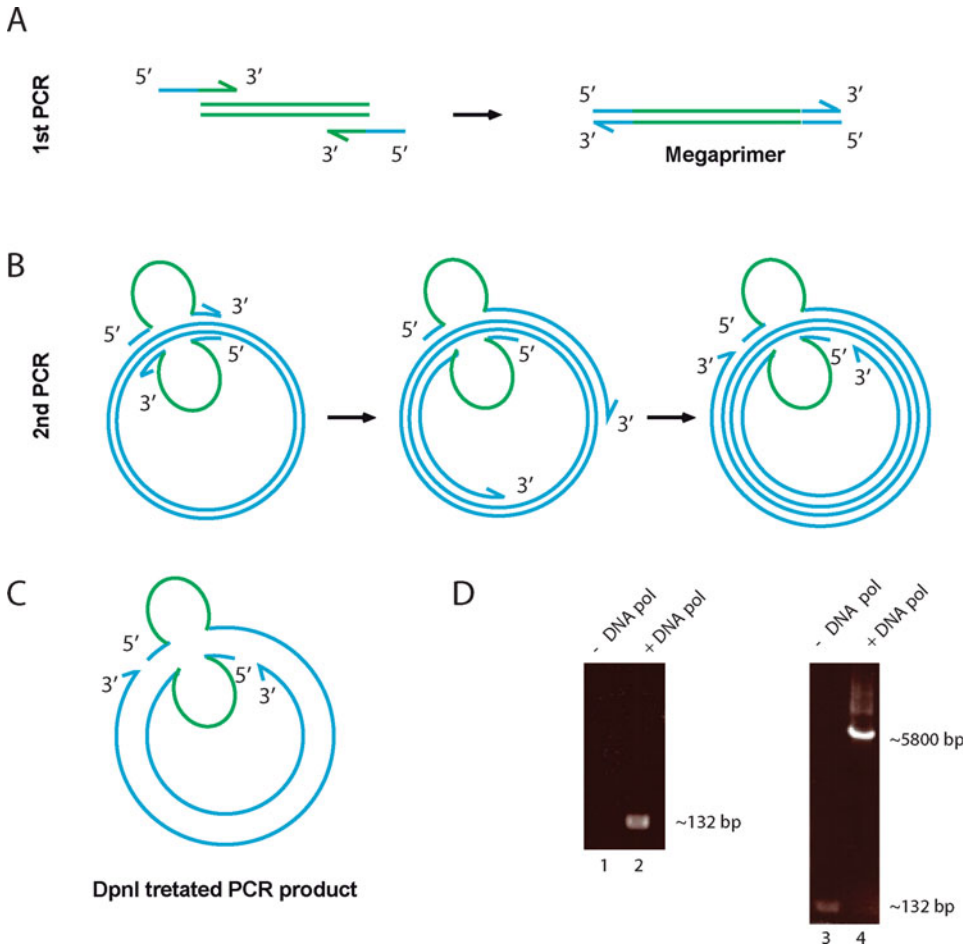


Fig. 6 Modification of an existing construct. In this example, RF cloning is used to introduce a cleavable affinity tag (3C-twinstrep) at the 3′ extremity of the MAT1 cDNA which is part of a multigene expression construct. The sequence of the modified plasmid where the existing part (in blue) and the insert (in green) are shown. The hybrid primers *MAT1-TwinStrep-FW* and *MAT1-TwinStrep-RV* used to amplify the insert (3C cleavage site and Twin-Strep tag) and introduce 20–22 bp overhangs matching the *MAT1* gene at the insertion point (immediately before the Stop codon) are also displayed

4. Run the PCR in a thermocycler by using the conditions suggested by the polymerase manufacturer. We typically denature the template DNA at 98 °C for 30 s, perform 25 cycles of amplification (denaturation at 98 °C for 30 s, annealing at the calculated Tm for 15 s and elongation at 72 °C for 60 s in the case of a 1 kb gene) followed by a final incubation at 72 °C for 10 min.
5. Load 1/10 of the PCR volume on a 1% agarose gel in parallel with a molecular weight marker to verify that the PCR amplification resulted in a single band of the expected size.

Fig. 7 Principle of site-directed insertion by RF cloning. (**a**) A first PCR amplifies the sequence to be inserted (green, double-stranded) and introduces overhangs at each end (light blue strands), which overlap the insertion site at the destination vector. The primers contain a target-specific sequence and an 18–25 bp extension overlapping the desired insertion sites at the destination vector. The double-stranded PCR product is then used as a megaprimer for the second amplification reaction. (**b**) In this step, each of the DNA strands anneals to the destination vector and extends in a linear amplification reaction using the destination vector (navy blue circumferences) as template, leading to a double-stranded nicked plasmid. (**c**) The parental DNA is then removed by DpnI treatment, and newly synthesized plasmid, containing the DNA insert, is transformed into *Escherichia coli* cells where the nicked DNA is sealed by endogenous repair activities. (**d**) Agarose gel electrophoresis of the product from the first (left panel) and second (right panel) PCRs. The amplified insert (lane 2, 132 bp) was used as a mega-primer to generate the desired vector (lane 4, 5800 bp). The Phusion DNA polymerase was omitted in the negative control (lanes 1 and 3). Note that the 132-bp fragment, visible in the negative control (lane 3), has been consumed in the presence of DNA polymerase (lane 4)

6. Purify the PCR product using PCR clean-up columns following the instruction manual, elute bound DNA in 30 μL, and determine the concentration of the extracted DNA spectrophotometrically.
7. Set the second 50 μL PCR with 100–200 ng of the megaprimer, 20 ng of the recipient plasmid, and 0.8 μL of Phusion polymerase (Thermo scientific™). Run the PCR in a thermocycler by using the conditions suggested by the polymerase manufacturer. As a negative control, set up the same reaction omitting DNA polymerase.
8. 1/10 of the second PCR can be loaded on an agarose gel with a molecular weight marker to verify that the PCR amplification resulted in a band of the expected size, but this step is optional (*see* **Note 9**).
9. Add 1 μL of DpnI restriction enzyme directly to the second PCR and incubate 1 h at 37 °C. DpnI enzyme will digest and destroy the plasmid template still present in the PCR.
10. Use 2 μL of the DpnI-treated PCR product from the second reaction to transform 50 μL of *E. coli* competent cells (Pir1 cells for plasmids with an R6Kγ replication origin). Plate the

transformation mixture on LB agar plates containing the appropriate antibiotic and incubate overnight at 37 °C. If all steps were successful, there will be no or little colonies in the negative control and at least one order of magnitude more colonies in the experiment.

11. Purify plasmid DNA from six colonies for restriction analysis and sequencing.

4 Notes

1. *E. coli* strain Pir1 (F-Δlac169 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMluI)::pir-116) or Pir2 (F-Δlac169 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMluI)::pir) is for use with vectors that contain the R6Kγ origin of replication. The *pir* gene encodes the replication protein π, which is required to replicate and maintain plasmids containing the R6Kγ origin. Other strains including BW23473 or BW23474 can also be used.
2. Refer to general guidelines for primer selection [21] and/or use a primer design tool available from vector mapping software or web-services (SnapGene or NEBuilder for the planning of homology-based cloning projects and RF-Cloning.org for designing hybrid RF cloning primers). Also check the propensity of the primers to form dimers or secondary structures. Structures with a ΔG greater than −6 kcal/mole suggest that redesign may be required.
3. It is important to have a sharp, specific PCR band for the SLIC reaction. Gel purification of PCR products considerably reduces the amount and the purity of the DNA. Large and difficult-to-amplify cDNAs (more than 2 kb) can be splitted into smaller fragments which are easier to prepare and can be assembled more efficiently by SLIC.
4. Incubation time controls the formation of compatible single-stranded ends and is thus a critical parameter for T4 polymerase treatment. Do not hesitate to optimize the amount of T4 DNA polymerase and incubation time when a new batch of polymerase is used or when the overhang length between PCR products and vector is changed.
5. To calculate the amount of each DNA fragment to be used at a given ratio and knowing the starting amount of vector, use the formula:

$$\text{Insert (ng)} = \text{ratio} \times \frac{\text{insert (bp)}}{\text{vector (bp)}} \times \text{vector (ng)}$$

For example, in a multiple fragment reaction, where we choose a vector:insert ratio of 1:2, using 250 ng of a vector 3 kb long, we would calculate the amount of 500-bp-long insert as:

$$\text{Insert (ng)} = 2 \times \frac{500}{3000} \times 250 = 83.3 \text{ ng of insert}$$

6. As an alternative to restriction digestion, the presence of the correct inserts can be checked by PCR or colony-PCR and amplification across two or more assembled fragments, for example, by using one primer that hybridizes on the coding sequence and another on the plasmid backbone.
7. When using the In-Fusion reaction, the step of DNA purification may be avoided if the PCR products show one clean band of the expected size after gel electrophoresis. These samples may be treated with Cloning Enhancer in the same PCR tube and used directly to set up the In-Fusion reaction. Add 2 μL of Cloning Enhancer to 5 μL of the PCR product and incubate at 37 °C for 15 min, followed by 80 °C for 15 min. Use a thermocycler for accurate temperature control.
8. It is important to have a sharp, specific PCR band as a megaprimer. The eventual gel purification of the PCR should be avoided since it can impact the efficiency of the following PCR amplification reaction. It is crucial to optimize the PCR conditions until a specific band is obtained.
9. Agarose gel electrophoresis after the second PCR is optional. Even if the discrete band of the high molecular weight PCR product is not clearly visible, there might be enough material to obtain several colonies. You can proceed immediately to *E. coli* transformation step after DpnI treatment.

Acknowledgments

This work was supported by the Centre National pour la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), Université de Strasbourg (UdS), the Association pour la Recherche sur le Cancer (ARC), the Ligue nationale contre le cancer, the Institut National du Cancer (INCa; INCA 9378), and the Agence National pour la Recherche (ANR-12-BSV8-0015-01 and ANR-10-LABX-0030-INRT under the program Investissements d'Avenir ANR-10-IDEX-0002-02). Instruct (R&D Project Funding), Instruct-ULTRA as part of the European Union's Horizon 2020 (Grant ID 731005), and Instruct-ERIC are also acknowledged.

References

1. Contreras-Gomez A, Sanchez-Miron A, Garcia-Camacho F, Molina-Grima E, Chisti Y (2014) Protein production using the baculovirus-insect cell expression system. Biotechnol Prog 30(1):1–18
2. Martinez-Solis M, Herrero S, Targovnik AM (2019) Engineering of the baculovirus expression system for optimized protein production. Appl Microbiol Biotechnol 103(1):113–123
3. Berger I, Poterszman A (2015) Baculovirus expression: old dog, new tricks. Bioengineered 6(6):316–322
4. Gupta K, Tolzer C, Sari-Ak D, Fitzgerald DJ, Schaffitzel C, Berger I (2019) MultiBac: baculovirus-mediated multigene DNA cargo delivery in insect and mammalian cells. Viruses 11(3):198
5. Berger I, Fitzgerald DJ, Richmond TJ (2004) Baculovirus expression system for heterologous multiprotein complexes. Nat Biotechnol 22(12):1583–1587
6. Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I (2006) Protein complex expression by using multigene baculoviral vectors. Nat Methods 3 (12):1021–1032
7. Trowitzsch S, Bieniossek C, Nie Y, Garzoni F, Berger I (2010) New baculovirus expression tools for recombinant protein complex production. J Struct Biol 172(1):45–54
8. Weissmann F, Petzold G, VanderLinden R, Huis In't Veld PJ, Brown NG, Lampert F, Westermann S, Stark H, Schulman BA, Peters JM (2016) biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. Proc Natl Acad Sci U S A 113(19): E2564–E2569
9. Zhang Z, Yang J, Barford D (2016) Recombinant expression and reconstitution of multiprotein complexes by the USER cloning method in the insect cell-baculovirus expression system. Methods 95:13–25
10. Osz-Papai J, Radu L, Abdulrahman W, Kolb-Cheynel I, Troffer-Charlier N, Birck C, Poterszman A (2015) Insect cells-baculovirus system for the production of difficult to express proteins. Methods Mol Biol 1258:181–205
11. Abdulrahman W, Radu L, Garzoni F, Kolesnikova O, Gupta K, Osz-Papai J, Berger I, Poterszman A (2015) The production of multiprotein complexes in insect cells using the baculovirus expression system. Methods Mol Biol 1261:91–114
12. Abdulrahman W, Iltis I, Radu L, Braun C, Maglott-Roth A, Giraudon C, Egly JM, Poterszman A (2013) ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. Proc Natl Acad Sci U S A 110(8): E633–E642
13. Kobbi L, Demey-Thomas E, Braye F, Proux F, Kolesnikova O, Vinh J, Poterszman A, Bensaude O (2016) An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. Proc Natl Acad Sci U S A 113(45):12721–12726
14. Radu L, Schoenwetter E, Braun C, Marcoux J, Koelmel W, Schmitt DR, Kuper J, Cianferani S, Egly JM, Poterszman A, Kisker C (2017) The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH. Nucleic Acids Res 45(18):10872–10883
15. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6 (5):343–345
16. Irwin CR, Farmer A, Willer DO, Evans DH (2012) In-fusion(R) cloning with vaccinia virus DNA polymerase. Methods Mol Biol 890:23–35
17. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4 (3):251–256
18. Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48(6):463–465
19. Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y (2010) Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. J Struct Biol 172(1):34–44
20. van den Ent F, Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods 67(1):67–74
21. Bustin SA, Mueller R, Nolan T (2020) Parameters for successful PCR primer design. Methods Mol Biol 2065:5–22

Chapter 3

Tagging Proteins with Fluorescent Reporters Using the CRISPR/Cas9 System and Double-Stranded DNA Donors

Sylvain Geny, Simon Pichard, Alice Brion, Jean-Baptiste Renaud, Sophie Jacquemin, Jean-Paul Concordet, and Arnaud Poterszman

Abstract

Macromolecular complexes govern the majority of biological processes and are of great biomedical relevance as factors that perturb interaction networks underlie a number of diseases, and inhibition of protein–protein interactions is a common strategy in drug discovery. Genome editing technologies enable precise modifications in protein coding genes in mammalian cells, offering the possibility to introduce affinity tags or fluorescent reporters for proteomic or imaging applications in the bona fide cellular context. Here we describe a streamlined procedure which uses the CRISPR/Cas9 system and a double-stranded donor plasmid for efficient generation of homozygous endogenously GFP-tagged human cell lines. Establishing cellular models that preserve native genomic regulation of the target protein is instrumental to investigate protein localization and dynamics using fluorescence imaging but also to affinity purify associated protein complexes using anti-GFP antibodies or nanobodies.

Key words CRISPR/Cas9, Genome editing, Double-stranded DNA donors, Fluorescent protein

1 Introduction

Molecular complexes of interacting proteins govern virtually all biological processes such as metabolism, cell signaling, DNA repair, and gene expression. Macromolecular assemblies are also of great biomedical relevance as their dysfunctions underlie a number of diseases, and deliberate inhibition of protein–protein interactions is an increasingly common strategy in drug discovery [1–3]. To fully understand their biological roles, it is essential to study the structure and function of intact protein assemblies. Although advanced recombinant protein technologies are available to reconstitute multiprotein complexes composed of ten or more subunits, many protein complexes are difficult to obtain using recombinant methods. An additional hurdle is that the subunit composition of complexes is not always known well enough to proceed to

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_3,

reconstitution of functional entities and therefore to structure determination.

The CRISPR/Cas9 system has revolutionized many fields of life sciences by making it possible to modify the genome sequence with unprecedented efficiency and precision [4, 5]. For example, it is now possible to insert fusion tags at endogenous loci of mammalian cells, providing an efficient way to undertake affinity purification of macromolecular complexes and/or visualize their distribution and dynamics in a cellular context [6, 7]. We have recently detailed a simplified protocol for CRISPR/Cas9-mediated gene tagging in human cell lines using chemically modified single-stranded oligonucleotides encoding a small affinity tag as donor template and a co-selection strategy targeting the *ATP1A1* gene [8] (to be published in MiB, Editor R Owens). Here we describe a procedure based on the use of the CRISPR/Cas9 system and double-stranded DNA donors for tagging the protein of interest with a fluorescent reporter and detail the generation of an U2-OS$^{XPB\text{-}GFP}$ knocked-in cell line expressing a XPB-GFP fusion protein. XPB is a subunit of the TFIIH complex, essential in initiation of DNA transcription by RNA polymerase II and DNA repair by nucleotide excision repair [9, 10]. This XPB-GFP-tagged U2-OS cell line was instrumental to establish a partnership between TFIIH and the histone acetyl transferase GCN5 and the impact of TFIIH on GCN5 activity with important consequences on gene expression and chromatin structure [11].

2 Materials

Procedures described here need access to standard equipment for molecular biology (PCR amplification, agarose gel analysis, bacteria transformation, access to a DNA sequencing/synthesis service, etc.), cell culture (cryo-container and liquid nitrogen source, temperature- and CO_2-controlled incubator, laminar flow hood, centrifuge with adaptor for 15 and 50 mL tubes, cell counter), cell microscopy (fluorescence microscope with 63× or higher magnification objectives, etc.), and protein analysis (refrigerated microcentrifuge, a small-scale ultrasonic homogenizer, nano-UV spectrometer, protein gel electrophoresis, and Western blotting transfer system, etc.). Basic knowledge in these fields is expected.

Experiments detailed below were performed in U2-OS cells (ATCC HTB-96). These adherent cells are grown at 37 °C, 5% CO_2 in McCoy's 5a or DMEM-based culture medium (*see* **Note 1**). For experiments in Subheadings 3.2–3.5, in addition to specific materials, you will need:

1. T75 cell culture flasks (75 cm^2) (Corning, cat. no. 431464 U).
2. McCoy's 5a or DMEM medium supplemented with 10% fetal bovine serum (FBS), 4 mM GlutaMAX™ (Gibco, cat. no. 35050061) and 1% PenStrep (Gibco, cat. no. 15140-122).
3. Phosphate-buffered saline (PBS).
4. Trypsin–EDTA solution (0.025% w/v Trypsin, 0.05 mM EDTA in 1× PBS) (Gibco, cat. no. R001100).

2.1 Reagents for the Preparation of sgRNA Expression Plasmids

1. Oligonucleotides for preparing (sgRNA-FW, sgRNA-RV) and sequencing (U6-seq) the single guide RNA expression plasmids (sgRNA) (Table 1).
2. Guide RNA expression plasmid: MLM3636 (Addgene plasmid #43860).
3. T4 DNA ligase with buffer for the preparation of the guide RNA expression plasmid (New England BioLabs, cat. no. M0202S).
4. BsmBI restriction enzyme (New England BioLabs, cat. no. R0580S).

Table 1
XPB locus

	U-2 $OS^{XPB::GFP}$
Chromosome/gene	Chr. 2/XPB
Homology arms	5′ Chr.2: 127257599–127257872 (hg 38) 3′ Chr.2: 127257596–127256813 (hg 38)
sgRNA	5′ **TAGGAAATGATGCTTAGGCA**ggg 3′
Selection marker	Puromycin
sgRNA-FW	5′ ACACCG **TAGGAAATGATGCTTAGGC** G 3′
sgRNA-RV	5′ AAAAC **TGCCTAAGCATCATTTCCTA** CG
U6-seq	5′ CAGGGTTATTGTCTCATGAGCGG-3′
XPB-FW	5′ AGACAGTAAGCGATCTGTAAACA 3′
XPB-RV	5′ ACCCCACTCCCCAAAAAGTT 3′
XPB-FW2	5′ TCCTCTTCTTTCAGGTGTGGA 3′
GFP-RV	5′ GAACTTCAGGGTCAGCTTGC 3′
Puro-FW	5′ GCAACCTCC CCTTCTACGAG 3′
XPB-RV2	5′ GCGAATATGCCTTATGTGTG 3′

Target sequence for the sgRNA used for editing the C-terminus of the XPB gene. Online web applications such as CRISPOR (http://crispor.tefor.net/) aimed at optimizing CRISPR knock-in tag experimentation identify and rank potential guide RNAs in an input sequence. Oligonucleotides for sense and antisense strands designed with CRISPOR software for cloning into MLM3636 (sgRNA-FW and sgRNA-RV) as well as for genomic screening (XPB-FW, RV) are detailed. Efficient transcription from the U6 promoter requires a 5′ G (underlined)

5. LB and LB agar medium, ampicillin, competent DH5α *E. coli* cells (New England BioLabs, cat. no. C2987I).
6. Plasmid DNA Mini and Midiprep purification kits (Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit and Macherey-Nagel™ NucleoBond™ Xtra Midi).

2.2 Reagents for Nucleofection and Assessment of Gene Editing

2.2.1 Nucleofection with AMAXA Nucleofector Machine

1. Expression plasmid for SpCas9: JDS246 (Addgene plasmid #43861).
2. Expression plasmid for sgRNA (Addgene plasmid #43860) (from Subheading 3.1, **step 4**).
3. AMAXA nucleofector machine® (Lonza) and consumables including cuvettes (Lonza, cat. no. AAB-1001), pipettes, and solution V kit (Lonza, cat. no. VVCA-1003).
4. Tissue culture plate, 6 wells (Eppendorf, cat. no. EP0030720121).

2.2.2 Pool and Clone Analysis

1. QuickExtract DNA solution (Viagen, cat. no. 302C).
2. Oligonucleotides for PCR amplification and sequencing (*see* Table 1).
3. Phusion high-fidelity DNA polymerase with buffer and dNTPs (New England BioLabs, cat. no. M053OS).
4. PCR isolation kit (Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit).

2.3 Reagents for Gene Editing and FACS Sorting

2.3.1 Nucleofection and Antibiotic Selection

In addition to the reagents described in Subheading 2.2.1

1. Donor plasmid (from Subheading 3.3, **step 2**).
2. Puromycin dihydrochloride (InvivoGen, cat. no. ant-pr-1).

2.3.2 FACS Sorting

Sorting of GFP-fluorescent cells was carried out on a Melody FACS (BD), and analysis of GFP-positive cells was carried on an Accuri C6 analyzer (BD). You will need:

1. Cytometry Falcon tubes (Becton Dickinson).
2. Tissue culture plate, 96 wells (Eppendorf, cat. no. EP0030730119).

2.4 Reagents for the Validation of Selected Clones

2.4.1 Genotyping of Single-Cell Clones

In addition to reagents described in Subheading 2.2.2, you will need:

1. Tissue culture plates, 6 and 96 wells (Eppendorf, cat. no. EP0030720121, EP0030730119).
2. Oligonucleotides from the inserted sequence and from the target chromosomal locus.

2.4.2 Western Blot Analysis

1. 150-mm Petri dishes (Falcon® 150 mm TC-treated dishes, Corning, cat. no. 353025).
2. Cell scrapers (Dominique Dutscher, cat. no. 353085).
3. PBS containing 30% w/v glycerol (Sigma Aldrich, cat. no. G5516).
4. RIPA buffer: 20 mM Tris-HCl or HEPES, pH 7.5, 120 mM KCl, 1% NP-40, 0.1% SDS, 1 mM EDTA, 0.5% Na-deoxycholate, supplemented with protein inhibitor cocktail (Roche™) and 0.5 mM 1,4-dithreothiol (DTT) (Sigma Aldrich).
5. Lysis buffer: 20 mM Tris or HEPES, 250 mM KCl, NP-40 0.05% supplemented with protein inhibitor cocktail (Roche™) and 0.5 mM DDT.
6. Protein assay dye reagent (Bio-Rad, cat. no. 5000006).
7. Laemmli buffer 4×: 60 mM Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 0.0005% Bromophenol Blue, 355 mM β-mercaptoethanol.
8. SDS PAGE gels, PVDF transfer membranes, 3MM Whatman paper.
9. TBST: 20 mM Tris/HCl pH 7.4, 150 mM NaCl, 0.1% Tween-20.
10. Non-fat dry milk or BSA powder.
11. Western blotting detection reagents (Amersham ECL Prime, cat no. 3030-931).
12. Primary antibodies against the target protein or against the GFP tag and corresponding secondary antibodies coupled to horseradish peroxidase (HRP). For the detection of XPB protein, we used a mouse monoclonal anti-XPB antibody (1B3, MABE1123, Sigma-Aldrich) and a donkey anti-mouse antibody coupled with HRP.

2.4.3 Fluorescence Microscopy

1. Glass-bottom dishes (Glass Bottom Dish 35 mm, Clinisciences/Ibidi, cat. no. 81218-200).
2. Epifluorescence inverted microscope with filters compatible with DAPI/Hoechst and the fluorescent reporter used (here GFP).
3. Paraformaldehyde (PFA) 16% ultrapure methanol free (Euromedex, cat. no. 15710).
4. Hoechst (Sigma Aldrich, cat. no. 32670).
5. Antibodies to confirm the expression of the target protein or for co-localization experiments.

3 Methods

Targeted gene knock-in is achieved through the following steps. First, the Cas9 endonuclease is directed by a guide RNA to a specific target site in the genome to generate a double-stranded break (DSB). The DSB is then repaired with a donor template through a homology-directed repair (HDR) pathway [12]. Below, we first describe the design of the single-guide RNAs and donor plasmid template to edit the genome (Subheadings 3.1–3.3) and the clonal isolation of CRISPR/Cas9-modified cells (Subheading 3.4). Selected clones are finally characterized at the genomic and protein levels (Subheading 3.5) (Fig. 1).

To facilitate the generation of homozygous knock-in cells, we favored a tagging strategy based on the use of a donor plasmid containing 300–1000 bp homology arms flanking the eGFP-2A-PURO coding sequence [13] (Fig. 2). This strategy allows to combine transient antibiotic selection with fluorescence-assisted cell sorting (FACS) to isolate CRISPR/Cas9-modified cells [14].

3.1 Design of the sgRNAs

The design of the sgRNA and HDR donor depends on the nature and position of the tag (*see* **Note 2**). As the efficiency of HDR insertion varies with the distance from the DSB, the DSB (generated 3 bp upstream of the PAM sequence corresponding to the sgRNA) should be as close as possible to the insertion site of tag. For example, to tag a protein at its C-terminal end, the DSB should be as close as possible to the stop codon of the gene of interest (Fig. 3).

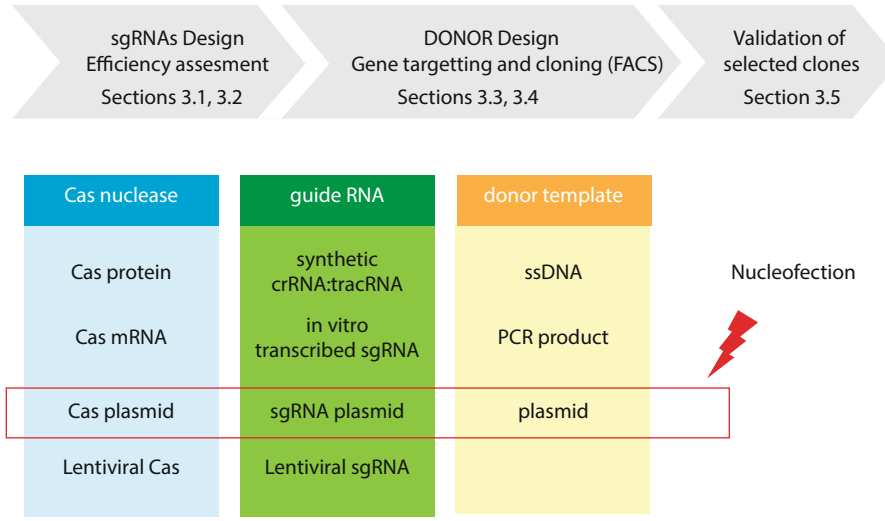


Fig. 1 Experimental workflow for CRISPR/Cas9-mediated genome editing. Several sources of Cas nuclease, guide RNA and donor template can be used. We detailed a protocol that utilizes electroporation (nucleofection) for the delivery of plasmidic reagents (donor, Cas9, sgRNA)

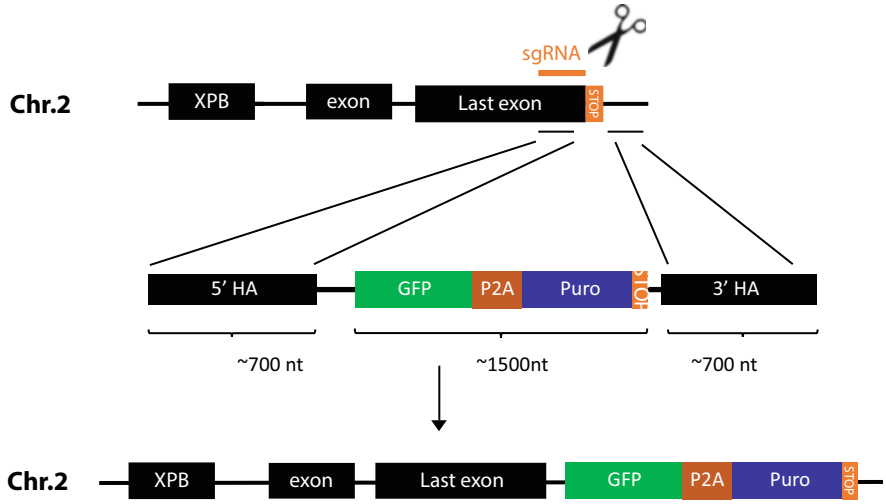


Fig. 2 Schematic of the tagging strategy for the XPB locus. To introduce a GFP tag at the C-terminal end of the XPB gene, we select an sgRNA to generate a CRISPR/Cas9-induced DSB in vicinity of the stop codon of the XPB gene. The donor plasmid contains the eGFP-2A-Puro coding sequence flanked by respectively 300 and 800 bp homology arms (5′ and 3′ HA). The eGFP-2A-Puro coding sequence is inserted at the position of the XPB STOP codon and in frame with the last exon of the target protein

1. Retrieve the nucleotide sequence of the gene of interest (GOI) and focus on the site of tag insertion. The UCSC genome browser is a convenient source for this step (genome.ucsc.edu).
2. Screen the genomic regions of interest using an online CRISPR design tool such as CRISPOR (http://crispor.tefor.net/) to identify and rank the guide sequences (Table 1). We recommend testing several sgRNAs with cut sites less than 20 bp away from the insertion site to find an optimal sgRNA. Guide RNAs for the human genome can also be visualized directly in the UCSC genome browser with pre-calculated data from CRISPOR (select "Full" display mode of the "CRISPR targets" in the "Genes and Gene Predictions" list of features).
3. Order the sense and antisense oligonucleotides for cloning the sgRNA sequences into an sgRNA expression plasmid such as MLM3636. Note that the oligonucleotides contain overhangs for ligation into the pair of BsmBI sites with the sense and antisense sequences matching the genomic target. The oligonucleotide sequences can be accessed by clicking on the "PCR/cloning primers" button in CRISPOR. If you have used the UCSC genome browser for sgRNA selection, you can directly transfer the sgRNA sequence to CRISPOR by a hyperlink feature or simply by copy-pasting to the CRISPOR input page.

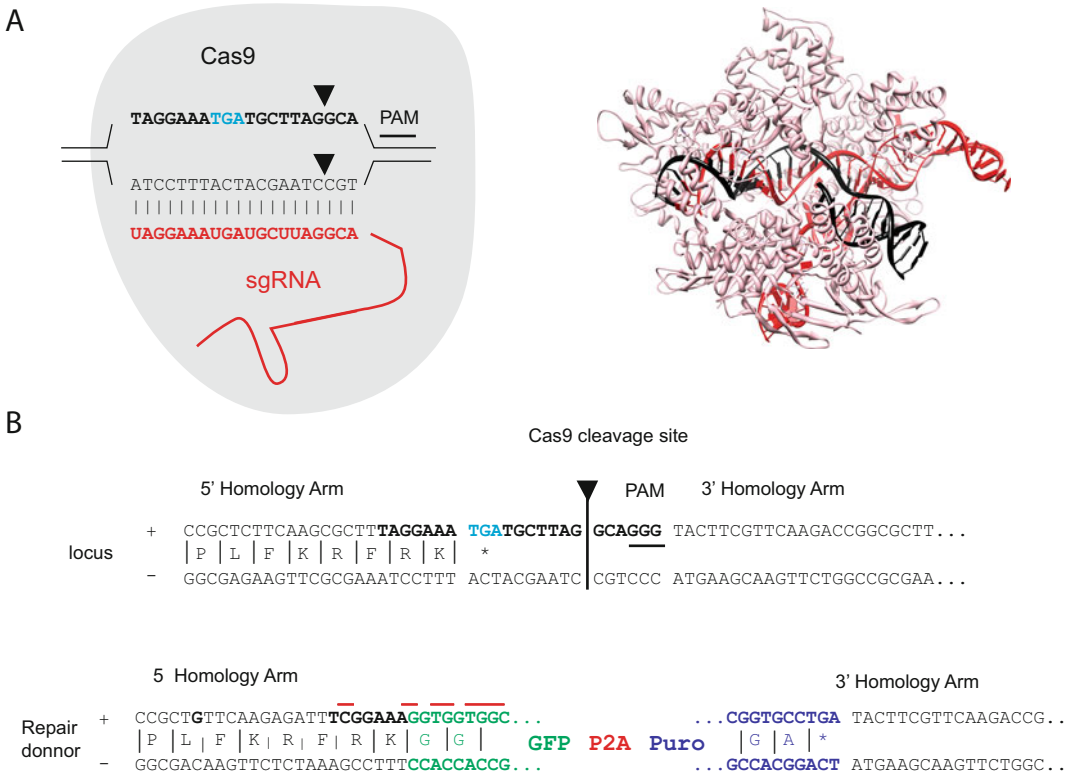


Fig. 3 Design of sgRNAs and HDR donor sequences. (**a**) Schematic representation of the Cas9 protein (gray) bound to sgRNA (red) and targeted to genomic DNA (black). The PAM sequence (GGG with overbar), the Cas9 cleavage sites (black triangles), and the stop codon of the XPB gene (TGA, in cyan) are indicated. A ribbon representation of the Cas9 nuclease from *Streptococcus pyogenes* bound to PAM-containing DNA target is shown in the right panel (PDB code, 4UN3; [30]). (**b**) Repair donor designed to insert the eGFP-2A-Puro coding sequence (all in frame with the last exon of the target protein) upstream of the XPB STOP codon (TGA). Nucleotides targeted by the sgRNA are in bold. After integration of the tag, the sequence is significantly modified (red line above the + strand), preventing cleavage of the repaired locus by Cas9

4. Generate the oligo duplex, clone into the BsmBI predigested plasmid MLM3636, sequence the resulting plasmid. Transform and amplify the guide expression plasmid in DH5 competent cells using a Midi or Maxi Prep kit depending on the amount of plasmid required.

3.2 Delivery of CRISPR Reagents and Assessment of Gene Editing

As knock-in using double-stranded donor template has relatively low efficiency, successful HDR-mediated experiments require experimental optimization. In particular, it is recommended to control and optimize delivery of CRISPR reagents by evaluating the levels of gene editing in absence of donor template. In some cases, GFP fluorescence can be used to directly characterize and optimize HDR events, but low to very low levels of fluorescent-tagged proteins driven by endogenous promoters can be

challenging to detect. An attractive alternative consists in delivering the nuclease with a single strand donor designed to fuse a HiBiT peptide to the target protein. The abundance of HiBiT-tagged proteins in transfected cells is proportional to the proportion of edited gene copies and is readily detected using the Nano-Glo HiBiT luciferase-based detection system (*see* ref. 8).

In this section, we detail the delivery of CRISPR/Cas9 reagents by nucleofection (*see* **Note 3**) and PCR-based analysis of the cleavage efficiency of guide RNAs.

3.2.1 Nucleofection

1. Thaw and maintain U2-OS cells using your favorite protocol (*see* **Note 1**). We usually split cells 1:5 every 3–4 days, verify that the viability is always higher than 95% and limit the number of passages to 20.
2. Two or three days before nucleofection, seed two 75-cm^2 flasks and incubate at 37 °C, 5% CO_2.
3. When 50–70% confluence is reached, detach the cells with trypsin, spin-down in a 15-mL tube at 90 g at room temperature for 10 min, remove the supernatant and re-suspend the cells in PBS to a concentration of 1–3 × 10^6 cells/mL.
4. Aliquot 10^6 cells into 1.5-mL microcentrifuge tubes (one tube per guide RNA to be tested plus an additional tube for a negative control), pellet the cells, discard supernatant completely, and resuspend cells in Nucleofector® solution (100 μL per sample).
5. Prepare the DNA mixes containing 2 μg of guide sgRNA plasmid (sgXPB) and 2 μg of plasmid expressing the SpCas9 protein (JDS246) in a total volume less than 10 μL.
6. Mix the DNA solution with 100 μL of cell suspension and transfer to a cuvette. Process the samples quickly to avoid storing the cells longer than 10 min in Nucleofector® Solution V.
7. Insert the cuvette into the Nucleofector®, select the cell-type-specific program X-001 and press the start button. Using the provided pipette, immediately remove the sample from the cuvette and transfer into the six-well plate containing 2 mL of culture medium.
8. Incubate the six-well plates at 37 °C, 5% CO_2 and proceed to analysis or collect the cells and store the cell pellet at −20 °C.

3.2.2 PCR-Based Analysis of the Cleavage Efficiency for the sgRNA

CRISPR/Cas9 cleavage is usually assessed by amplifying the target region by PCR and analyzing the mutation rate in the resulting product. This can be performed using various assays such as the polymerase chain reaction (PCR)/restriction enzyme (RE) assay, T7 endonuclease I (T7EI) assay, Surveyor nuclease assay or with TIDE [15–18].

Table 2
Composition of the PCR mix used to amplify the genomic site XPB

Genomic DNA (50 ng/μL)	1 μL
dNTPs (25 mM)	0.4 μL
DNA polymerase (2000 U/mL)	0.5 μL
Buffer PCR 5×	10 μL
Forward Primer (10 μM)	2.5 μL
Reverse Primer (10 μM)	2.5 μL
H_2O	33.1 μL

Use the following PCR program: 98 °C/30 s; 98 °C/10 s; 65 °C/20 s; 72 °C/20 s (30 cycles); 72 °C/5 min; 4 °C/hold

1. Three days after transfection, wash cells with PBS, extract genomic DNA (typically from 1×10^5 cells) using QuickExtract DNA extraction kit or similar according to the manufacturer's recommendations. Dilute genomic DNA at a final concentration around 50 ng/μL.
2. Design a pair of PCR primers (XPB-FW and XPB-RV) which hybridize on either side of the target site to amplify a ~500 bp stretch of DNA. The sgRNA cut site should ideally be located ~200 bp downstream from one of the primers for convenient analysis by DNA sequencing.
3. Amplify the locus of interest by PCR (*see* Table 2), run 2 μL of PCR products on 1% agarose gel TBE 0.5× at 100 V for 90 min and visualize DNA using a UV transilluminator. A sharp single band of the expected size should be visible in control samples. Additional bands may be detected in test samples corresponding to insertion and/or deletion.
4. Purify the PCR products using an appropriate PCR purification kit and measure its concentration.

A wide panel of assays can be used to analyze PCR products and verify the efficacy of programmable nucleases. The enzymatic Surveyor and T7 endonuclease I cleavage assays [19, 20] are commonly used, but these approaches are semiquantitative and their sensitivity is limited [15]. We favor the TIDE assay that accurately quantifies editing efficacy and simultaneously identifies the predominant types of insertions and deletions (indels) in the pool of treated cells [21].

5. Send the PCR products (control and experimental samples) for Sanger sequencing and retrieve the sequence trace files in .ab1 or .scf format.

6. Upload the trace files and the guide RNA sequence (20 nt) into the TIDE web tool (available at http://tide.nki.nl and https://deskgen.com) and perform the analysis with default parameters.

Use the TIDE assay to compare the efficiency of the different guides and select the best experimental conditions for further experiments. An example of TIDE analysis is provided on the TIDE website.

3.3 Design of Plasmid Donor

Several types of DNA repair template can be utilized in HDR-mediated CRISPR/Cas9 gene editing. Donor plasmids, with homology arms of 300–1000 bp allow for large insertions in the 1–2 kbp range. Other types of DNA repair templates such as linear DNAs with short homology arms obtained by PCR or gene synthesis can also be used for inserting fluorescent labels or other protein tags such as the Halo or Snap tags [22, 23]. Linear DNA, however, may concatemerize before insertion [24], which may perturb gene expression or make the modification unstable during cell replication. Different types of donor DNA will give similar or different KI efficiencies depending on the gene targeted and cell line used and may be tested if needed.

This section details the design and preparation of the donor plasmid which requires the introduction of the coding sequence for the insertion flanked by two homology arms of a few hundred bases (typically between 0.3 and 1.0 kbp).

1. Define the DNA fragment that you wish to introduce, which in this example case consists in the eGFP-P2A-PURO coding sequence where eGFP, the 2A self-cleaving peptide (P2A), and the puromycin N-acetyl transferase coding sequence are separated by flexible glycine/serine-rich linkers (Fig. 4). References [25, 26] provide useful information on the use of 2A peptides in a polycistronic vector and the design of protein linkers, respectively. Note that when using 2A self-cleaving peptide, antibiotic resistance is expressed from the target gene, which minimizes the selection of clones with non-targeted integration. However, in some cases, expression levels of the target gene may be insufficient to provide antibiotic resistance, and it may be necessary to provide antibiotic resistance from a constitutive promoter; higher numbers of clones with random integration of the donor plasmid into the cell genome may be isolated [27].
2. Retrieve the sequence of a fragment encompassing upstream and downstream homology arms of approximatively 700 bp and replace the STOP codon of the GOI by the eGFP-P2A-PURO coding sequence. Carefully verify that the eGFP-P2A-PURO coding sequence is in frame with the last exon of the

ggtggtggcggttcaggcggaggtggctctggcggtggcggatcggccATGGTGAGCAAGGGCGAGGAGCTGTTCACCGG
GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACC
CTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGA
AGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC
GTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGtacagcggtcgcgactctagagtcgacggttctGG
TAAGCAAACTTTGAATTTTGACCTTCTTAAGCTTGCGGGAGACGTCGAGTCCAACCCCGGGCCCgcatgcaagcttcagc
tgaagcttaccATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTC
GCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCA
AGAACTCTTCCTCACGCGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTCTGGA
CCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTG
GCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGT
CTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGC
CCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAG
GTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGA

Fig. 4 eGFP-2A-Puro nucleotide sequence. cDNAs encoding gGFP and the puromycin resistance genes are in green and blue, respectively, the P2A peptide in red and linker regions in black

target protein and that the guide RNA or PAM sequence has been modified to prevent cleavage by Cas9. If that is not the case, introduce point mutations to avoid unwanted mutations being introduced after integration.

Having precisely defined the donor sequence, different options can be considered to prepare the donor plasmid.

3. Purchasing the designed sequence from your preferred supplier is the easiest approach. Alternatively, you can amplify the eGFP-P2A-PURO coding sequence from an existing plasmid, the homology arms from genomic DNA, and assemble the three fragments into your preferred cloning vector using restriction enzyme-free technologies (*see* **Note 4**).

3.4 Gene Targeting and Isolation of Gene Edited Cells by FACS Sorting

When the CRISPR/Cas9 reagents are ready, you can proceed to genome editing with CRISPR/Cas9 to generate cell lines that express an endogenously tagged protein. The strategy detailed below relies on antibiotic selection combined with fluorescence-activated cell sorting (FACS) for enrichment in HDR during genome editing. FACS is an efficient technology which allows to simultaneously measure the knock-in efficiency by counting fluorescently positive cells and sort isolated cells into 96-well plates. As cell sorting relies on the expression levels of the fluorescent-tagged proteins driven by endogenous promoters, isolation of positive clones can be challenging for proteins expressed at low levels.

3.4.1 Nucleofection and Antibiotic Selection

1. Proceed as detailed in Subheading 3.2.1 with a master mix containing 2 μg of the selected guide sgRNA plasmid (sgXPB), 2 μg of plasmid expressing the SpCas9 protein (JDS246), and 6 μg of donor plasmid in a total volume lower than 10 μL.
2. Three days after transfection, replace the culture medium and add puromycin to a concentration of 5 μg/mL and further incubate for an additional week. The use of puromycin selection can be avoided if the signal of the fluorescently tagged protein is sufficiently strong and the proportion of fluorescent cells is sufficient for FACS sorting.

3.4.2 Isolation of Gene Edited Cells by FACS Sorting

Proceed to FACS sorting 7 days after the beginning of the puromycin treatment (or 3 days after nucleofection in absence of antibiotic selection). Use non-transfected cells as negative control for gating.

1. Prepare 96-well plates containing 100 μL of growth medium in each well.
2. Wash cells with PBS, add 1 mL of trypsin and transfer the resuspended cells in a 1.5-mL microcentrifuge tube containing 0.5 mL of medium.
3. Centrifuge the cell suspension for 5 min at 90 × *g* at room temperature, remove the supernatant, resuspend the cells in 1 mL PBS containing 2% FBS, and filter the cells through a 50-μm cell strainer into a sterile flow cytometry tube.
4. Turn on the cytometer, run the cleaning/calibration procedures and load settings to detect GFP. Refer to the user manual of your instrument for detailed procedures.
5. Run the non-transfected cells to check the cell sorter parameters and establish gates to select single live cells (Fig. 5). For single cells, plot forward scatter linear (FCS) on the *y*-axis and the fluorescent channel (GFP-FL) on the *x*-axis. Ideally, set a background threshold where less than 0.1% of the non-transfected control cells are counted as GFP positive.

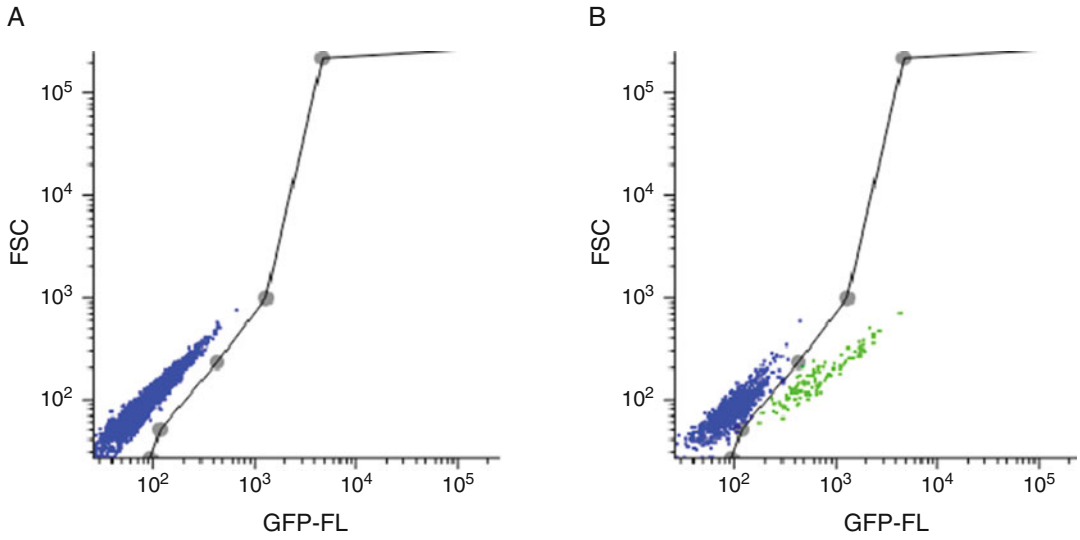


Fig. 5 FACS analysis of wild-type and modified U2-OS cells. Non-transfected U2-OS cells (**a**) and CRISPR/Cas9 modified cells in which a GFP tag is fused to the C-terminus of XPB (**b**) were analyzed for the expression of a green fluorescent protein (**b**). Both the plots show the green fluorescence intensity (GFP-FL) versus the forward scatter (FSC). U2-OS wild-type cells served as a negative control (blue dots) and define background signal; gates for the GFP-positive cells (green dots) were set accordingly. The GFP-positive cells were sorted into 96-well plates to obtain single clones expressing XPB-eGFP

6. Repeat with the cell population to be sorted and verify that the CRISPR/Cas9-modified cells can be differentiated from parental cells. Run the transfected cells to quantify the amount of fluorescently labeled cells and define the collection gate. The fluorescent signal depends on expression levels of the target gene and can be relatively dim (Fig. 6b).
7. Sort the positive cells into the 96-well plates and incubate at 37 °C, 5% CO_2.
8. After ~5 days, identify 24 single-cell clones for validation and keep them growing. Maintain the cells in 96-well plates until you can proceed to validation. Split cells by trypsinization when required.

3.5 Validation of Selected Clones

This section details: (1) genotyping of the knock-in cell clones by PCR to verify that the fluorescent marker was correctly inserted at the target locus and to test for homozygosity, (2) Western blot analysis to verify that the GFP fusion protein is effectively expressed, and (3) fluorescence microscopy to validate expression and localization of the tagged protein.

3.5.1 Validation at the Genomic Level

1. Split cells from each positive well into two 96-well plates. The day after, spin down and wash the cells in one of the plates with PBS.

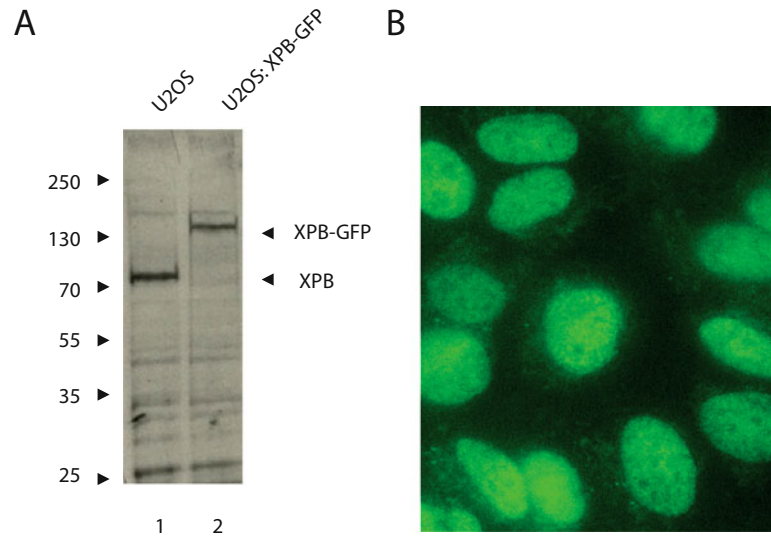


Fig. 6 Validation of engineered cell lines. (**a**) Western blot analysis of engineered U2-OS cells in which wild-type XPB was replaced by C-terminal GFP-tagged version (U2-OS$^{XPB::GFP}$, lane 2). Unmodified U2-OS cells were used as control (U2-OS, lane 1) (Adapted from Sandoz et al. 2019). (**b**) GFP fluorescence of living U2-OS$^{XPB::GFP}$ cells observed using an inverted microscope equipped with a 63× immersion objective

2. Extract genomic DNA directly in the wells using 25 μL of commercially available DNA lysis buffer.
3. Proceed to PCR amplification from genomic DNA of each clone with the primer pairs XPB-FW/XPB-RV or XPB-FW/GFP-RV following the conditions used for PCR analysis of the pool of transfected cells. Run 2 μL of the PCR on 1% agarose TBE 1× gel at 100 V for 1 h. Visualize the PCR products using a UV transilluminator (*see* **Note 5**).
4. Sequence the PCR fragments of clones to ensure correct tag sequence insertion.

3.5.2 Western Blot Analysis

1. Seed five 150-mm Petri dishes containing 25 mL of growth medium and incubate at 37 °C, 5% CO_2 until a confluency of 80% is reached.
2. Wash the plate with PBS containing 30% w/v glycerol, detach the cells with a scraper, and after centrifugation at 1000 × *g* for 10 min at 4 °C, snap freeze the pellet in liquid nitrogen and store at −80 °C. We usually prepare batches of 15×10^6 and 25×10^6 cells.
3. Resuspend 15×10^6 cells in 150 μL of RIPA buffer and incubate for 10 min by pipetting up and down.

4. Sonicate two times for 30 s on ice. We use a Bioruptor™ sonication system (amplitude 30 and 0.5 s pulse on ice) (optional).
5. Centrifuge for 15 min at 14,000 × *g* at 4 °C, collect the supernatant, and estimate the total protein concentration using a Bradford assay. A concentration of 3–4 mg/mL is expected.
6. Heat 20 μg of total proteins from the soluble extract mixed with 2× Laemmli buffer at 95 °C for 5 min and centrifuge samples at 10,000 × *g* for 30 s to bring down the condensate and remove insoluble debris.
7. Load the centrifuged sample on an SDS polyacrylamide gel, electrophorese, and transfer proteins from the gel matrix to a nitrocellulose or PVDF membrane using your favorite device.
8. Block the membrane for 1 h at room temperature or overnight at 4 °C using 3% w/v dry skimmed milk or BSA solution in PBS, incubate the membrane with an appropriate dilution of a primary antibody directed against the affinity tag or the tagged subunit in the same buffer for 1 h at room temperature or overnight at 4 °C.
9. Wash the membrane three times in TBST for 5 min each, incubate with the recommended dilution of conjugated secondary antibody in TBST at room temperature for 1 h and develop the Western blot.

The Western blot analysis of protein lysates from the parental and from a modified cell line where XPB and XPB-GFP are detected with an anti-XPB antibody is shown in Fig. 6a. As expected in the case of a homozygous modification, the 90 kDa wild-type XPB protein (lane 1) is replaced by the 130 kDa XPB-GFP fusion (lane 2). A heterozygous modification would result in the detection of both a 90 kDa and a 130 kDa protein.

3.5.3 Fluorescence Microscopy

We typically select four positive clones and include the non-modified cell line as control. The target protein being fused to GFP, expression of a fluorescent fusion protein can be checked using an epifluorescence inverted microscope. Live or fixed cell imaging can be performed.

1. Seed glass-bottom dishes/plates with clones that have been tested positive until 70–80% confluency is reached.
2. For live cell imaging, change the medium to a phenol red-free medium and observe cells using an inverted epifluorescence microscope with adapted filters (i.e., an FITC filter for GFP detection). Maintain cells at 37 °C in the incubation chamber of the microscope. Record images with a 63×/1.4 NA

immersion objective. Cells can be counter-stained with Hoechst (0.1 μg/mL) to observe subcellular localization of the fusion protein.

3. For imaging of fixed cells, wash the cells with PBS, fix them with 4% PFA in PBS and again twice with PBS. Record images with a 63×/1.4 NA immersion objective. Although fixed cells can be stored for a few weeks at 4 °C, try to observe them 1–2 days after fixation. Cells can be counter-stained with antibodies to confirm expression of the target protein or for co-localization experiments.

A live cell image of U2-OS$^{::XPB\text{-}GFP}$ cells in which the XPB gene is fused to eGFP is shown in Fig. 6b.

4 Notes

1. Detailed information is available on the ATCC website (https://www.lgcstandards-atcc.org/). A guide containing general technical information for working with animal cells in culture, including media, sub-culturing, cryopreservation, and thawing is also accessible.
2. Affinity tags can be introduced at the N- or C- end of the target protein or inserted into internal loops. It is wise to perform extensive literature search for functionality studies and expression data in human cells as well as in other species such as mouse, *C. elegans*, or yeast to avoid trying to tag an extremity that would compromise protein function. In the absence of information on the accessibility of the extremities of the target protein, a C-terminal tag is generally preferred [28].
3. Nucleofection relies on permeabilization of the cellular membranes by electroporation. Although Amaxa nucleofection is often superior to other widely used transfection approaches, lipid-based transfection with reagents such as Lipofectamine 3000™ or jet-CRISPR™ can also be used.
4. Homology-based cloning techniques including sequence and ligation–independent cloning (SLIC) and commercially available technologies such as In-Fusion and Gibson assembly are perfectly adapted to assemble three to four DNA fragments in a single step reaction. These technologies are now routinely used in many laboratories (*see* ref. [29] in this issue) and can be used to rapidly generate the donor plasmid. The homology arms can be amplified from genomic DNA and the insert from plasmids which can be obtained from non-profit organizations (e.g., *see* Addgene plasmids #112848, #52379). Note that a growing number of companies also offer donor plasmid design and preparation as a service.

5. With the XPB-FW2/XPB-RV2 primers which hybridize on either side of the insertion site, both knock-in and non-modified alleles should be amplified: A 2800 bp PCR product is expected if the eGFP-P2A-PURO was successfully inserted while a 1300 bp fragment should be observed for a non-modified allele, allowing to discriminate between homozygous, heterozygous, and non-modified clones. With the XPB-FW/GFP-RV and Puro-FW/XPB-RV2 pairs which amplify 540 bp and 1000 bp fragments only modified alleles will be detected as the GFP-RV and Puro-RV2 oligonucleotides are complementary to the eGFP-P2A-PURO cDNA from the inserted sequence.

Acknowledgments

We thank IGBMC cell culture and imaging facilities for assistance and fruitful discussions. This work was supported by the Centre National pour la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), Université de Strasbourg (UdS), Association pour la Recherche sur le Cancer (ARC), the Ligue nationale contre le cancer, Institut National du Cancer (INCa; INCA 9378), Agence National pour la Recherche (ANR-12-BSV8-0015-01 and ANR-10-LABX-0030-INRT under the program Investissements d'Avenir ANR-10-IDEX-0002-02), Instruct (R&D Project Funding), and Instruct-ULTRA as part of the European Union's Horizon 2020 (Grant ID 731005) by Instruct-ERIC.

References

1. Villicana C, Cruz G, Zurita M (2014) The basal transcription machinery as a target for cancer therapy. Cancer Cell Int 14(1):18
2. Scott DE, Bayly AR, Abell C, Skidmore J (2016) Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. Nat Rev Drug Discov 15 (8):533–550
3. Bushweller JH (2019) Targeting transcription factors in cancer—from undruggable to reality. Nat Rev Cancer 19(11):611–624
4. Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346 (6213):1258096
5. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337 (6096):816–821
6. Dalvai M, Loehr J, Jacquet K, Huard CC, Roques C, Herst P, Cote J, Doyon Y (2015) A scalable genome-editing-based approach for mapping multiprotein complexes in human cells. Cell Rep 13(3):621–633
7. Dewari PS, Southgate B, McCarten K, Monogarov G, O'Duibhir E, Quinn N, Tyrer A, Leitner MC, Plumb C, Kalantzaki M, Blin C, Finch R, Bressan RB, Morrison G, Jacobi AM, Behlke MA, von Kriegsheim A, Tomlinson S, Krijgsveld J, Pollard SM (2018) An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein. elife 7
8. Geny S, Pichard S, Poterszman A, Concordet J (in press) Gene tagging with the CRISPR-Cas9 system to facilitate macromolecular complex purification. Methods Mol Biol

9. Compe E, Egly JM (2016) Nucleotide excision repair and transcriptional regulation: TFIIH and beyond. Annu Rev Biochem 85:265–290
10. Kolesnikova O, Radu L, Poterszman A (2019) TFIIH: a multi-subunit complex at the crossroads of transcription and DNA repair. Adv Protein Chem Struct Biol 115:21–67
11. Sandoz J, Nagy Z, Catez P, Caliskan G, Geny S, Renaud JB, Concordet JP, Poterszman A, Tora L, Egly JM, Le May N, Coin F (2019) Functional interplay between TFIIH and KAT2A regulates higher-order chromatin structure and class II gene expression. Nat Commun 10(1):1288
12. Yeh CD, Richardson CD, Corn JE (2019) Advances in genome editing through control of DNA repair pathways. Nat Cell Biol 21 (12):1468–1478
13. Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, Kuhn R (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol 33 (5):543–548
14. Sluch VM, Chamling X, Wenger C, Duan Y, Rice DS, Zack DJ (2018) Highly efficient scarless knock-in of reporter genes into human and mouse pluripotent stem cells via transient antibiotic selection. PLoS One 13(11):e0201683
15. Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM (2018) A survey of validation strategies for CRISPR-Cas9 editing. Sci Rep 8(1):888
16. Vouillot L, Thelie A, Pollet N (2015) Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. G3 (Bethesda) 5 (3):407–415
17. Renaud JB, Boix C, Charpentier M, De Cian A, Cochennec J, Duvernois-Berthet E, Perrouault L, Tesson L, Edouard J, Thinard R, Cherifi Y, Menoret S, Fontaniere S, de Croze N, Fraichard A, Sohm F, Anegon I, Concordet JP, Giovannangeli C (2016) Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. Cell Rep 14(9):2263–2272
18. Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 9(10):2395–2410
19. Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res 19 (7):1279–1288
20. Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ (2010) A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649:247–256
21. Brinkman EK, Chen T, Amendola M, van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res 42(22):e168
22. Paix A, Rasoloson D, Folkmann A, Seydoux G (2019) Rapid tagging of human proteins with fluorescent reporters by genome engineering using double-stranded DNA donors. Curr Protoc Mol Biol 129(1):e102
23. Paix A, Schmidt H, Seydoux G (2016) Cas9-assisted recombineering in C. elegans: genome editing using in vivo assembly of linear DNAs. Nucleic Acids Res 44(15):e128
24. Auer TO, Del Bene F (2014) CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. Methods 69(2):142–150
25. Liu Z, Chen O, Wall JBJ, Zheng M, Zhou Y, Wang L, Ruth Vaseghi H, Qian L, Liu J (2017) Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep 7(1):2193
26. Chen X, Zaro JL, Shen WC (2013) Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev 65(10):1357–1369
27. Koch B, Nijmeijer B, Kueblbeck M, Cai Y, Walther N, Ellenberg J (2018) Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing. Nat Protoc 13(6):1465–1487
28. Kimple ME, Brill AL, Pasker RL (2013) Overview of affinity tags for protein purification. Curr Protoc Protein Sci 73:19–23
29. Rossolillo P, Kolesnikova O, Essabri K, Ramon Zamorano G, Poterszman A (in press) Production of multiprotein complexes using the Baculovirus expression system: homology based and Restriction Free cloning strategies for construct design. Methods Mol Biol
30. Anders C, Niewoehner O, Duerst A, Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513(7519):569–573

Chapter 4

Validation of the Production of Antibodies in Different Formats in the HEK 293 Transient Gene Expression System

Jens König, Michael Hust, and Joop van den Heuvel

Abstract

Mammalian cells are the most commonly used production system for therapeutic antibodies. Protocols for the expression of recombinant antibodies in HEK293-6E cells in different antibody formats are described in detail. As model, antibodies against Kallikrein-related peptidase 7 (KLK7) were used. KLK7 is a key player in skin homeostasis and represents an emerging target for pharmacological interventions. Potent inhibitors can not only help to elucidate physiological and pathophysiological functions but also serve as a new archetype for the treatment of inflammatory skin disorders. Phage display-derived affinity-matured human anti-KLK7 antibodies were converted to scFv-Fc, IgG, and Fab formats and transiently produced in the mammalian HEK293-6E system. For the production of the corresponding antigen—KLK7—the baculovirus expression vector system (BEVS) and virus-free expression in Hi5 insect cells were used in a comparative approach. The target proteins were isolated by various chromatographic methods in a one- or multistep purification strategy. Ultimately, the interaction between anti-KLK7 and KLK7 was characterized using biolayer interferometry. Here, protocols for the expression of recombinant antibodies in different formats are presented and compared for their specific features. Furthermore, biolayer interferometry (BLI), a fast and high-throughput biophysical analytical technique to evaluate the kinetic binding constant and affinity constant of the different anti-KLK7 antibody formats against Kallikrein-related peptidase 7 is presented.

Key words Transient gene expression, Recombinant antibody formats, scFv, scFv-Fc, Fab, IgG, Biolayer interferometry, Kallikrein-related protease 7

1 Introduction

1.1 Aim of the Study

The functional and structural analyses of mammalian protein and protein complexes require simple, robust, and efficient protein production systems. Virus-free transient gene expression in insect cell lines and mammalian cell lines as well as the baculoviral expression vector system allows the purification of ample amounts of high-quality protein [1, 2]. Furthermore, adequate biochemical and biophysical methods are essential for the kinetic analysis of the binding affinity. Biolayer interferometry is a simple method

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_4,

which can be used in high throughput to determine the kinetic binding constant of interacting proteins [3].

Here, we describe the expression and binding studies of the Kallikrein-related protease 7 (KLK7) with improved inhibitory anti-KLK7 antibodies. The functional integrity of the skin epidermis is highly dependent on cutaneous protease activity. The human kallikrein-related peptidases (KLKs) are key players in skin homeostasis [4] and therefore important pharmacological therapeutic targets [5].

In our approach, genetically engineered antibodies are investigated for the potential as KLK inhibitors [6]. Biologics like recombinant antibodies have high specificity. Inhibitory antibodies have broad application and high acceptance in the community and industry as a therapeutic tool. Efficient screening for appropriate recombinant antibodies involves the small single-chain variable fragment format (scFv), which can be isolated using the in vitro phage display technology [7, 8].

For therapeutic use, however, the IgG molecule is still the most widely accepted format. It allows for bivalent binding and immune effector functions, along with a generally high stability, long serum half-life, and low immunogenicity [9–11]. Accordingly, fragments are routinely converted to IgG format after completed scFv selection and engineering. Nevertheless, there are other recombinant formats that should be considered depending on the application. Previous studies have shown that the conversion into the IgG format can affect antigen binding both positively and negatively [12, 13]. So, a structural compromise like the IgG-like scFv-Fc format might better preserve the binding properties of the selected scFv fragment. Due to their smaller size, recombinant antigen-binding fragments (like Fabs) of antibodies generally have better tissue penetration, which may be particularly beneficial in applications for passing the skin barrier [14].

There are a variety of expression systems available today to express the different antibody formats. Each has its own respective advantages in terms of cost, ease of use, and the post-translational modification profiles [15].

1.2 Expression of Different Recombinant Antibody Formats

Below, the set of expression vectors to generate different antibody formats and purification recombinant proteins expressed in HEK293-6E cells is described. The HEK293-6E cell line is especially qualified for fast and simple protein production as this cell line is compatible with the plasmid-based expression using polyethylenimine (PEI) [1, 2, 16]. The VH and VL coding fragments are cloned in specific vector pairs to be able to generate scFv, homodimeric scFv-Fc, heterodimeric Fab, or tetrameric IgG antibodies [9–11] (*see* Fig. 1)

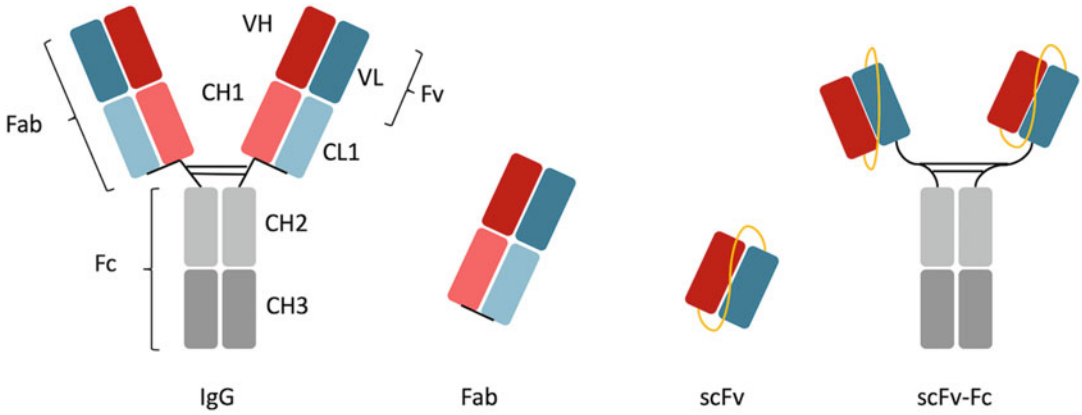


Fig. 1 Recombinant antibody formats compared to IgG. IgGs are large heterotetrameric molecules and depend on homodimerization of the Fc region (fragment crystallizable) of two identical heavy chains (HCs) and the subsequent assembly of two identical light chains (LCs) via disulfide linkages. The modular structure of multiple immunoglobulin domains allows for recombinant antibody formats with preserved antigen binding in a less complex dimeric (Fab, scFv-Fc) or single-chain structure (scFv). Red and grey: heavy chain; blue: light chain; orange: artificial peptide linker

1.3 Analysis of the Binding Kinetics of the Antibodies and KLK7 Using Biolayer Interferometry (BLI)

Binding affinities of antibodies to their target protein (antigen) can be determined by the dip and read biosensor technology of the OctetRed System. This technology uses biolayer interferometry (BLI) to measure protein–protein interactions [3]. The data are generated in real time, allowing kinetic analysis as well as quantitative analysis of bound protein. Here, we describe the application to determine the binding affinity K_D of immobilized biotinylated-KLK7 to streptavidin sensors to the different antibody formats. After initial equilibration of the biosensor, the ligand (KLK7) is tightly bound to the tip. Binding of protein results in a change of the thickness of the biolayer at the tip of the biosensor. The increase in thickness of the biolayer results in a change of the interference between two reflected beams from the internal reference layer and the surface of the tip used for specific binding to the protein samples. The binding of protein is detected and reported as a change in wavelength. In this way, the association and dissociation of the antibody to the immobilized KLK7 can be followed in time for a range of concentrations using parallel tips. The kinetic constants can be calculated from these curves presuming a 1:1 interaction stoichiometry, which allows proper fitting of the curves.

2 Materials

In the first part, we present an overview of the vectors used, to enable the expression of the scFv fragment derived from the phage display analysis into different antibody formats. Furthermore, we

describe the necessary materials needed for fast and efficient protein expression of the materials required for kinetic analysis of the binding of the different antibody format to the KLK7 target protein by biolayer interferometry (BLI).

2.1 Expression Vectors

All cloning steps were performed using standard restriction enzyme digestion with subsequent ligation. Alternatively, the sequence- and ligation-independent cloning (SLIC) technique according to Li et al. [17] was used to generate the expression clones. DNA preparation for transient gene expression was done using the PureYield Plasmid Midiprep kit (Promega) or the endotoxin-free Plasmid DNA Gigaprep kit (Qiagen). The construction of the cloning vectors is described in detail in Jäger et al. [16].

1. *pCSE2.6-hIgG1-Fc-LUP37-A10* (Expression vector for the synthesis of the antibody LUP37-A10 in the scFv-Fc format).
2. *pCSE2.6-hIgG1-Fc-LUP37-B10* (Expression vector for the synthesis of the antibody LUP37-B10 in the scFv-Fc format).
3. *pCSE2.6-hIgG1-Fc-LUP37-C11* (Expression vector for the synthesis of the antibody LUP37-C11 in the scFv-Fc format).
4. *pCSE2.6-hIgG1-Fc-LUP37-D11* (Expression vector for the synthesis of the antibody LUP37-D11 in the scFv-Fc format).
5. *pCSL3l* (Expression vector for the synthesis of antibody light chains. VL domains were cloned into this vector for production of IgG and Fab antibodies).
6. *pCSEH1c* (Expression vector for the synthesis of antibody heavy chains. VH domains were cloned into this vector for production of IgG antibodies).
7. *pCSE2.5-Fab-h-HIS-XP* (Expression vector for the synthesis of antibody heavy chains lacking the domains CH2 and CH3. VH domains were cloned into this vector for the production of antibodies in the Fab format).
8. *pUC57-His8-proKLK-3C-hKLK7* (Cloning vector obtained from GenScript carrying the synthetic codon optimized DNA sequence His8-proKLK-3C-hKLK7).
9. *pOpIE2- His8-proKLK-3C-hKLK7* (The sE2-dHVR1 sequence of pOpIE2-sE2-dHVR1 was exchanged with the insert His8-proKLK-3C-hKLK7 to get a Hi5 expression vector for human proKLK7).
10. *pTTo/eGFPq* (Control vector for the determination of transfection efficiency in HEK293-6E cells via flow cytometry).
11. *pOpiE2-eGFP-HA* (Control vector for the determination of transfection efficiency in Hi5 cells via flow cytometry).

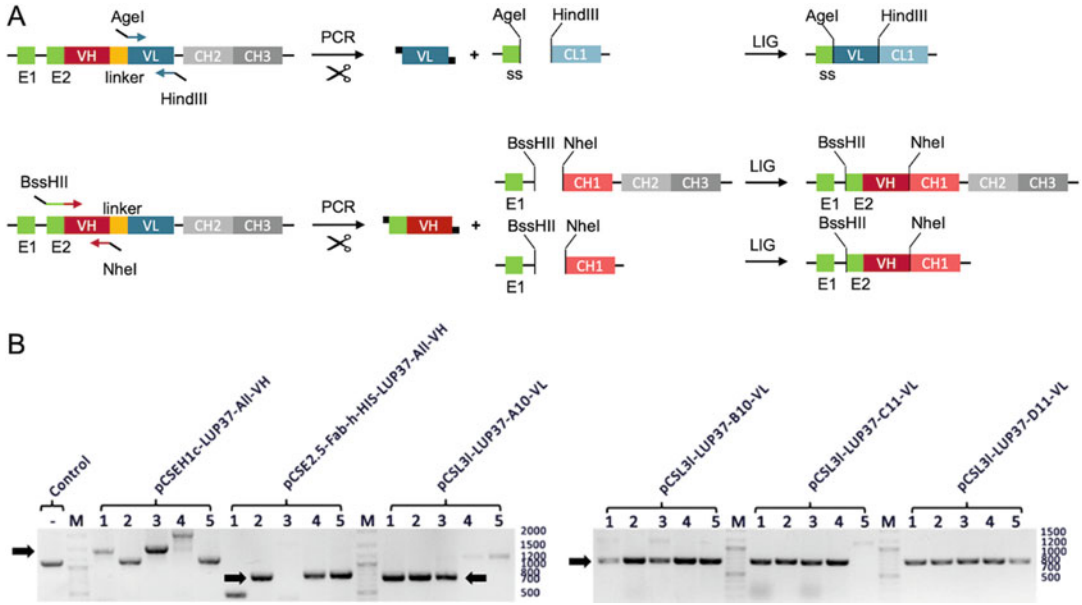


Fig. 2 Vector cloning for transient scFv-Fc, IgG, and Fab expression in HEK293-6E. (**a**). Schematic overview of the cloning strategies for the generation of pCSL3I-LUP37-A10-VL, pCSL3I-LUP37-B10-VL, pCSL3I-LUP37-C11-VL, pCSL3I-LUP37-D11-VL, pCSEH1c-LUP37-All-VH, and pCSE2.5-Fab-h-HIS-LUP37-All-VH ("All" in the plasmid names means that all four different LUP isolates (A10, B10, C11 and D11) were cloned in the same way) starting from the corresponding pCSE2.6-hIgG1-Fc-LUP37 vectors. Shown are the expression cassettes consisting of signal peptide (green), Ig domains, and artificial peptide linker (orange), as well as the particular restriction sites used for cloning. (**b**) Analysis of colony-PCR products via agarose gel electrophoresis. Black arrows indicate the expected lengths of positive amplificates

The transfer of the scFv gene fragment into the vector pCSE2.6-hIgG1-Fc-XP [16, 18] can be done by simple restriction digestion and ligation using the isolated scFv phage DNA from the library. However, for the generation of the other formats, the VH and VL DNA fragments have to be individually isolated by PCR amplification with the addition of new compatible cloning sites. The successful cloning is shown in Fig. 2.

2.2 Cell Lines for Protein Production

1. The HEK293-6E cell line was licensed from National Research Council (NRC), Biotechnological Research Institute (BRI), Montreal, Canada.
2. The *Trichoplusia ni* Hi5 insect cell line (BTI-Tn-5B1-4) was isolated by the *Boyce Thompson Institute* for Plant Research, Ithaca, USA.
3. The *Spodoptera frugiperda* Sf21 (DSMZ #ACC119) insect cell line (IPLB-Sf21-AE) was used for amplification of the baculovirus.

2.3 Cell Culture Conditions

1. 125 mL up to 1 L polycarbonate shake flasks (Corning).
2. Climo-shaker ISF1-XC CO_2 orbital shaker with 50 mm orbit (Kühner).
3. Complete cultivation medium for HEK293-6E: FreeStyle F17 (Gibco, Life Technologies/Thermo Scientific) supplemented with 7.5 mM glutamine, 0.1% Pluronic-F68, and 25 μg/mL G418. Prepare by adding the following components to 1 L fresh F17 medium: 40 mL of 200 mM stock solution of L-glutamine, 0.5 mL of 50 mg/mL G418-solution, and 10 ml of 10% (w/v) Pluronic-F68 (20 g in 200 mL of water and filter through 0.2 μm). Mix well and store at 4 °C.
4. Complete cultivation medium for Hi5 insect cells: EX-CELL 405.
5. Complete cultivation medium for Sf21 insect cells: EX-CELL 420.

2.4 Transfection Reagents and Additional Chemicals for Protein Production

Prepare all solutions using ultrapure water and analytical grade reagents. All reagents will be sterile filtered and stored at 4 °C (unless otherwise mentioned).

1. 1 mg/mL 25 kDa Polyethylenimine (PEI_25), linear (Polysciences #23966): Dissolve 0.05 g of PEI in 50 mL of water and heat until it is completely dissolved. Store aliquots at −70 °C.
2. Use 15-mL polystyrene tubes for preparing DNA/PEI_25-complexes.
3. 20% (w/v) Tryptone N1 (Organotechnie S.A.S., La Courneuve, France): Weigh 100 g Tryptone N1. Make up to 500 mL with water.
4. 300 g/L Glucose: Weigh 150 g glucose and make up to 500 mL with water.
5. 75 mM Valproic acid (20×). Dissolve 1.2 g of valproic acid sodium salt (MW 155.19) and make up to 100 mL in F17 medium (supplemented).
6. Guava EasyCyte™ Mini (Merck, Millipore).
7. CASY Cell Counter (Innovatis).

2.5 Biolayer Interferometry

Prepare all solutions using ultrapure water and analytical grade reagents. All reagents are sterile filtered and stored at 4 °C (unless otherwise mentioned).

1. 1 mM EZ-Link Sulfo-NHS-LC-LC-Biotin solution (Thermo Fisher).
2. Zeba spin desalting columns (Thermo Fisher).
3. OctetRed 96 System (Pall, Fortebio).

4. Streptavidin (SA) Dip and Read Biosensors (Pall, Fortebio).
5. 96-well flat bottom polypropylene microplate, black.
6. PBS (1×): 138 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2mM KH_2PO_4, pH7.4).
7. Kinetics buffer (1× PBS, pH 7.4) optionally with 0.1% BSA and 0.2% Tween 20.
8. 100 mM 10× glycine buffer pH 1.5.

3 Methods

Carry out all procedures at room temperature unless otherwise specified. Use sterile equipment, flasks, and centrifuge tubes throughout the experiment.

3.1 Transient Protein Expression in HEK293-6E Cells

The protocol is described in Karste et al. [1]. The employed vectors contain the OriP to ensure optimal long-term expression levels. Here the expression is described for 500 mL.

1. *[Day 3, e.g., Friday]* Prepare a 400 mL culture containing 0.3 × 10^6 HEK293-6E c/mL. Incubate the culture for 72 h at 37 °C, 110 rpm, and 5% CO_2.
2. *[Day 0, e.g., Monday]* Count the cells of the preparatory culture and prepare a 250 mL culture containing 1.8 × 10^6 HEK293-6E c/mL by centrifuging the required volume of the cell suspension at 180 × *g* for 4 min. Discard the supernatant carefully and resolve the cell pellet in 250 mL fresh F17. Place the cells back into the incubator (37 °C, 90 rpm, and 5% CO_2) until use (*see* **Note 1**).
3. Prepare the DNA solution in 12.5 mL F17 in a 50-mL corning flask (*see* **Note 2**). Hereto, mix 237 μg of the target plasmid (s) containing your gene(s) of interest with 13 μg control plasmid pTTo-GFPq coding, e.g., for eGFP. Mix by agitating the tube briefly.
4. Prepare the PEI_25 solution in 12.5 mL F17 in a 50-mL Corning flask (*see* **Note 2**). Hereto, pipette 625 μL PEI of the 1 mg/mL PEI_25 stock solution into the. Mix by vortexing briefly.
5. Mix the PEI_25 solution with the DNA solution directly (*see* **Note 3**).
6. Incubate the PEI_25–DNA solution for 15 min at RT to preform PEI–DNA complexes.
7. Pipette the PEI–DNA complexes (25 mL) to the cells and mix gently. Incubate the cells until harvest at 37 °C, 90 rpm, and 5% CO_2.

8. *[Day 1–7, e.g., Tuesday to Monday]* Take samples daily and determine transfection efficiency in the flow cytometer (Guava EasyCyte) or/and determine target protein expression by a suitable technique (SDS-PAGE, slot blot, or Western blot).
9. *[Day 2, e.g., Wednesday]* Add 250 mL fresh F17 to the cells as well as 12.5 mL of the TN1-Stock (20%) = 0.5%.
10. *[Day 3, e.g., Thursday]* Add 7.5 mL of the glucose stock (300 g/L) = 4.5 g/L.
11. *[Day 4, e.g., Friday]* Add 25 mL of the valproic acid stock (75 mM) = 3.75 mmol/L.
12. *[Day 7, e.g., Monday]* Harvest secreted target proteins by centrifuging the cell suspension for 15 min at 1000 × *g* and store the supernatant at 4 °C until purification. Depending on the protein, it might be required to add protease inhibitor. For intracellular proteins, snap freeze the cell pellet and store at −80 °C until cell lysis and purification.

The cloned expression vectors were transfected single or in pairs into HEK293-6E cells, and the culture was grown for 168 h. The viability of the cells remained >80%, and the efficiency of transfection varied from 20% to 80% for the different experiments (*see* Fig. 3). All the supernatants contained sufficient antibody for subsequent purification of the protein samples. The quality of the purified antibodies was validated by SDS-PAGE. Total volumetric yield and specific cellular yield of the recombinant antibodies are presented in Fig. 4.

3.2 Transient Protein Expression in Hi5 Cells

The protocol is described in Karste et al. MiMB [1]. Here, we show the expression and purification of the secreted KLK7 in Hi5 cells in 500 mL scale. The employed vectors contain the optimal expression cassette with the strong OpIE2 promoter (*see* **Note 4**, [20]).

1. *[Day 1, e.g., Friday]* Prepare a 150 mL preculture starting at 0.3–0.4 × 10^6 Hi5 c/mL. Incubate the culture for 72 h at 27 °C and 110 rpm (*see* **Note 5**).
2. *[Day 0, e.g., Monday]* Count the cells of the preparatory culture and prepare a 100 mL culture containing 5 × 10^6 Hi5 c/mL by centrifuging the required volume of the cell suspension at 180 × *g* for 4 min. Discard the supernatant and resolve the cell pellet in 100 mL fresh EX-CELL 405.
3. Incubate the culture at 27 °C and 110 rpm for 1 h.
4. Mix 425 µg of the expression plasmid and 25 µg of the control plasmid, e.g., pOpIE2-eGFP-HA, coding for eGFP (*see* **Note 6**).

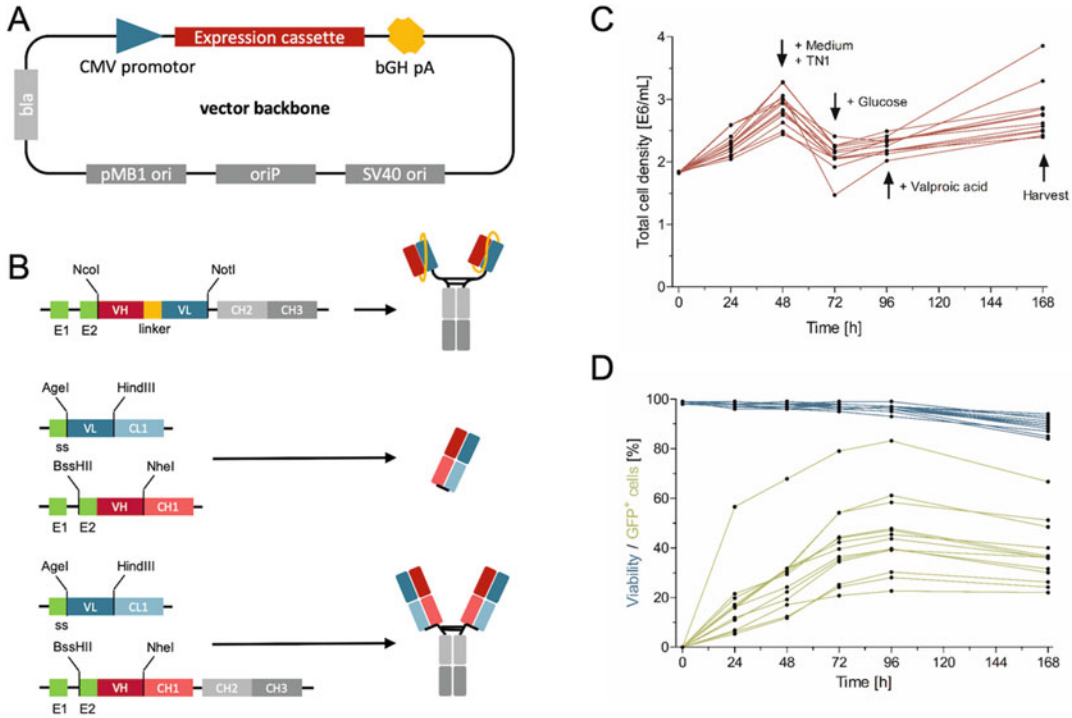


Fig. 3 Transient antibody production in HEK293-6E cells. (**a**) Schematic vector map of the mammalian HEK293-6E expression vector with exemplary expression cassette (found in pTT5 or pFlpBtMll [2]). (**b**) Schematic illustration of corresponding gene expression cassettes used for the production of recombinant antibodies in the scFv-Fc, Fab, and IgG format. (**c**/**d**) Monitoring of growth, viability, and transfection efficiency (GFP-positive cells) over time. Volume doubling and supplementations at 48, 72, and 96 h post-transfection are indicated by black arrows. Mean values are presented

5. Pipette the DNA mix directly to the prepared cells and mix gently (*see* **Note 7**).
6. Immediately pipette 4 mL PEI of the 1 mg/mL PEI stock solution to the cells and mix gently.
7. Incubate the culture at 27 °C and 110 rpm for 3 up to 5 h.
8. Add 400 mL fresh EX-CELL 405 (*see* **Note 8**).
9. *[Day 1–3, Tuesday to Thursday]* Take samples daily, monitor cell numbers and viability, and determine transfection efficiency in the cytometer. If cells reach a density above 3×10^6 c/mL, adjust the concentration back to 2×10^6 c/mL by adding fresh culture medium.
10. *[Day 3, Thursday]* Harvest secreted target proteins by centrifuging the cell suspension 15 min at 1000 × *g* and store the supernatant at 4 °C after sterile filtration (0.45 μm) until purification.

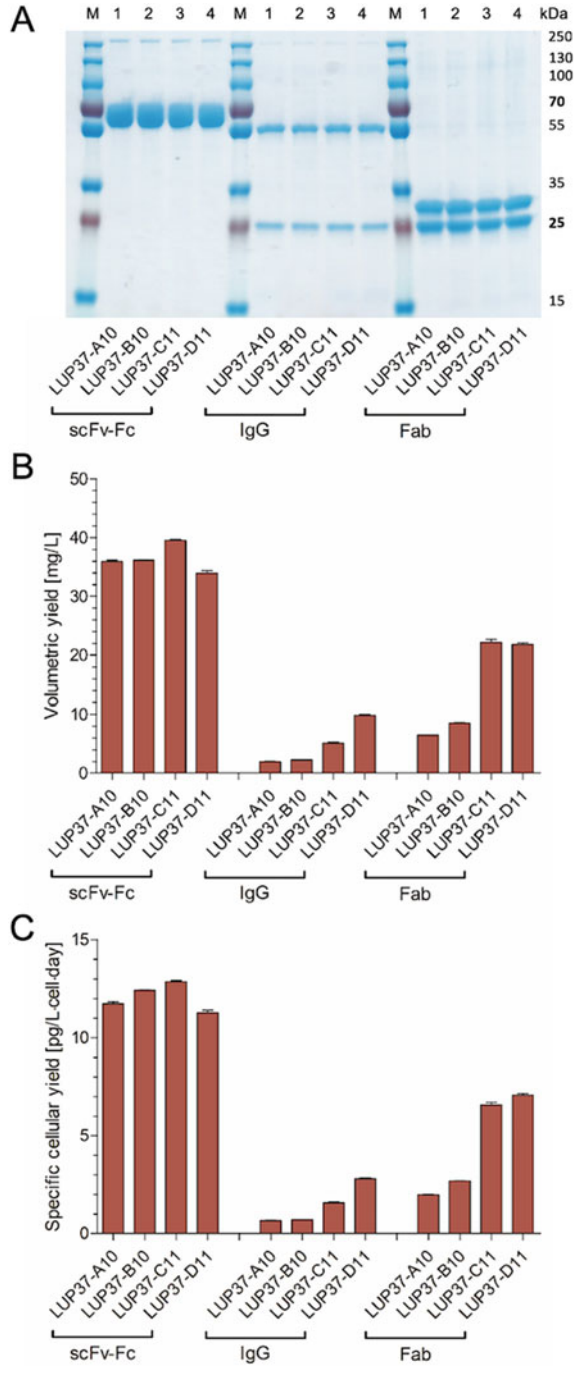


Fig. 4 Yields of antibody production in HEK293-6E cells. (**a**) Stained protein gel of purified scFv-Fc, IgG, and Fab antibodies. Samples of 3.5 μg/lane were separated for 50 min on a 12% SDS-PAGE gel at 160 V and analyzed via Coomassie staining. (**b**, **c**) Column charts of volumetric and specific cellular recombinant protein yields with SEM error bars. Yield was calculated on the basis of the total amount of purified protein and final production volume. For the calculation of the specific cellular yield, maximal cell number and cultivation time were additionally considered

3.3 Baculoviral Expression of hKLK7 in Hi5 Cells

The protocol describes the expression and purification of the secreted KLK7 in Hi5 cells in 500 mL scale using the BEVS technology (*see* **Note 9**).

1. *[Day 3, Friday]* Prepare a 250 mL preculture starting with 0.3–0.4 $\times$ 10^6 Hi5 c/mL. Incubate the culture for 72 h at 27 °C and 110 rpm (*see* **Note 5**).
2. *[Day 0, e.g., Monday]* Count the cells of the preparatory culture and prepare a 500 mL culture containing 1.0–1.5 $\times$ 10^6 Hi5 c/mL by diluting the required volume of the cell suspension into fresh EX-CELL 405 medium. Infect the cells at a MOI of 0.5–2.0 with the baculoviral stock solution (*see* **Note 10**). Mix gently and incubate the culture at 27 °C and 110 rpm for 72 h.
3. *[Day 1–3, Tuesday to Thursday]* Take samples daily, monitor cell numbers, viability, and cell diameter in the CASY Cell Counter to determine the progress of the baculoviral infection. If cells reach a density above 3 $\times$ 10^6 c/mL, adjust the concentration back to 1 $\times$ 10^6 c/mL by adding fresh EX-CELL 405 culture medium.
4. *[Day 3, Thursday]* Harvest secreted target proteins 48 h after total proliferation stop by centrifuging the cell suspension for 15 min at 1000 $\times$ *g* and store the supernatant at 4 °C after sterile filtration (0.45 μm) until purification. Prevent microbial growth by adding 0.05% sodium azide.

3.4 Preparation of the Baculoviral Stock Solution

1. *[Day 3, Friday]* Prepare a 150 mL Sf21 preculture starting with 0.3–0.4 $\times$ 10^6 c/mL in EX-CELL 420 medium. Incubate the culture for 72 h at 27 °C and 110 rpm (*see* **Note 5**).
2. *[Day 0, e.g., Monday]* Count the cells of the preparatory culture and prepare a 500 mL culture containing 0.5 $\times$ 10^6 Hi5 c/mL by diluting the required volume of the cell suspension into fresh EX-CELL 420 medium. Infect the cells with a MOI of 0.2 with the baculoviral transfection supernatant or use 5% v/v of a nontitered viral stock. Mix gently and incubate the culture at 27 °C and 110 rpm for 4 days.
3. *[Day 1–4, Tuesday to Friday]* Monitor the cell number, viability, and the cell diameter (CASY cell counter). The cells should be split to 1 $\times$ 10^6 c/mL by adding fresh EX-CELL 420 medium once they reach 2 $\times$ 10^6 c/mL.
4. *[Day 4, e.g., Friday]* The viral supernatant was harvested 48 hours after an increase in cell diameter of 3 μm was observed.

3.5 Protein Purification by Affinity Chromatography

Purification and maintenance of the columns and chromatography systems were done according to the protocols recommended by the manufacturing company.

1. *[HisTrap Excel column chromatography]* The His-tagged Fab antibodies and proKLK7 were isolated using the ÄKTA start system from (GE Healthcare) using HisTrap Excel columns for direct loading of cell culture supernatants. The pre-filtered (0.45 μm) supernatants were loaded onto a 1 mL HisTrap Excel column equilibrated with buffer A (20 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl). Washing was done using buffer A containing 10–30 mM imidazole. The protein was eluted using buffer A containing 250 mM imidazole. Eluted fractions were analyzed by SDS-PAGE electrophoresis. Protein-containing fractions were pooled and dialyzed against PBS using Spectra/Por dialysis tubing (Spectrum, MWCO 6-8000) to remove imidazole from the sample. The purification of KLK7 from virus-free transient gene expression in Hi5 resulted in higher amounts and higher purity of recombinant protein compared to the BEVS expressed KLK7. (*see* Fig. 5).
2. *[Protein A column chromatography]* The semi-automated Profinia Protein purification system was used for the purification of the Fc-tag containing recombinant antibody formats (IgG, scFv-Fc) using the customized "Protein A and Desalting"

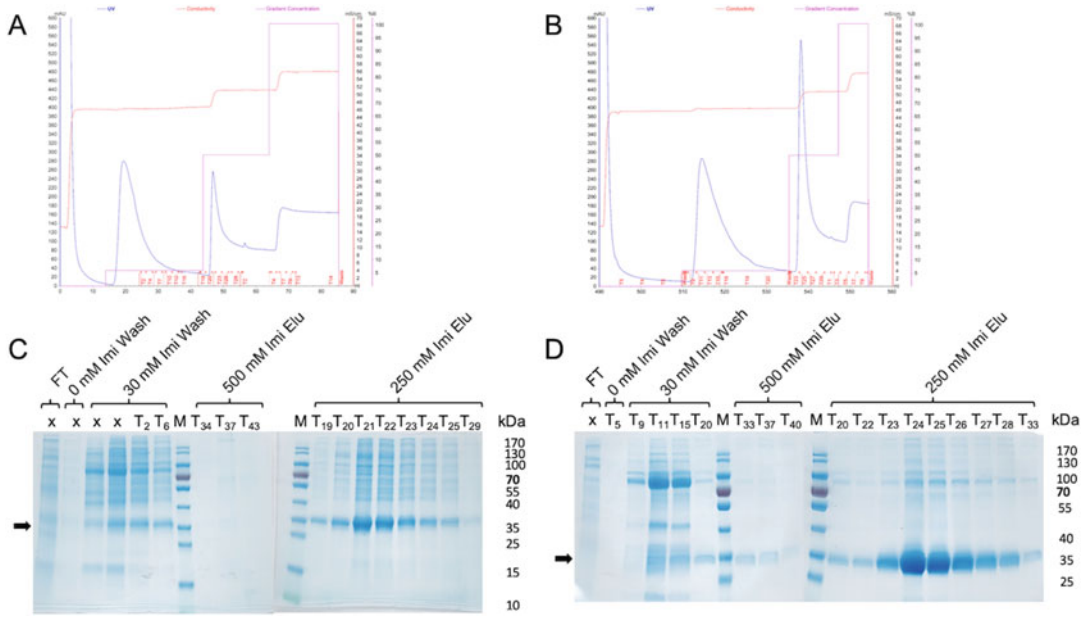


Fig. 5 Affinity chromatography. The cell culture supernatants were filtrated (0.45 m) and loaded onto the HisTrap Excel column. (**a**, **b**) Comparison of the chromatograms of KLK7 affinity purification of the production via BEVS and virus-free expression in Hi5 cells (**c**, **d**, respectively). Peak fractions were detected by SDS-PAGE. 20 μL samples of the collected 1 mL fractions were applied per lane of a 12% SDS-gel and separated for 50 min at 160 V. Black arrows indicate the bands resembling pro-KLK7. The imidazole gradient is shown on the right axis in % B (100% B corresponds to 500 mM Imidazole)

method. The pre-filtered (0.45 μm) supernatants were loaded in manual mode (for volumes >50 mL) onto a 1 mL pre-packed protein A cartridge. Following washing, the antibodies were eluted by pH shift and directly loaded onto a 10 mL P6 cartridge for desalting. The protein fraction was stored at 4 °C for further analysis

3.6 Biolayer Interferometry

Protocols were set up using the Acquisition and Analysis User Guide for the OctetRed 96 (Fortebio, Pall) and according to the methods described in detail in [19].

1. Purified pro-KLK7 samples were activated for 16 h at 4 °C using 0.02 U/g bovine Enterokinase (Sigma).
2. The activated KLK7 was biotinylated using 1 mM EZ-link Sulfo-NHS-LC-LC-Biotin (Thermo Fisher). The protein was cross-linked in at molar ratio of 3:1 (linker: KLK7) for 30 min at room temperature.
3. Non-reacted biotinylating reagent was removed by gel filtration in a Zeba spin desalting column (Thermo Fisher) according to the manufacturer's protocol. The biotinylated KLK7 was diluted to 12.5 μg/mL. Antibodies were diluted in a range of concentration from 1 to 500 nM. The sample plate was prepared as described in Fig 6. Each well was filled with 200 μL of buffer or sample.
4. Prepare a plate with eight streptavidin biosensors in row 1 and preincubate the tips in 200 μL PBS for 10 min. Start the acquisition software (8.2) of the OctetRed 96 and set up and kinetic experiment with regeneration of the tips according to the Octet Data Acquisition User Guide. Detailed protocols are described by Kamaraswamy and Tobias [19].
5. The setup for the biolayer interferometric kinetic measurements is described in detail in Table 1. Preincubated streptavidin biosensors were equilibrated for 60 s in PBS before the biotine-KLK7 was immobilized on the sensor for 300 s. Baseline, association, and dissociation to the antibody were monitored over time as shown in Table 1. The tips were subsequently regenerated by a threefold cycle of regeneration in 10 mM 1× glycine buffer, pH 1.5 followed by neutralization in PBS. The tips were reused to measure the remaining KLK7—anti-KLK7 antibody kinetics.
6. The data were processed using the Octet Data Analysis software package (8.2). The association step was aligned to the baseline. The relative signal of antibody binding to the immobilized KLK7 is shown in Fig 7. Make sure that the antibody does not show nonspecific binding to the tips by monitoring the signal of a tip which is not charged with biotinylated KLK7.

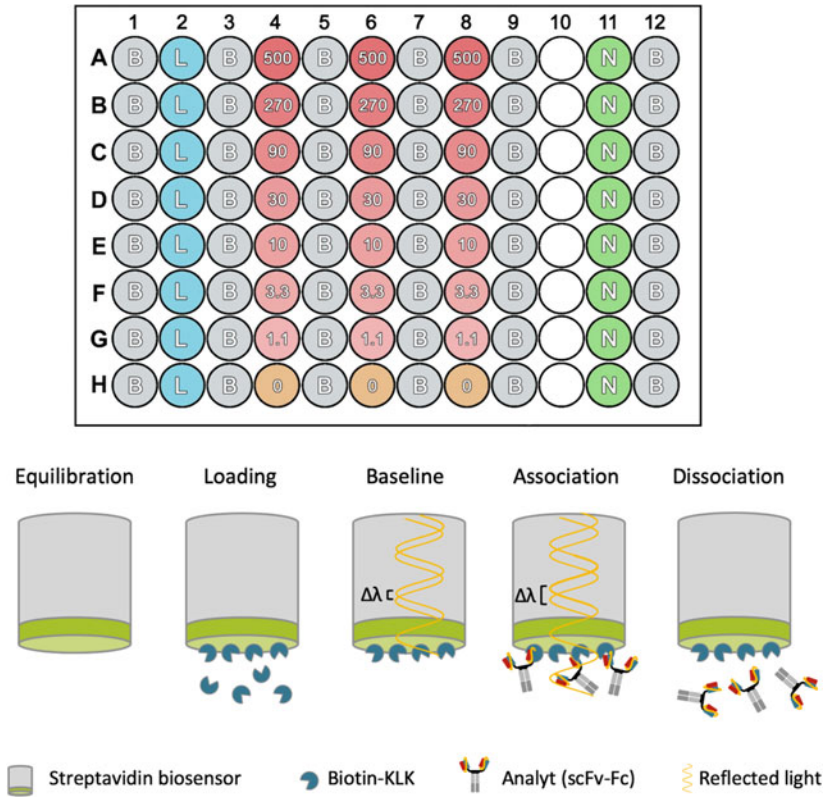


Fig. 6 Plate map diagram and principle of KLK7—anti-KLK7 kinetic measurements. Plate map: column 1, 3, 5, 7, 9, 12 (gray = PBS), column 2 (blue = Biotin-KLK7 (12.5 μg/mL)), column 6, 8, 10 (red = antibody dilution series (1.1-500 nM)), well H4, H6, H8 (orange = reference for subtraction of the background), column 11 (green = glycine buffer for regeneration). Biolayer interferometry principle: Top-emitting light is reflected at the internal reference layer and the outer reflection surface of the biosensor tip. Binding molecules increase the optical thickness and cause a shift of the resultant wave due to changes in the wave interference pattern. This shift in wavelength is measured in real-time

Table 1
Settings for the biolayer interferometric kinetic measurements

				Sample plate column		
Step	**Step name**	**Time [s]**	**Flow [rpm]**	**Run 1**	**Run 2**	**Run 3**
1	Equilibration	60	1000	1	1	1
2	Loading	300	1000	2	2	2
3	Baseline	60	1000	3	3	3
4	Association	600	1000	4	6	8
5	Dissociation	600	1000	5	7	9
6	Regeneration	30	1000	11,12	11,12	11,12

The experiment was repeated after regeneration of the tips using the different antibody formats in column 4, 6, and 8 respectively

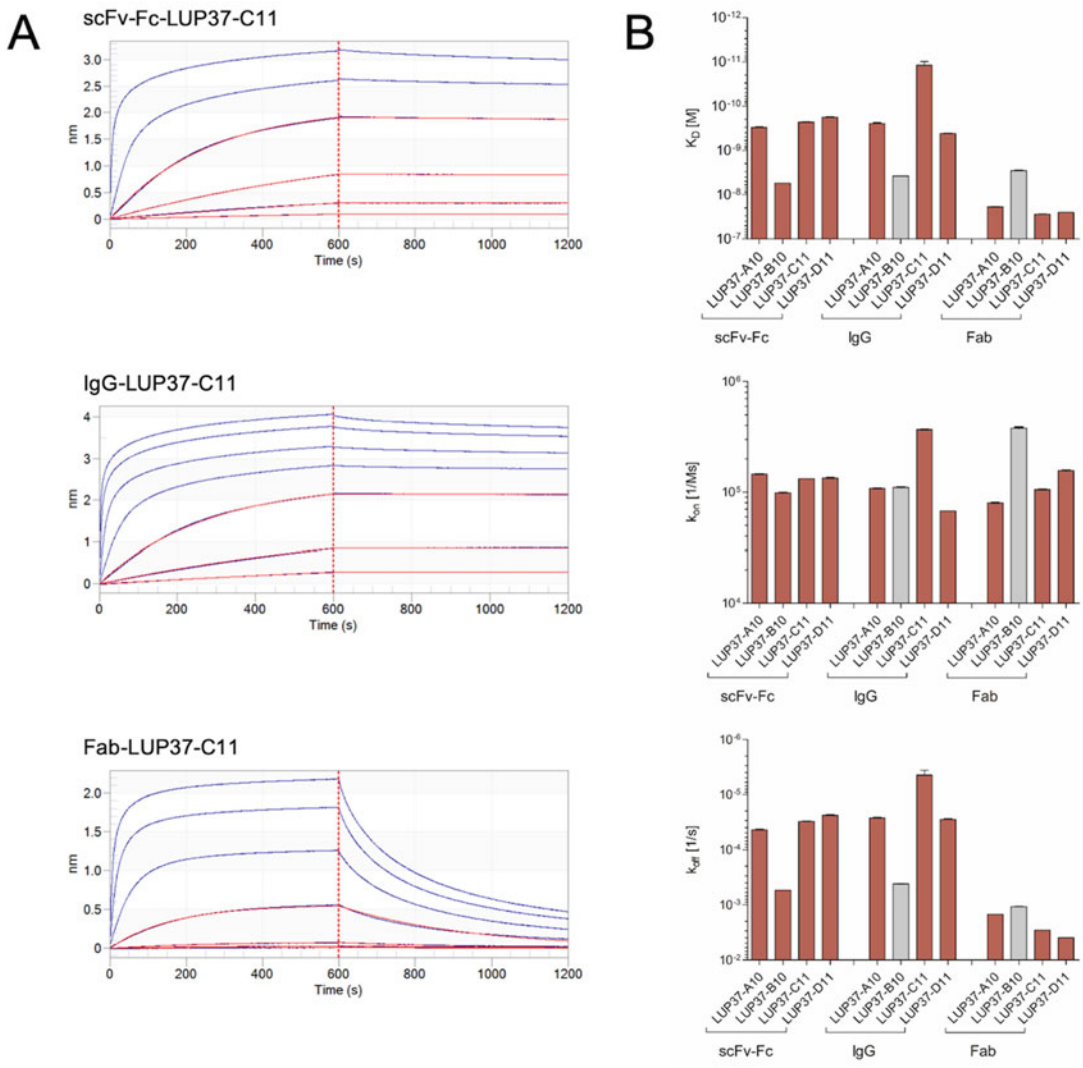


Fig. 7 Biolayer interferometry-based affinity ranking for recombinant scFv-Fc, IgG, and Fab antibodies. Streptavidin Dip and Read tips were loaded with biotinylated KLK7 and assayed for their response to various anti-KLK7 antibody levels. (**a**) Nonlinear fit of grouped association and dissociation curves based on a 1:1 binding model for the three recombinant antibody formats of LUP37-C11. (**b**) Summary of apparent affinity and rate constants in a grouped column chart with error bars representing the SEM. Antibodies with insufficient fits (X^2: > 3 and R^2: < 0.95) are displayed in gray

7. Select parameters for the kinetic analysis and determination of the affinity constant as described in [19]. After curve-fitting, the lower concentrations are used for determining the kinetic and affinity constants (Table 2). The results shown in Table 2 indicate that the apparent affinities of the three scFv-Fc antibodies (A10, C11 and D11) are in a similar middle picomolar range. Conversion into the IgG format showed similar affinities for the antibodies, only C11 improved a factor of 19 in affinity.

Table 2
Summary of antibody yields and apparent affinity to KLK7

Antibody	Volume yield (mg/L)	K_D (nM)	k_{on} (1/M·s)	k_{off} (1/s)
scFv-Fc-LUP37-A10	7.36	0.304	1.45×10^5	4.39×10^{-5}
scFv-Fc-LUP37-B10	7.42	5.53	9.84×10^4	5.44×10^{-4}
scFv-Fc-LUP37-C11	8.10	0.231	1.32×10^5	3.05×10^{-5}
scFv-Fc-LUP37-D11	6.96	0.179	1.34×10^5	2.39×10^{-5}
IgG-LUP37-A10	0.40	0.247	1.08×10^5	2.67×10^{-5}
IgG-LUP37-B10	0.47	–	–	–
IgG-LUP37-C11	1.05	0.012	3.68×10^5	4.43×10^{-6}
IgG-LUP37-D11	2.02	0.419	6.74×10^4	2.82×10^{-5}
Fab-LUP37-A10	1.33	18.8	7.95×10^4	1.50×10^{-3}
Fab-LUP37-B10	1.75	–	–	–
Fab-LUP37-C11	4.55	27.8	1.05×10^5	2.91×10^{-3}
Fab-LUP37-D11	4.47	25.3	1.56×10^5	3.96×10^{-3}

However, the Fab format substantially reduced the affinity by two orders of magnitude. On account of the chosen orientation of the assay, the affinity constant determination can be affected by avidity effects. *See* **Note 11**.

4 Notes

1. Replacing used culture medium for fresh F17 (without antibiotic), ensuring a sufficient amount of nutrients to keep the cells growing until the addition of fresh culture medium at day 2.
2. Using polystyrene tubes instead of polypropylene tubes is important to ensure that the PEI-DNA complexes do not stick to the tube what would lead to a lower transfection rate due to a lower total amount of available complex.
3. Directly mix the DNA with PEI to form stable PEI-DNA complexes. Be sure that the DNA is not precipitating. Turbidity or visible aggregates indicate that the quality of the DNA preparation is not sufficient. Make sure the DNA sample is not contaminated with chromosomal *E. coli* DNA.
4. The optimal combination in our facility comprises the OpIE2 promoter, the IE1 terminator, and a FlashBac compatible backbone [20].

5. This step must ensure that the cells are still in the exponential growth phase, and this will increase the transfection efficiency, baculoviral infection, and protein expression vastly.
6. Using higher amounts of DNA (up to 3 μg per 1×10^6 cells) leads to higher yields if the PEI amount is increased accordingly and the PEI:DNA ratio of 4:1 is kept stable. Further increasing the amount of DNA will decrease cell viability due to the required higher PEI concentration, as PEI is known to be cytotoxic in high concentrations. This will decrease the overall productivity.
7. Direct transfection without preforming the DNA-PEI complexes led to the highest transfection rates and protein levels in our lab. In contrast, Shen et al. observed no differences between direct transfection or transfection after preforming the DNA-PEI complexes [21]. The difference might be caused by different performance of other batches of Hi5 cells. However, a direct transfection is easier and faster. Thus, the described direct transfection is the optimal method in our hands.
8. Diluting with fresh media is important to keep the cells in the exponential growth phase until the time of harvest.
9. Optimized protocols are validated and reviewed in the benchmarking study by Stolt-Bergner et al. [22].
10. For optimal expression results, the titer of the viral stock solution should be determined by a plaque assay. Additionally, the best MOI for expression should be determined in a small-scale expression analysis prior to scaling up by varying the MOI from 0.2 to 2.0.
11. Avidity effects can be caused by a multivalent molecule, like an antibody, when the antigen is immobilized. For accurate determination of the kinetic constants, the orientation of the assay must also be reverted by using anti-human IgG Fc Capture (AHC) biosensors for the immobilization of the IgG and Fc-tagged antibody format using KLK7 as analyte. The Fab format cannot be analyzed in this way and needs direct chemical cross-linking to suitable tips.

Acknowledgments

We greatly thank Luciano Puzer from the Universidade Federal do ABC, Sao Paulo, Brazil for his valuable scientific input and the biomaterials he made available for this research project. We also thank Daniela Gebauer, Nadine Konisch, and Anke Samuels for their excellent technical support during the experiments. This work was supported by the Helmholtz Protein Sample Production Facility (PSPF).

References

1. Karste K, Bleckmann M, van den Heuvel J (2017) Not limited to E. coli: versatile expression vectors for mammalian protein expression. Methods Mol Biol 1586:313–324
2. Meyer S, Lorenz C, Baser B et al (2013) Multihost expression system for recombinant production of challenging proteins. PLoS One 8 (7):e68674
3. Kamat V, Rafique A (2017) Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. Anal Biochem 536:16–31
4. Egawa G, Kabashima K (2018) Barrier dysfunction in the skin allergy. Allergol Int 67:3–11
5. Sotiropoulou G, Pampalakis G (2012) Targeting the kallikrein-related peptidases for drug development. Trends Pharmacol Sci 33:623–634
6. Goettig P, Magdolen V, Brandstetter H (2010) Natural and synthetic inhibitors of kallikrein-related peptidases (klks). Biochimie 92:1546–1567
7. Russo G, Meier D, Helmsing S et al (2018) Parallelized antibody selection in microtiter plates. Methods Mol Biol 1701:273–284
8. Frenzel A, Hust M, Schirrmann T (2016) Phage display-derived antibodies in clinical development and therapy. mAbs 8:1177–1194
9. Strohl WR (2018) Current progress in innovative engineered antibodies. Prot Cell 9:86–120
10. Frenzel A, Hust M, Schirrmann T (2013) Expression of recombinant antibodies. Front Immunol 4:217
11. Ha J-H, Kim J-E, Kim Y-S (2016) Immunoglobulin Fc heterodimer platform technology: From design to applications in therapeutic antibodies and proteins. Front Immunol 7:394
12. Thie H, Toleikis L, Li J et al (2011) Rise and fall of an anti-MUC1 specific antibody. PLoS One 6:e15921
13. Wenzel EV, Bosnak M, Tierney R et al (2020) Human antibodies neutralizing diphtheria toxin in vitro and in vivo. Sci Rep 10:571
14. Li Z, Krippendorff B-F, Sharma S et al (2016) Influence of molecular size on tissue distribution of antibody fragments. mAbs 8:113–119
15. Khan KH (2013) Gene expression in mammalian cells and its applications. Adv Pharm Bull 3:257–263
16. Jäger V, Büssow K, Wagner A et al (2013) High level transient production of recombinant antibodies and antibody fusion proteins in HEK293 cells. BMC Biotechnol 13:52
17. Li MZ, Elledge SJ (2012) SLIC: a method for sequence- and ligation-independent cloning. Methods Mol Biol 852:51–59
18. Fühner V, Heine PA, Helmsing S et al (2018) Development of neutralizing and non-neutralizing antibodies targeting known and novel epitopes of TCDB of Clostridioides difficile. Front Microbiol 9:2908
19. Kumaraswamy S, Tobias R (2015) Label-Free kinetic analysis of an antibody-antigen interaction using biolayer interferometry. Methods Mol Biol 1278:165–182
20. Bleckmann M, Schürig M, Chen FF et al (2016) Identification of essential genetic baculoviral elements for recombinant protein expression by transactivation in Sf21 insect cells. PloS One 11:e0149424
21. Shen X, Pitol AK, Bachmann V et al (2015) A simple plasmid-based transient gene expression method using High Five cells. J Biotechnol 216:67–75
22. Stolt-Bergner P, Benda C, Bergbrede T et al (2018) Baculovirus-driven protein expression in insect cells: a benchmarking study. J Struct Biol 203:71–80

Chapter 5

The Intervening Removable Affinity Tag (iRAT) System for the Production of Recombinant Antibody Fragments

Norimichi Nomura, Yayoi Nomura, Yumi Sato, and So Iwata

Abstract

Fv and Fab antibody fragments are versatile co-crystallization partners that aid in the structural determination of otherwise "uncrystallizable" proteins, including human/mammalian membrane proteins. Accessible methods for the rapid and reliable production of recombinant antibody fragments have been long sought. In this chapter, we describe the concept and protocols of the intervening removable affinity tag (iRAT) system for the efficient production of Fv and Fab fragments in milligram quantities, which are sufficient for structural studies. As an extension of the iRAT system, we also provide a new method for the creation of genetically encoded fluorescent Fab fragments, which are potentially useful as molecular devices in various basic biomedical and clinical procedures, such as immunofluorescence cytometry, bioimaging, and immunodiagnosis.

Key words Recombinant antibody, Polyprotein, Secretory expression, Fv fragment, Fab fragment, Crystallography, Genetically encoded fluorescent antibody, Therapeutic antibody, Biosimilar, Immunodiagnostics

1 Introduction

Crystallography of biomedically relevant proteins often remains stagnant because of challenges in preparing diffraction-quality crystals. In most cases, their molecular flexibility, conformational heterogeneity, or polydispersal characteristics in solution hinder protein crystallization. Antibody-aided crystallography is a powerful strategy to manage this problem and can be applied to various membrane proteins as well as heavily glycosylated and/or multidomain soluble proteins [1–7]. Fv and Fab antibody fragments assist crystallization by increasing the hydrophilic surface area available for rigid crystal lattice formation. The bound antibody fragments reduce the inherent protein flexibility and conformational heterogeneity, thereby increasing the chance of successful crystallization of difficult target proteins. In addition to crystallography, Fab and Fv fragments have also been proven to be powerful as

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_5,

fiducial markers for single-particle cryo-electron microscopy (cryo-EM) to aid in structure determination [8, 9].

The preparation of milligram quantities of purified Fv and Fab fragments is the most critical prerequisite once the strategies for antibody-aided structural studies are adopted. Recently, we developed an efficient production strategy for recombinant Fv antibody fragments and named it the "iRAT system" [10]. A primary feature of the iRAT system is the secretory expression of a single synthetic polyprotein consisting of a variable light (V_L) domain, an intervening removable affinity tag (iRAT), and a variable heavy (V_H) domain using the Gram-positive bacterium *Brevibacillus choshinensis*. This ensures 1:1 stoichiometric expression of the V_L and V_H domains from the monocistronic construct, followed by proper folding and assembly of the two domains, each of which contains an intrachain disulfide bond. A polyhistidine-maltose binding protein (His_6-MBP) cassette, embedded within the intervening region of the two variable domains, serves as an affinity purification tag that can be removed by tobacco etch virus (TEV) protease cleavage during the purification process to yield tag-free Fv fragments suitable for crystallization trials (Fig. 1a). We have validated the iRAT-based production of multiple Fv fragments, including a crystallization chaperone for a mammalian membrane protein as well as

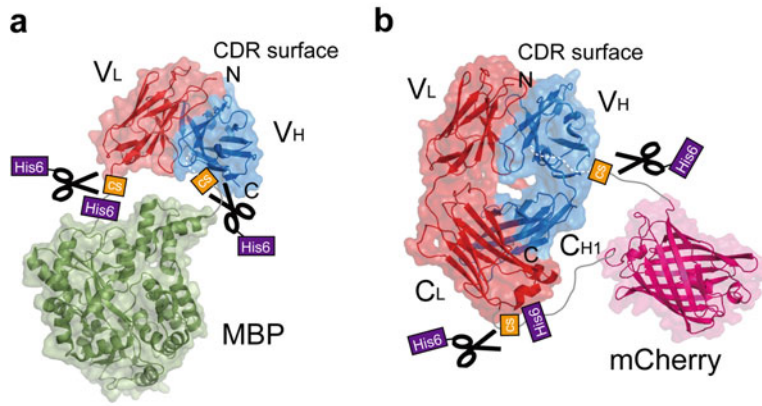


Fig. 1 Concept of the iRAT production system. (**a**) iRAT-synthetic polyproteins with the built-in MBP facilitate production of Fvs. (**b**) iRAT-synthetic polyproteins with the built-in mCherry and long Gly/Ser-rich tethering linkers facilitate production of Fabs or Fvs. The resulting polyproteins are recovered from the culture supernatant of *Brevibacillus* or *Sf*9, purified by IMAC and excised with TEV-His_6 (indicated by the "scissors" icon). After the removal of the iRAT portion and TEV-His_6 by reverse IMAC, non-tagged antibody fragments are further purified by using SEC. The antibody light chain-derived domains are red, iRAT-MBP is green, iRAT-mCherry is magenta, and the antibody heavy chain-derived domains are blue. The polyprotein N-terminus, polyprotein C-terminus, and CDR surface are marked. The TEV cleavage site is indicated by the "cs" in orange boxes. Notably, iRAT-mCherry polyproteins can be utilized as genetically encoded fluorescent antibody fragments if not treated with TEV

FDA-approved therapeutic antibodies. The resulting Fv fragments are functionally active and are crystallized in complex with the target proteins [4, 10–12].

Another important feature of the iRAT system is its high extensibility. Antibodies in different modalities can be produced via the original and modified iRAT system. Beyond the production of Fv fragments, we have succeeded in producing Fab fragments through a process of reorganization of the iRAT segment. Interestingly, when a polyhistidine-mCherry (His_6-mCherry) cassette is incorporated within the long intervening region, tethering the light and heavy chains of Fab fragments, the resultant polyproteins fold and assemble into functional fluorescent Fab fragments (Fig. 1b). The iRAT concept could provide promising clues to design and create antibody-derived artificial molecular devices that are useful in various basic biomedical and clinical procedures, such as fluorescence-activated cell sorting (FACS), fluorescence-linked immunosorbent assay (FLISA), immunohistochemistry (IHC), fluorescent bioimaging, and immunodiagnostics.

In this chapter, we present reliable and broadly applicable iRAT protocols to prepare milligram quantities of Fv and Fab antibody fragments using *Brevibacillus* and *Sf*9 insect cells, respectively, with an average duration. Anticipated results are also shown, taking examples of the iRAT-mediated production of certolizumab Fv and Fab. Certolizumab (Cimzia®) is an FDA-approved humanized monoclonal antibody that targets tumor necrosis factor α (TNFα) and is widely used to treat rheumatoid arthritis and Crohn's disease. Biosimilars of such known therapeutic antibodies for research use can be easily prepared via the iRAT system. Both Fv production using *Sf*9 cells and Fab production using *Brevibacillus* can be performed similarly, although the detailed protocols are not described here due to space limitations.

2 Materials

2.1 Molecular Biology

1. PrimeSTAR Max Premix (2×) (Takara Bio/Clontech).
2. Gibson Assembly Master Mix (2×) (New England Biolabs).
3. TaKaRa Ex Taq (5 U/μL) (Takara Bio/Clontech).
4. 10× Ex Taq Buffer.
5. dNTPs mixture (2.5 mM each).
6. Wizard SV gel and PCR clean-up system (Promega).
7. MinElute PCR purification kit (Qiagen).
8. QIAquick spin miniprep kit (Qiagen).
9. Thermal cycler.
10. NanoDrop spectrophotometer.

2.2 Protein Expression Using Brevibacillus Cells

1. *Brevibacillus choshinensis* HPD31-SP3 (Takara Bio/Clontech).
2. TM medium: 10 g/L glucose, 10 g/L polypeptone, 5 g/L meat extract, 2 g/L yeast extract, 10 mg/L $FeSO_4 \cdot 7H_2O$, 10 mg/L $MnSO_4 \cdot 4H_2O$, 1 mg/L $ZnSO_4 \cdot 7H_2O$, pH 7.0.
3. MT medium: 10 g/L glucose, 10 g/L polypeptone, 5 g/L meat extract, 2 g/L yeast extract, 10 mg/L $FeSO_4 \cdot 7H_2O$, 10 mg/L $MnSO_4 \cdot 4H_2O$, 1 mg/L $ZnSO_4 \cdot 7H_2O$, 4.1 g/L $MgCl_2$, pH 7.0.
4. MT/neomycin agar plate: 10 g/L glucose, 10 g/L polypeptone, 5 g/L meat extract, 2 g/L yeast extract, 10 mg/L $FeSO_4 \cdot 7H_2O$, 10 mg/L $MnSO_4 \cdot 4H_2O$, 1 mg/L $ZnSO_4 \cdot 7H_2O$, 4.1 g/L $MgCl_2$, 10 mg/L neomycin, 15 g/L agar, pH 7.0.
5. 2SY/neomycin medium: 20 g/L glucose, 40 g/L Bacto Soytone, 5 g/L yeast extract, 0.15 g/L $CaCl_2 \cdot 2H_2O$, 50 mg/L neomycin.
6. PGH buffer: 1 mM HEPES-KOH (pH 7.0), 15% (w/v) glycerol, 15% (w/v) PEG6000.
7. pBIM2 vector (Fig. 2).
8. Synthetic or natural antibody-encoding cDNA of interest (*see* **Note 1**).
9. BIM2-F primer: 5′-ACTAGTGAAAATTTATATTTTCAAGG TCACCATCACCATCACCATTCCTCTGGTGGCGGT.
10. BIM2-R primer: 5′-AAGCTTCGATTGGAAGTACAGGTTT TCGGAAGATCTAGAGGAACCACCCCCACCCCCGAG.
11. pNY326-F primer: 5′-GAATTCGGTACCCCGGGTTCGA.
12. pNY326-R primer: 5′-GGATCCTGCAGCGAAAGCCATGGGAGC.
13. BacillusExp-F primer: 5′-CACGCGCTTGCAGGATTCGG.
14. BacillusExp-R primer: 5′- CAATGTAATTGTTCCCTACCT GC.
15. GenePulser electroporator (Bio-Rad).
16. 0.2-cm electroporation cuvette.
17. 125-mL baffled bottom Erlenmeyer flask.
18. 500-mL baffled bottom Erlenmeyer flask.
19. 2.5-L Tunair baffled shaker flask.
20. Orbital shaker incubator.
21. Centrifuge with adaptors for 1-L, 250-mL, 50-mL, and 15-mL tubes.

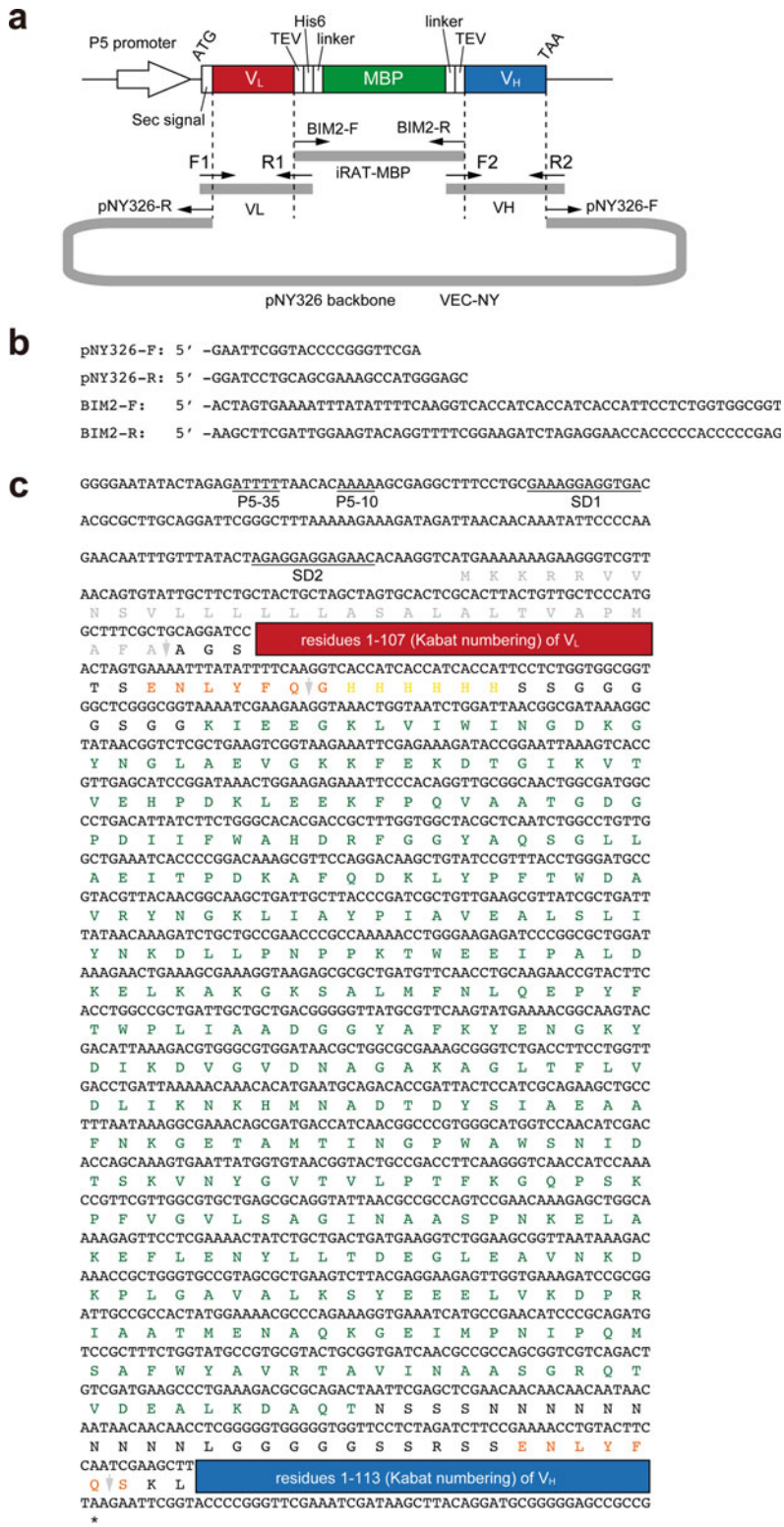


Fig. 2 Structure of the iRAT-based Fv expression vector pBIM2. (**a**) Schematic illustration of the region containing the promoter, secretion signal, and iRAT-

2.3 Protein Expression Using Sf9 Cells and Baculoviruses

1. *E. coli* TOP10 chemically competent cells (Thermo Fisher Scientific).
2. *E. coli* DH10Bac chemically competent cells (Thermo Fisher Scientific).
3. SOC medium: 20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.
4. LB/ampicillin plate: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 100 mg/L ampicillin, 15 g/L agar.
5. LB/ampicillin medium: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 100 mg/L ampicillin.
6. LB/KTGIB agar plate: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 50 mg/L kanamycin, 10 mg/L tetracycline, 7 mg/L gentamicin, 40 mg/L IPTG, 50 mg/L X-Gal, 15 g/L agar.
7. LB/KTG medium: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 50 mg/L kanamycin, 10 mg/L tetracycline, 7 mg/L gentamicin.
8. pFBGP67-iRATmC (Fig. 4).
9. IRAT70-F primer: 5′-ACTAGTGAAAATTTATATTTTCAAGGTCACCATCACCATCACCATGGTGGCTCCTCTGGT.
10. IRAT70-R primer: 5′- AAGCTTCGATTGGAAGTACAGGTTTTCACCACCACTACTACTGGAACCCGAAGAGCCGGA.
11. FB-F1 primer: 5′- TAAAAGCTTGTCGAGAAGTACTAGAGGATCATAATCAGCCATACCACA.
12. gp67-R primer: 5′- CGCCGCAAAGGCAGAATGCGCCGCCGCCGCCAAAAGCACATATAAAAC.
13. SFseq-F primer: 5′-ACTGTTTTCGTAACAGTTTTGTAATA.
14. SFseq-R primer: 5′- GTGGTATGGCTGATTATGATCCTCTAGTACTTCT.
15. *Sf*9 cells adapted for suspension growth.
16. PSFM-J1 medium (Wako).
17. Fetal bovine serum (FBS).

Fig. 2 (continued) polyprotein construct. (**b**) Universal PCR primers for the construction of expression plasmid. (**c**) Sequence corresponding to the region shown in (**a**). P5–35 and P5–10 represent the −35 and −10 box of the promoter P5, respectively. SD1 and SD2 are Shine-Dalgarno sequences. The Sec secretion signal, TEV recognition sequence, His_6 tag, and MBP are shown in gray, orange, yellow, and green, respectively. Gray arrowheads indicate the signal cleavage site and TEV cleavage sites

18. PSFM-J1/FBS-PS medium: PSFM-J1 medium, 2% FBS, 50,000 U/L penicillin G, 50 mg/L streptomycin.
19. 1× Grace's medium (Thermo Fisher Scientific).
20. FuGENE HD transfection reagent (Promega).
21. LUNA automated cell counter and cell counting slides (Thermo Fisher Scientific).
22. Trypan blue: 0.4% solution in PBS, pH 7.2.
23. Six-well tissue plate with flat bottom.
24. 125-mL Erlenmeyer flask (Vent filter cap/Baffle bottom) (Corning).
25. 3-L Fernbach flask (Vent filter cap/Baffle bottom) (Corning).
26. Orbital shaker incubator.
27. Centrifuge with adaptors for 1-L and 50-mL tubes.
28. Temperature-controlled room or incubator set at 27 °C.

2.4 Protein Purification and Characterization

1. Ammonium sulfate.
2. TBS: 10 mM Tris-HCl (pH 7.5), 150 mM NaCl.
3. Buffer A: 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole.
4. Buffer B: 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 250 mM imidazole.
5. Buffer C: 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM imidazole.
6. Buffer D: 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 500 mM imidazole.
7. Ni-NTA Superflow resin (Qiagen).
8. Ni Sepharose excel resin (GE Healthcare).
9. Glass Econo-column (Bio-Rad).
10. HisTrap HP column (5 mL) (GE Healthcare).
11. HiLoad16/60 Superdex200 column (GE Healthcare).
12. Dialysis tubing (MWCO 10 K) (Spectra/Por).
13. Amicon Ultra-15 (MWCO 10 K) centrifugal concentrator (Merck).
14. Amicon Ultra-15 (MWCO 30 K) centrifugal concentrator (Merck).
15. His_6-tagged TEV protease (TEV-His_6) (*see* **Note 2**).
16. Chromatograhy system (e.g., BioLogic DuoFlow system (Bio-Rad)).

3 Methods

3.1 Original iRAT System for the Production of Recombinant Fv Fragments

Fv fragments (with a molecular weight of ~25 kDa) are heterodimeric molecules consisting of two domains, V_L and V_H, that assemble through hydrophobic interactions. Each of the V_L and V_H domains contains an intrachain disulfide bond for stabilization. These structural properties require a sophisticated apparatus for secretion, folding, and assembly, as well as an oxidizing environment for disulfide bond generation. The expression of Fv fragments via the *B. choshinensis* Sec-dependent secretion pathway ensures the formation of the intrachain disulfide bonds to stabilize the structure. A reason to include MBP in the iRAT segment is that the distance between the N- and C-termini (43 Å) is similar to the distance between the C-terminus of V_L and the N-terminus of V_H (approximately 35–40 Å) for most Fv fragments.

3.1.1 Preparation of Brevibacillus Electrocompetent Cells

1. *Day 1*: Inoculate a single colony of *B. choshinensis* into 10 mL of TM medium and grow overnight at 37 °C and 200 rpm.
2. *Day 2*: Inoculate 100 mL of TM medium in a baffled bottom 500-mL Erlenmeyer flask with 1 mL of overnight culture and incubate at 37 °C and 200 rpm until an OD_{600nm} of 3.0–4.0 (*see* **Note 3**).
3. Centrifuge at 5000 × *g* for 10 min at 4 °C.
4. Resuspend the cell pellet in 20 mL of PGH buffer.
5. Centrifuge at 8000 × *g* for 20 min at 4 °C.
6. Remove the supernatant thoroughly and resuspend the cell pellet carefully in 2 mL of PGH buffer.
7. Dispense into 100 μL aliquots and flash-freeze with liquid nitrogen.
8. Store at −80 °C.

3.1.2 Construction of the Brevibacillus Expression Plasmid

The pBIM2 vector (Fig. 2a) for the expression of Fv fragments is designed based on the backbone of the *Brevibacillus* plasmid pNY326. The V_L-insertion site, iRAT cassette, and V_H-insertion site are tandemly arranged in this order and are introduced downstream of the constitutively active P5 promoter and Sec secretion signal sequence. The iRAT-MBP cassette encodes the N-terminal TEV cleavage site, His_6 tag, a 9-residue Gly/Ser-rich linker, the MBP coding sequence (residues 27–392 from UniProt ID: P0AEX9), a 25-residue Asn/Gly/Ser-rich linker, and a C-terminal TEV cleavage site (Fig. 2c). Residues 1–107 (Kabat numbering scheme) of the V_L domain and 1–113 of the V_H domain derived from monoclonal antibodies are used for Fv fragment expression.

Table 1
Preparation of PCR tubes

Component	Volume	Final concentration
H_2O	23.4 μL	–
PrimeSTAR Max Premix (2×)	25 μL	1×
Forward primer (100 pmol/μL)	0.3 μL	0.6 μM
Reverse primer (100 pmol/μL)	0.3 μL	0.6 μM
Template (5 ng/μL)	1 μL	0.1 ng DNA/μL
Total	50 μL	

Table 2
PCR primer template combination

PCR product				
Name	**Size**	**Forward primer**	**Reverse primer**	**Template**
V_L	360 bp	F1	R1	Native or synthetic cDNA encoding V_L
iRAT-MBP	1300 bp	BIM2-F	BIM2-R	pBIM2 vector
V_H	380 bp	F2	R2	Native or synthetic cDNA encoding V_H
VEC-NY	3400 bp	pNY326-F	pNY326-R	pBIM2 vector

1. *Day 1*: Perform PCR amplifications of V_L, the iRAT-MBP segment, V_H, and the pNY326 vector backbone (VEC-NY). Add the appropriate components to four 0.2 mL PCR tubes as indicated in Table 1. The PCR primer–template combinations are listed in Table 2. The primers for the V_L and V_H domains include 5′ overhangs (27 bp) complementing the upstream and downstream sequences to the pBIM2 vector junctions. Mix the contents gently, and place the tubes in a thermal cycler.
2. Run the PCR according to the following parameters: (1) 2 min at 95 °C; (2) 30 cycles of a sequence composed of 10 s at 98 °C/5 s at 56 °C/5 s/kbp at 72 °C; and (3) 10 min at 72 °C.
3. Check the PCR products by loading an aliquot (5 μL) onto a 1% agarose gel with a DNA ladder in a separate lane. After ethidium bromide staining, observe the gel by UV illumination.
4. Purify the remaining PCR product (45 μL) using a Wizard SV gel and PCR clean-up system, according to the manufacturer's instructions. Use the PCR products directly for purification. DNA extraction from the agarose gel slice is not necessary.

Table 3
Gibson assembly reaction

Component	Size of DNA fragment	Amount per reaction
V_L	360 bp	0.5 pmol[a]
iRAT-MBP	1300 bp	0.5 pmol[a]
V_H	380 bp	0.5 pmol[a]
VEC-NY	3400 bp	0.16 pmol[a]
Gibson Assembly Master Mix (2×)	---	10 μL
Sterile MilliQ water	---	To make 20 μL total volume
Total		20 μL

[a]pmol = (weight in ng) × 1000/(base pairs × 650 Da)

5. Measure the DNA concentration using a NanoDrop spectrophotometer.
6. Set up the Gibson assembly reaction (20 μL) as in Table 3.
7. Incubate for 60 min at 50 °C.
8. Add 80 μL of sterile MilliQ water to make a total volume of 100 μL.
9. Purify the reaction product using a MinElute PCR purification kit, according to the manufacturer's instructions. Carefully elute the DNA with 10 μL of sterile MilliQ water.
10. Thaw 100 μL of *Brevibacillus choshinensis* electrocompetent cells on ice (approximately 10 min), and mix the cells by flicking gently.
11. Transfer the cells to a chilled 1.5-mL centrifuge tube. Add 10 μL of the purified Gibson assembly reaction product.
12. Carefully transfer the cell/DNA mix into a chilled 0.2-cm cuvette without introducing bubbles, and ensure that the cells deposit across the bottom of the cuvette. Electroporate using the following conditions for the Bio-Rad GenePulser electroporator: 1.5 kV, 1000 Ω, and 25 μF. The typical time constant is ~19 ms.
13. Immediately add 1 mL of pre-warmed MT medium to the cuvette, gently mix by pipetting up and down twice, and then transfer to the 17 mm × 100 mm round-bottomed culture tube.
14. Shake (200 rpm) at 37 °C for 1 h.
15. Concentrate the cells via centrifugation as appropriate, and then spread 100 μL of cells onto a pre-warmed MT/neomycin agar plate.

Table 4
Colony PCR mix

Component	Volume/reaction	Final concentration
H_2O	3.4 μL	–
10× Ex Taq Buffer	1 μL	1×
dNTPs mixture (2.5 mM each)	1 μL	250 μM each
BacillusExp-F (1 pmol/μL)	2 μL	0.2 μM
BacillusExp-R (1 pmol/μL)	2 μL	0.2 μM
TaKaRa Ex Taq (5 U/μL)	0.1 μL	0.05 U/μL
Brevibacillus cell lysate	0.5 μL	–
Total	10 μL	

16. Incubate the plate overnight at 37 °C.
17. *Day 2*: Perform colony PCR to check the size of the insert DNA. Aliquot 100 μL of sterile MilliQ water into 0.2 mL PCR tubes (*see* **Note 4**).
18. To create a stock of each individual colony, pick a small colony using a sterile pipette tip, streak the pipette tip onto another MT/neomycin plate, and grow at 37 °C for more than 6 h.
19. Dip the same pipette tip into MilliQ water in a 0.2-mL PCR tube.
20. Boil the *Brevibacillus* cells on a thermal cycler by incubating the tubes at 95 °C for 5 min.
21. Centrifuge the boiled samples at 6000 × *g* for 1 min at room temperature.
22. Use 0.5 μL of the supernatant to prepare the colony PCR mix (10 μL) as in Table 4.
23. Run the PCR according to the following parameters: (1) 5 min at 95 °C; (2) 30 cycles of a sequence composed of 30 s at 94 °C/1 min at 55 °C/1 min at 72 °C; and (3) 5 min at 72 °C.
24. Load 10 μL of each colony PCR product directly onto a 1% agarose gel, alongside an appropriate DNA ladder. After ethidium bromide staining, observe the gel by UV illumination. The expected size of the colony PCR product is approximately 2200 bp.
25. Select several positive colonies and use them to inoculate 6 mL of 2SY/neomycin medium. Shake the cultures overnight at 30 °C and 200 rpm.
26. *Day 3*: Add 1 mL of each saturated overnight culture to 0.5 mL of 50% glycerol in a 2-mL cryovial and gently mix. Freeze the

glycerol stock tube at −80 °C. The stock is now stable for years, as long as it is kept at −80 °C.

27. Extract and purify the plasmid DNA from the remaining 5 mL of each culture using the QIAquick spin miniprep kit, according to the manufacturer's instructions. Measure the DNA concentration using a NanoDrop spectrophotometer.
28. Using the vector primers, BacillusExp-F and BacillusExp-R, perform sequence analysis to confirm that the nucleotide sequence of the plasmids is correct.

3.1.3 Secretory Expression Using Brevibacillus Cells

1. *Day 1*: (Pre-culture) Open the tube of the glycerol stock of *Brevibacillus choshinensis* harboring a positive pBIM2 expression plasmid and scrape some of the frozen bacteria off using a sterile loop. Inoculate the bacteria into 50 mL of 2SY/neomycin medium in a 125-mL baffled-bottom Erlenmeyer flask and grow for 40–55 h at 30 °C and 200 rpm.
2. *Day 3*: (Large-scale culture) Inoculate 1 L of 2SY/neomycin medium into a 2.5-L Tunair baffled shaker flask with 10 mL of saturated pre-culture.
3. Grow the cells for 48–60 h at 30 °C and 200 rpm.
4. *Day 5*: Centrifuge the culture at 6000 × *g* for 15 min at 4 °C. Recover the supernatant and discard the cell pellet.

3.1.4 Purification of Tag-Free Fv Fragments

1. *Day 1*: (Ammonium sulfate precipitation) To the 1 L of *Brevibacillus* culture supernatant, add 390 g of ammonium sulfate powder slowly but steadily with thorough mixing to adjust to a final concentration of 60% saturation. Do not allow clumps to form.
2. Allow a precipitate to form for 1 h at 4 °C with stirring.
3. Recover the precipitate by centrifugation at 10,000 × *g* for 20 min at 4 °C.
4. Dissolve the precipitate by adding 10 mL of TBS.
5. Dialyze overnight against 1 L of TBS at 4 °C.
6. Centrifuge the dialyzed sample at 15,000 × *g* for 10 min at 4 °C. Recover the supernatant and discard the precipitate.
7. *Day 2*: (Immobilized metal affinity chromatography (IMAC)) Use 5 mL of Ni-NTA resin (10 mL of 50% slurry) per 1 L of *Brevibacillus* culture, pour the slurry into a glass Econo-Column, and equilibrate with five column volumes (CVs) of buffer A.
8. Add the supernatant to the column and collect the flow-through fraction.
9. Wash the column with 10 CVs (50 mL) of buffer A.

10. Elute the polyprotein of interest from the column by adding three CVs (15 mL) of buffer B and collect the eluate into a 50-mL Falcon tube.
11. Measure the protein concentration of each fraction using a NanoDrop spectrophotometer.
12. (TEV cleavage and dialysis) Add an appropriate amount (typically at a protease to target protein ratio of 1:100 (w/w) or 3 mg) of TEV-His_6 to the elution fraction.
13. Transfer the mixture into dialysis tubing, and dialyze overnight at 4 °C against 2 L of TBS.
14. *Day 3*: (Reverse IMAC) Equilibrate a 5 mL HisTrap HP column with 25 mL of buffer C.
15. Remove the dialysate from the tubing and inject it into the column using a 10 mL syringe.
16. Collect the flow-through fraction (FT).
17. Wash the column three times, each wash containing 1 CV (5 mL) of buffer C and collect as W1, W2, and W3 fractions, respectively.
18. Elute the bound material (excised iRAT-MBP segment, TEV--His_6, and contaminating proteins) from the column by adding three CVs (15 mL) of buffer D and collect as E fraction.
19. Measure the protein concentration of each fraction using a NanoDrop spectrophotometer.
20. Mix the FT and W1 fractions and concentrate the protein using a Amicon Ultra-15 (MWCO 10 K) centrifugal concentrator to a final volume of 5 mL.
21. (Size-exclusion chromatography (SEC)) Inject the 5 mL of concentrated protein sample onto a HiLoad16/60 Superdex200 column pre-equilibrated with TBS at a flow rate of 1 mL/min.
22. Collect 2 mL fractions. Analyze the UV trace, and pool the peak fractions.
23. Perform SDS-PAGE to check the purity.
24. Measure the protein concentration of each fraction using a NanoDrop spectrophotometer to estimate the protein yield.
25. Use an aliquot to evaluate the biophysical and/or functional proprieties of the purified tag-free Fv fragment (Fig. 3 as a typical example). Ideally, binding assays, such as surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), and ELISA, should provide valuable information.
26. Concentrate the remaining purified Fv fragment using an Amicon Ultra-15 (MWCO 10 K) centrifugal concentrator to an appropriate protein concentration (e.g., 10 mg/mL).

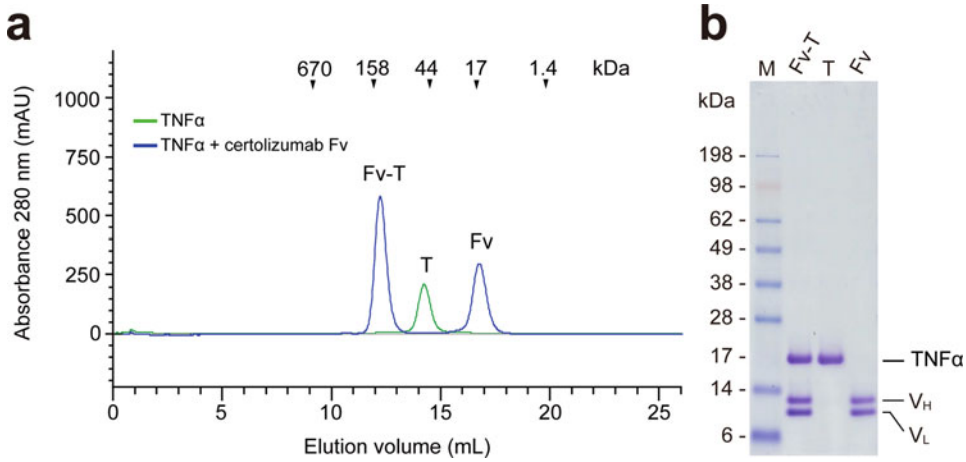


Fig. 3 Certolizumab Fv produced via the *Brevibacillus*-iRAT system. (**a**) Typical profile in size exclusion chromatography of the certolizumab Fv-TNFα complex in the presence of excess certolizumab Fv on a Superdex200 Increase 10/300GL column (blue line). As a reference, TNFα alone was separated on the same column (green line). Elution volumes of protein standards are indicated at the top. *Peak Fv-T* certolizumab Fv-TNFα trimer complex, *peak T* free TNFα trimer, and *peak Fv* free certolizumab Fv. (**b**) SDS-PAGE analysis of the corresponding peak fractions in (**a**). The theoretical molecular masses of the certolizumab Fv and TNFα trimer are 24.6 kDa (V_L: 11.6 kDa, V_H: 13.0 kDa) and 51.9 kDa (17.3 kDa protomer ×3), respectively

27. Re-measure the protein concentration of each fraction using a NanoDrop spectrophotometer.
28. Prepare 100 μL aliquots and flash-freeze with liquid nitrogen.
29. Store at −80 °C.

3.2 Extended iRAT System for the Production of Recombinant Fab Fragments and Their Fluorescent Derivatives

Fab fragments (with a molecular weight of ~50 kDa) are heterodimeric molecules consisting of two chains, each of which has two domains: the light chain variable domain (V_L) linked to the light chain constant domain (C_L) and the heavy chain variable domain (V_H) linked to the heavy chain constant domain 1 (C_H1). Fab fragments contain five disulfide bonds (except IgA and IgE which have six disulfides) and need to be targeted to the Gram-negative bacterial periplasm, the Gram-positive bacterial extracellular space, or the eukaryotic endoplasmic reticulum. The linear distance between the C-terminus of the Fab light chain (Fab_L) and N-terminus of the Fab heavy chain (Fab_H) for most Fab fragments is approximately 65 Å. Appropriate long flexible linkers are needed for successful "molecular surgery" to anastomose the two termini without violating the overall fold and biological function of the Fab fragments. According to Koerber et al. a 70-residue Gly/Ser-rich linker (sc70) is available to express functional single-chain Fab [13]. Here we provide an example of which the mCherry red fluorescent protein, instead of MBP, is installed into an intermediate point of the sc70 linker to create a new iRAT segment. This construct yields a polyprotein that works as a genetically encoded

fluorescent Fab with complete target-binding activity before removal of the iRAT segment as described in the Subheading 3.2.3 (*see* Fig. 5). Once the iRAT segment is excised by TEV protease, a large amount of tag-free Fab fragment can be prepared as described in the Subheading 3.2.4 (*see* Fig. 6). Notably, this newly innovated iRAT segment has a moonlighting function as a fluorescent tag (*see* **Note 5**). The nested mCherry facilitates estimation of the expression levels of antibody fragments by simply applying the crude culture supernatant to fluorescence size-exclusion chromatography (FSEC).

iRAT-mediated Fab production can be performed using either *Brevibacillus* or *Sf*9 insect cells (*see* **Note 6**). When *Sf*9 expression is adopted, the culture supernatant can be directly loaded onto a column for IMAC using Ni Sepharose excel resin, which saves time and labor for pretreatment, such as ammonium sulfate precipitation and dialysis.

3.2.1 Construction of the Baculovirus Strain

The pFBgp67-iRATmC vector (Fig. 4a) for the expression of Fab fragments is designed based on the backbone of the pFastBac1, an entry vector to produce Bac-to-Bac baculoviruses. The Fab_L insertion site, iRATmC cassette, and Fab_H insertion site are tandemly arranged in this order and are introduced downstream of the constitutively active P_{PH} promoter and gp67 secretion signal sequence. The iRATmC cassette encodes the N-terminal TEV cleavage site, His_6 tag, a 35-residue Gly/Ser/Thr-rich linker, the mCherry coding sequence (residues 3–236 from UniProt ID: X5DSL3), a 35-residue Gly/Ser-rich linker, and a C-terminal TEV cleavage site (Fig. 4c).

1. *Day 1*: Perform PCR amplifications of the Fab_L, iRATmC, Fab_H, and pFastBac vector backbone containing the gp67 signal sequence (VEC-FB). Add the appropriate components to four 0.2-mL PCR tubes as in Table 5. The PCR primer-template combinations are listed in Table 6. The primers for Fab_L and Fab_H domains include 5′ overhangs (27 bp) complementing the upstream and downstream sequences to the pFBgp67-iRATmC vector junctions. Mix the contents gently, and place the tubes in a thermal cycler.
2. Run the PCR according to the following parameters: (1) 2 min at 95 °C; (2) 30 cycles of a sequence composed of 10 s at 98 °C/5 s at 56 °C/5 s/kbp at 72 °C; and (3) 10 min at 72 °C.
3. Check the PCR products by loading an aliquot (5 μL) on a 1% agarose gel with a DNA ladder in a separate lane. After ethidium bromide staining, observe the gel by UV illumination.
4. Purify the PCR product (45 μL) using a Wizard SV gel and PCR clean-up system, according to the manufacturer's instructions. Use the PCR products directly for purification. DNA extraction from the agarose gel slice is not necessary.

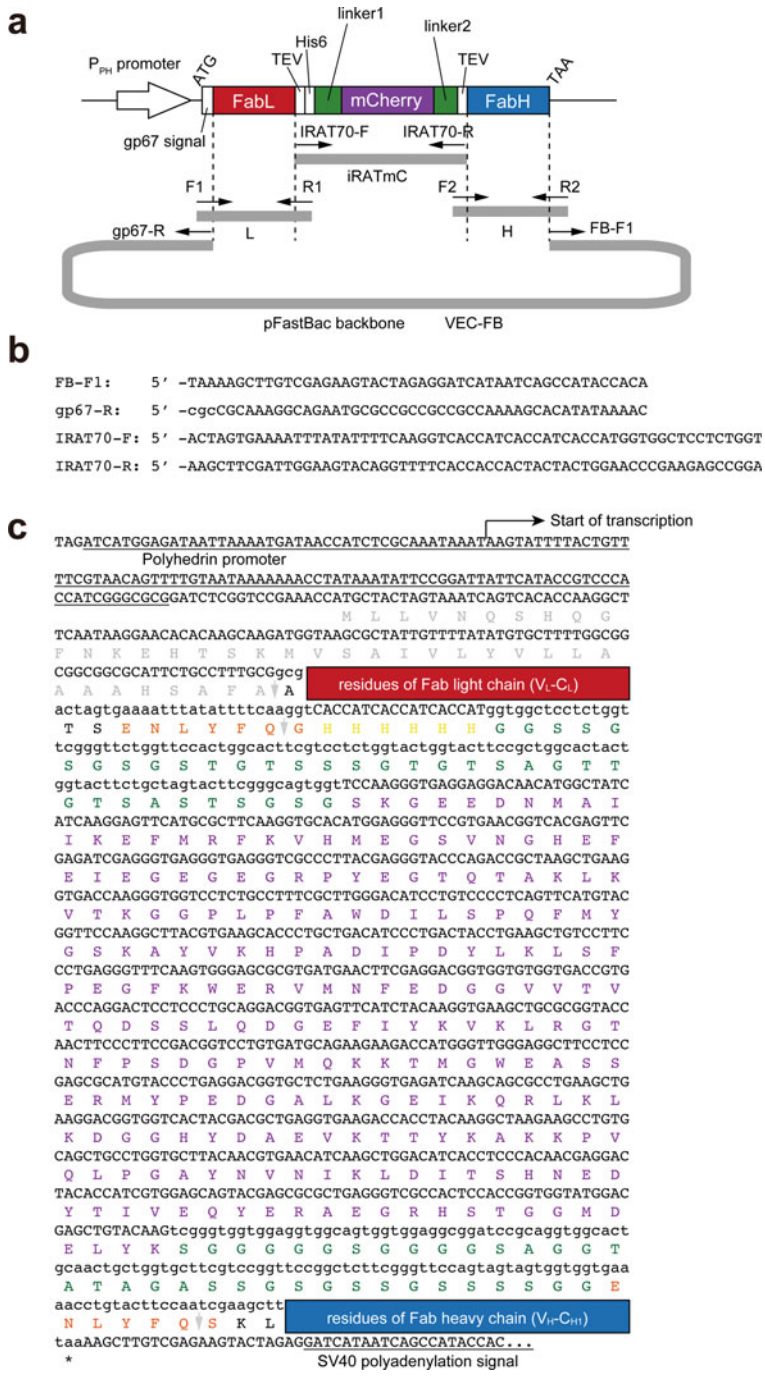


Fig. 4 Structure of the Bac-to-Bac entry vector pFBgp67-iRATmC for the iRAT-based Fab expression. (**a**) Schematic illustration of the region containing the promoter, secretion signal, and iRAT-polyprotein construct. (**b**) Universal PCR primers for construction of expression plasmid. (**c**) Sequence corresponding to the region shown in (**a**). The gp67 secretion signal, TEV recognition sequence, His_6 tag, and mCherry are shown in gray, orange, yellow, and purple, respectively. Gray arrowheads indicate the signal cleavage site and TEV cleavage sites. The sc70-derived linker regions are shown in green

Table 5
PCR amplification

Component	Volume	Final concentration
H_2O	23.4 μL	–
PrimeSTAR Max Premix (2×)	25 μL	1×
Forward primer (100 pmol/μL)	0.3 μL	0.6 μM
Reverse primer (100 pmol/μL)	0.3 μL	0.6 μM
Template (5 ng/μL)	1 μL	0.1 ng DNA/μL
Total	50 μL	

Table 6
PCR primer–template combinations

PCR product		Forward primer	Reverse primer	Template
Name	Size			
Fab_L	700 bp	F1	R1	Native or synthetic cDNA encoding Fab light chain
iRATmC	980 bp	IRAT70-F	IRAT70-R	pFBgp67-iRATmC
Fab_H	700 bp	F2	R2	Native or synthetic cDNA encoding Fab heavy chain
VEC-FB	4800 bp	FB-F1	gp67-R	pFBgp67-iRATmC

5. Measure the DNA concentration using a NanoDrop spectrophotometer.
6. Set up the Gibson assembly reaction (20 μL) as in Table 7.
7. Incubate for 60 min at 50 °C.
8. Add 60 μL of sterile MilliQ water to the reaction product.
9. Aliquot 10 μL of the diluted reaction product, and chill in a 1.5-mL centrifuge tube.
10. Thaw the *E. coli* TOP10 competent cells (100 μL) on ice.
11. Add 100 μL of competent cells to the DNA. Mix gently by pipetting up and down to mix the cells and DNA. Do not vortex.
12. Place the mixture on ice for 30 min.
13. Heat shock at 42 °C for 30 s.
14. Place the mixture on ice for 2 min.
15. Add 900 μL of pre-warmed SOC medium to the tube.

Table 7
Gibson assembly reaction

Component	Size of DNA fragment	Amount per reaction
Fab_L	700 bp	0.5 pmol[a]
iRATmC	980 bp	0.5 pmol[a]
Fab_H	700 bp	0.5 pmol[a]
VEC-FB	4800 bp	0.16 pmol[a]
Gibson Assembly Master Mix (2×)	–	10 μL
Sterile MilliQ water	–	To make a final volume of 20 μL
Total		20 μL

[a]pmol = (weight in ng) × 1000/(base pairs × 650 Da)

Table 8
Colony PCR mix

Component	Volume/reaction	Final concentration
H_2O	3.9 μL	—
10× Ex Taq Buffer	1 μL	1×
dNTPs mixture (2.5 mM each)	1 μL	250 μM each
SFseq-F (1 pmol/μL)	2 μL	0.2 μM
SFseq-R (1 pmol/μL)	2 μL	0.2 μM
TaKaRa Ex Taq (5 U/μL)	0.1 μL	0.05 U/μL
Total	10 μL	

16. Incubate the tube at 37 °C for 60 min with vigorously shaking (250 rpm).
17. Spread 100 μL of the cells and DNA mixture onto an LB/ampicillin plate.
18. Incubate overnight at 37 °C.
19. *Day 2*: Perform colony PCR to check the size of the insert DNA. Prepare the colony PCR mix (10 μL) in 0.2 mL PCR tubes as in Table 8.
20. To create a stock of each individual colony, pick a small colony using a sterile pipette tip, streak the pipette tip onto another LB/ampicillin plate, and grow at 37 °C for more than 6 h.

21. Run the PCR according to the following parameters: (1) 2 min at 94 °C; (2) 25 cycles of a sequence composed of 30 s at 94 °C/1 min at 55 °C/1 min at 72 °C; and (3) 5 min at 72 °C.
22. Load 10 μL of each colony PCR product directly onto a 1% agarose gel, alongside an appropriate DNA ladder. After ethidium bromide staining, observe the gel by UV illumination. The expected size of the colony PCR product is approximately 2500 bp.
23. Select several positive colonies, and use them to inoculate 6 mL of LB/ampicillin medium. Shake the cultures overnight at 37 °C and 250 rpm.
24. *Day 3*: Add 1 mL of each saturated overnight culture to 0.5 mL of 50% glycerol in a 2-mL screw top tube or cryovial and gently mix. Freeze the glycerol stock tube at −80 °C. The stock is now stable for years, as long as it is kept at −80 °C.
25. Extract and purify the plasmid DNA from the remaining 5 mL of each culture using a QIAquick spin miniprep kit, according to the manufacturer's instructions. Measure the DNA concentration using a NanoDrop spectrophotometer.
26. Using the vector primers, SFseq-F and SFseq-R, perform sequence analysis to confirm that the nucleotide sequence of the plasmids is correct.
27. *Day 4*: Chill 1 μL of the positive pFBgp67-iRATmC-based entry plasmid in a 1.5-mL centrifuge tube.
28. Thaw the *E. coli* DH10Bac competent cells (10 μL) on ice.
29. Add 10 μL of DH10Bac competent cells to the plasmid DNA. Mix gently by pipetting up and down to mix the cells and DNA. Do not vortex.
30. Place the mixture on ice for 30 min.
31. Heat shock at 42 °C for 45 s.
32. Place the mixture on ice for 2 min.
33. Add 90 μL of pre-warmed SOC medium to the tube.
34. Incubate the tube at 37 °C for 4 h with vigorous shaking (250 rpm).
35. Dilute the transformation culture with SOC medium as appropriate, and then spread 100 μL of cells onto an LB/KTGIB agar plate.
36. Incubate the plate for 2 days at 37 °C.
37. *Day 6*: Pick several white colonies, and use them to inoculate 6 mL of LB/KTG medium. Shake the cultures overnight at 37 °C and 250 rpm.

38. *Day 7*: Add 1 mL of each saturated overnight culture to 0.5 mL of 50% glycerol in a 2-mL cryovial and gently mix. Freeze the glycerol stock tube at −80 °C. The stock is now stable for years, as long as it is kept at −80 °C.
39. Extract and purify the bacmid DNA from the remaining 5 mL of each culture using a QIAquick spin miniprep kit, according to the manufacturer's instructions. Measure the DNA concentration using a NanoDrop spectrophotometer. The recombinant bacmid DNA is now ready for transfection into insect cells to generate the baculovirus.
40. Take an aliquot of a properly maintained *Sf*9 stock culture, count cells, and determine their viability (*see* **Note 7**).
41. Seed 1×10^6 of *Sf*9 cells in 2 mL of PSFM-J1/FBS-PS medium per well of six-well plate.
42. Incubate the cells at 27 °C until they attach (~15 min).
43. While waiting for the cells to attach, add 6 μL of FuGENE transfection reagent to 140 μL of 1× Grace's medium in a 1.5-mL centrifuge tube for each transfection. Add 1 μg of bacmid DNA to the FuGENE/1× Grace media mixture, incubate for 15 min at 27 °C, and then add 1× Grace's medium to make a total volume of 500 μL.
44. Remove 2 mL of the supernatant from the seeded cells and replace with 500 μL of the FuGENE/bacmid DNA mixture dropwise onto the cells. Ensure even dispersal.
45. Incubate the cells for 4 h at 27 °C, aspirate the transfectant-DNA suspension, add 2 mL of fresh PSFM-J1/FBS-PS medium, and return to 27 °C (make sure to put the plates in a sealable plastic bag to prevent strong evaporation of the reagent/medium).
46. Incubate the cells for 96 h at 27 °C.
47. *Day 11*: Collect the supernatant containing P0 virus (~2 mL from each well), and filter the medium containing P0 virus into a 2-mL centrifuge tube using a 3-mL syringe fitted with a small 0.2-μm filter. This virus is a stock of P0 virus that should be stored at 4 °C, protected from light.
48. Prepare a 125-mL Erlenmeyer flask (Vent filter cap/Baffle bottom) containing 50 mL of *Sf*9 cells in the PSFM-J1/FBS-PS medium at a density of 1.5×10^6 cells/mL in the exponential growth phase.
49. Infect with 500 μL of P0 virus.
50. Incubate the cells for 96 h at 27 °C with agitation (120 rpm).
51. *Day 15*: Transfer the suspension into a 50-mL tube, and centrifuge for 10 min at 2380 × *g* at 4 °C.
52. Collect the supernatant into a fresh 50-mL tube.

53. Store at 4 °C protected from light. This virus is the P1 virus stock. P1 is sufficient for initial protein expression studies. If large volumes of virus are required, amplify P1 to obtain P2.
54. Determine the titer of the P1 virus using the endpoint dilution assay.

3.2.2 Secretory Expression Using Sf9 Insect Cells/Baculoviruses

1. *Day 1*: Expansion of *Sf*9 cells should be prepared in advance, so that a sufficient number of cells is available on this day. To expand *Sf*9 cells, determine the total number of cells and percent viability using a LUNA automated cell counter and Trypan blue exclusion, and ensure that the density of the cells is 1.5×10^6 cells/mL. Based on the volume of cells needed (0.5–2 L), calculate the volume of the medium needed to dilute the culture. For 0.5 L, 7.5×10^8 cells and ~ 0.4 L of PSFM-J1/FBS-PS medium will be needed and a 3-L Fernbach flask (Vent filter cap/Baffle bottom).
2. Add P1 virus at a multiplicity of infection (MOI) of 1 to infect *Sf*9 cells at a density of 1.5×10^6 cells/mL (*see* **Note 8**).
3. Incubate the cells at 27 °C with agitation (120 rpm).
4. *Days 2–4*: Monitor the mCherry-derived fluorescence in the culture supernatant over the elapsed time. Take an aliquot of the *Sf*9 culture, spin down the cells using a table top centrifuge, and use 200 μL of the supernatant to measure the fluorescence (excitation: 587 nm; emission maximum: 610 nm).
5. *Day 4*: Collect the culture supernatant 96 h after infection by centrifugation for 15 min at 6000 × *g* at 4 °C. Discard the cell pellet. Adjust the pH of the supernatant to ~8.0 by adding 1/20 volume of 1 M Tris-HCl (pH 8.0).

3.2.3 Purification of Genetically Encoded Fluorescent Fab Fragments

1. *Day 1*: (IMAC) Use 20 mL of Ni Sepharose excel resin (40 mL of 50% slurry) per 1 L of *Sf*9 culture, pour the slurry into a glass Econo-Column, and equilibrate with five CVs of buffer A.
2. Add the *Sf*9 culture supernatant directly to the column, and collect the flow-through fraction.
3. Wash the column with 10 CVs of buffer A.
4. Elute the polyprotein of interest from the column by adding three CVs of buffer B and collect the eluate.
5. Measure the protein concentration of the eluate using a Nano-Drop spectrophotometer.
6. Concentrate the protein using an Amicon Ultra-15 (MWCO 30 K) centrifugal concentrator to the volume of 5 mL.
7. (SEC) Inject the 5 mL of concentrated protein sample onto a HiLoad16/60 Superdex200 column pre-equilibrated with TBS at a flow rate of 1 mL/min.

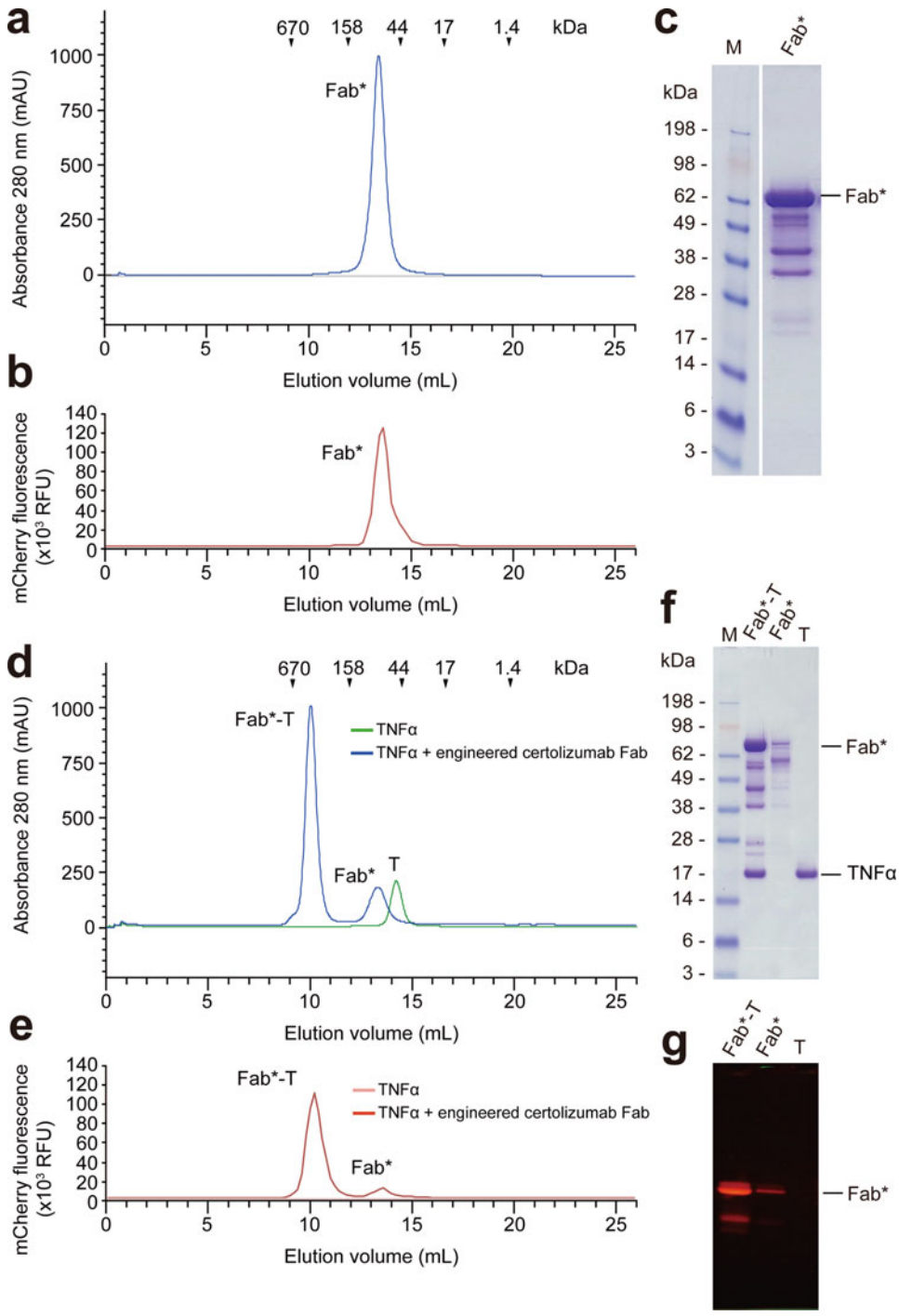


Fig. 5 iRAT-based production and characterization of mCherry-engineered certolizumab Fab. (**a**) The concentrated final product from the Subheading 3.2.3 was analyzed by SEC on a Superdex200 Increase 10/300GL column. (**b**) FSEC profile. The fractions collected in (**a**) were subjected to mCherry fluorescence measurement (excitation: 587 nm, emission maximum: 610 nm). (**c**) SDS-PAGE analysis of the peak fraction (labeled by Fab*) of SEC in (**a**). The existence of multiple extra bands could be attributed to DsRed-specific fragmentation

8. *Day 2*: Collect 2 mL fractions. Analyze the UV trace, and pool the peak fractions.
9. Perform SDS-PAGE to check the purity.
10. Measure the protein concentration of each fraction using a NanoDrop spectrophotometer to estimate the protein yield.
11. Use an aliquot to evaluate the biophysical and/or functional properties of the purified mCherry-engineered Fab fragment (Fig. 5 as a typical example). Ideally, binding assays, such as SPR, ITC, or ELISA, should provide valuable information.
12. Concentrate the remaining purified fluorescent Fab fragment using an Amicon Ultra-15 (MWCO 30 K) centrifugal concentrator to an appropriate protein concentration.
13. Re-measure the protein concentration of each fraction using a NanoDrop spectrophotometer.
14. Prepare 100 μL aliquots and flash-freeze with liquid nitrogen.
15. Store at −80 °C.

3.2.4 Purification of Tag-Free Fab Fragments

1. *Day 1*: (IMAC) After collecting the *Sf*9 culture supernatant, perform the same procedures as **steps 1–5** in Subheading 3.2.3.
2. (TEV cleavage and dialysis) Add an appropriate amount (typically at a protease to target protein ratio of 1:100 (w/w) or 3 mg) of TEV-His_6 to the elution fraction.
3. Transfer the mixture into dialysis tubing, and dialyze overnight at 4 °C against 2 L of TBS.
4. *Day 2*: (Reverse IMAC) Equilibrate a 5-mL HisTrap HP column with 25 mL of buffer C.
5. Remove the dialysate from the tubing and inject it into the column using a 10-mL syringe.

Fig. 5 (continued) caused by boiling before SDS-PAGE as described in [18]. (**d**) SEC profile of the mCherry-engineered certolizumab Fab-TNFα complex in the presence of excess engineered Fab on a Superdex200 10/300GL column (blue line). As a reference, TNFα alone was separated on the same column (green line). Elution volumes of protein standards are indicated at the top. *Peak Fab*-T* engineered certolizumab Fab-TNFα trimer complex, *peak T* free TNFα trimer, and *peak Fab** free engineered certolizumab Fab. (**e**) FSEC profile. The fractions collected in (**d**) were subjected to mCherry fluorescence measurement. Red line: mixture of the engineered certolizumab Fab and TNFα trimer; pink line: TNFα trimer alone. (**f**) SDS-PAGE analysis of the corresponding peak fractions in (**d**). (**g**) In gel fluorescence analysis of mCherry-engineered certolizumab Fab. The same peak fractions as in (**f**) were added 1:1 with buffer containing 50 mM Tris-HCl (pH 7.6), 50 mM DTT, 5% glycerol, 5% SDS, 5 mM EDTA, and 0.02% bromophenol blue. Without boiling the sample, electrophoresis was performed on a Novex WedgeWell 10% Tris-glycine gel (Thermo Fisher Scientific) at 100 V for 130 min. For mCherry, protein bands were observed on a LED transilluminator. The theoretical molecular masses of the red fluorescent certolizumab Fab and TNFα trimer are 81.2 kDa and 51.9 kDa (17.3 kDa protomer ×3), respectively

6. Collect the flow-through fraction (FT).
7. Wash the column three times, each wash containing one CV (5 mL) of buffer C and collect as W1, W2, and W3 fractions, respectively.
8. Elute the bound material (excised fluorescent iRATmC segment, TEV-His_6 and contaminating proteins) from the column by adding three CVs (15 mL) of buffer D and collect as the E fraction.
9. Measure the protein concentration of each fraction using a NanoDrop spectrophotometer.
10. Mix the FT and W1 fractions and concentrate the protein using an Amicon Ultra-15 (MWCO 30 K) centrifugal concentrator to a volume of 5 mL.
11. (SEC) Inject the 5 mL of concentrated protein sample onto a HiLoad16/60 Superdex200 column pre-equilibrated with TBS at a flow rate of 1 mL/min.
12. *Day 3*: Collect 2 mL fractions. Analyze the UV trace, and pool the peak fractions.
13. Perform SDS-PAGE to check the purity.
14. Measure the protein concentration of each fraction using a NanoDrop spectrophotometer to estimate the protein yield.
15. Use an aliquot to evaluate the biophysical and/or functional properties of the purified tag-free Fab fragment (Fig. 6 as a typical example). Ideally, binding assays, such as SPR, ITC, or ELISA, should provide valuable information.
16. Concentrate the remaining purified Fab fragment using a Amicon Ultra-15 (MWCO 10 K) centrifugal concentrator to an appropriate protein concentration (e.g., 10 mg/mL).
17. Re-measure the protein concentration of each fraction using a NanoDrop spectrophotometer.
18. Prepare 100 μL aliquots, and flash-freeze with liquid nitrogen.
19. Store at −80 °C.

4 Notes

1. Antibody sequencing from monoclonal hybridoma cell lines can be performed as described previously [14]. Amino acid sequence data for the known therapeutic antibodies can be retrieved from DrugBank (http://www.drug-bank.ca/).
2. The TEV-His_6 was prepared as described previously [15]. It is not necessary to add any additional DTT to the TEV-His_6 from that already contained in its storage buffer.

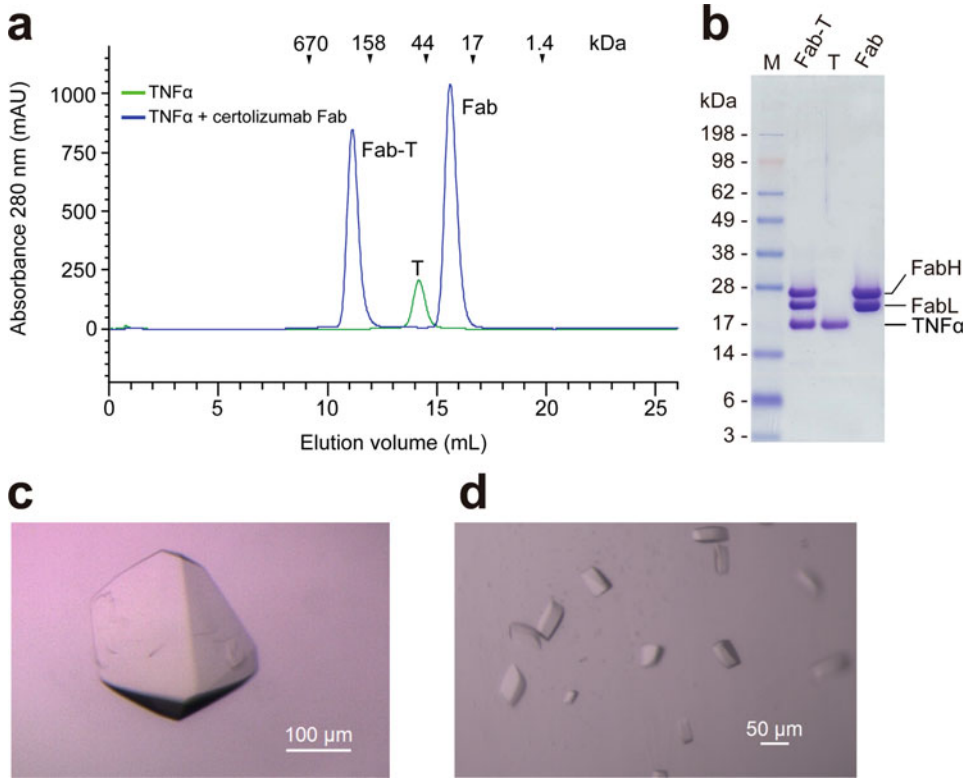


Fig. 6 Certolizumab Fab produced via the *Sf*9-iRAT system. (**a**) Typical profile in SEC of the certolizumab Fab-TNFα complex in the presence of excess certolizumab Fab on a Superdex200 10/300GL column (blue line). As a reference, TNFα alone was separated on the same column (green line). Elution volumes of protein standards are indicated at the top. *Peak Fab-T* certolizumab Fab-TNFα trimer complex, *peak T* free TNFα trimer, and *peak Fab* free certolizumab Fab. (**b**) SDS-PAGE analysis of the corresponding peak fractions in (**a**). The theoretical molecular masses of the certolizumab Fab and TNFα trimer are 47.6 kDa (Fab light chain: 23.1 kDa, Fab heavy chain: 24.5 kDa) and 51.9 kDa (17.3 kDa protomer ×3), respectively. (**c**) Crystal of the certolizumab Fab apo. (**d**) Crystal of the certolizumab Fab-TNFα complex

3. Typically, it takes 6–7 h to reach the desired OD_{600nm}.
4. We usually screen 12–24 neomycin-resistant colonies. The positive rate is approximately 65–70%.
5. The built-in mCherry within the iRAT segment can be replaced with another fluorescent protein such as EGFP or mBanana, depending on the experimental design.
6. Bypassing the papain digestion process is one technical advantage of the iRAT-mediated Fab production. In conventional preparation for Fabs from whole IgG, papain digestion often results in a combination of under- and over-digestion and requires specific optimization depending on different IgG subclasses.

7. Maintenance and management of insect cell culture can be performed as described previously [16].
8. We usually add ~1.5 mL of P1 baculovirus stock per 1 L of *Sf*9 culture.

Acknowledgments

We thank Shoichiro Horita, Hidetsugu Asada, and Tomoko Uemura for sharing their experiences on *Sf*9-baculovirus expression. The creation of the iRAT-based fluorescent Fab was conceived after being inspired by the pioneering work on multicolored fluorescent scFv by Markiv et al. [17]. This work was funded by the Strategic Basic Research Program from the Japan Science and Technology Agency (JST), the Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS) from the Japan Agency of Medical Research and Development (AMED), the Research on Development of New Drugs from the AMED, and the Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) (Nos. 15K06968 and 18K05334).

References

1. Griffin L, Lawson A (2011) Antibody fragments as tools in crystallography. Clin Exp Immunol 165:285–291
2. Lieberman RL, Culver JA, Entzminger KC et al (2011) Crystallization chaperone strategies for membrane proteins. Methods 55:293–302
3. Hino T, Arakawa T, Iwanari H et al (2012) G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. Nature 482:237–240
4. Nomura N, Verdon G, Kang HJ et al (2015) Structure and mechanism of the mammalian fructose transporter GLUT5. Nature 526:397–401
5. Asada H, Horita S, Hirata K et al (2018) Crystal structure of the human angiotensin II type 2 receptor bound to an angiotensin II analog. Nat Struct Mol Biol 25:570–576
6. Nagarathinam K, Nakada-Nakura Y, Parthier C et al (2018) Outward open conformation of a Major Facilitator Superfamily multidrug/H^+ antiporter provides insights into switching mechanism. Nat Commun 9:4005
7. Toyoda Y, Morimoto K, Suno R et al (2019) Ligand binding to human prostaglandin E receptor EP4 at the lipid-bilayer interface. Nat Chem Biol 15:18–26
8. Wu S, Avila-Sakar A, Kim J et al (2012) Fabs enable single particle cryoEM studies of small proteins. Structure 20:582–592
9. Koehl A, Hu H, Maeda S et al (2018) Structure of the μ-opioid receptor-Gi protein complex. Nature 558:547–552
10. Nomura Y, Sato Y, Suno R et al (2016) The intervening removable affinity tag (iRAT) production system facilitates Fv antibody fragment-mediated crystallography. Protein Sci 25:2268–2276
11. Horita S, Nomura Y, Sato Y et al (2016) High-resolution crystal structure of the therapeutic antibody pembrolizumab bound to the human PD-1. Sci Rep 6:35297
12. Ono M, Horita S, Sato Y et al (2018) Structural basis for tumor necrosis factor blockade with the therapeutic antibody golimumab. Protein Sci 27:1038–1046
13. Koerber JT, Hornsby MJ, Wells JA (2015) An improved single-chain Fab platform for efficient display and recombinant expression. J Mol Biol 427:576–586
14. Fields C, O'Connell D, Xiao S et al (2013) Creation of recombinant antigen-binding molecules derived from hybridomas secreting specific antibodies. Nat Protoc 8:1125–1148

15. Lucast LJ, Batey RT, Doudna JA (2001) Large-scale purification of a stable form of recombinant tobacco etch virus protease. Biotechniques 30:544–546, 548, 550 passim
16. Abdulrahman W, Radu L, Garzoni F et al (2015) The production of multiprotein complexes in insect cells using the baculovirus expression system. Methods Mol Biol 1261:91–114
17. Markiv A, Beatson R, Burchell J et al (2011) Expression of recombinant multi-coloured fluorescent antibodies in gor-/trxB- E. coli cytoplasm. BMC Biotechnol 11:117
18. Gross LA, Baird GS, Hoffman RC et al (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral. Proc Natl Acad Sci U S A 97:11990–11995

Chapter 6

Isolation of Artificial Binding Proteins (Affimer Reagents) for Use in Molecular and Cellular Biology

Anna A. S. Tang, Christian Tiede, Michael J. McPherson, and Darren C. Tomlinson

Abstract

Artificial binding proteins have been validated as alternatives to antibodies in their use as research reagents in molecular and cellular biology. For example, they have been used as inhibitors of protein–protein interactions to modulate activity, to facilitate crystallization, and as probes for cellular imaging.

Phage display is a widely used approach for isolating target-specific binding reagents, and it has even been used to isolate isoform-specific binding proteins and binders that can distinguish between highly homologous protein domains.

Here, we describe methods that have been employed in isolating highly specific artificial binding proteins against a wide range of target proteins.

Key words Artificial binding protein, Affimer, Research reagents, Protein–protein interaction, Inhibitor, Phage display, Panning, Phage ELISA

1 Introduction

Antibodies are the most widely used research reagents for protein binding; however, they have numerous limitations and issues with validation and reproducibility [1–3]. Consequently, artificial binding proteins such as DARPins [4], monobodies [5], and Affimer reagents [6–9] have been developed as promising alternatives with the added benefits of bacterial production for ease of manufacturability and mammalian cell expression to study protein function. Affimer reagents are based on two types of closely related scaffolds—a mammalian scaffold based on Stefin A [6, 7] and the Adhiron scaffold based on a consensus sequence of phytocystatins [8]. Both scaffolds were engineered to constrain two variable peptide sequences for molecular recognition (Fig. 1) [8]. Affimer reagents have been utilized as affinity reagents in various applications, for example, as detection reagents for diagnostics, as inhibitors of protein–protein interactions to modulate activity, as

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_6,

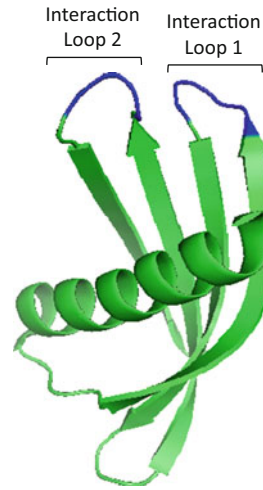


Fig. 1 The full-length Adhiron scaffold as determined by X-ray crystallography at 1.75 Å (PDB: 4N6T). The ligand binding sequences within each interaction loop (VVAG in loop 1 and PWE in loop 2) are highlighted in dark blue, and these were replaced with nine randomized amino acids each (excluding cysteine residues) in the library design [8]

chaperones to facilitate crystallization, and as probes for cellular imaging [8–30].

Phage display remains the most widely used approach for isolating protein binding reagents, despite the growing number of alternative methods such as ribosome display [31], mRNA display [32], and cis display [33]. Phage display was first described by George P. Smith in 1985 [34] and has since been adapted to display large libraries of antibody fragments, peptides, and artificial binding proteins. Large libraries of Affimer reagents have been produced that display Affimer proteins on a truncated pIII coat protein on filamentous (Ff) phage [8]. Affimer reagents that specifically bind a wide range of target proteins have been successfully isolated from these libraries by phage display.

Selection of Affimer reagents by phage display is usually conducted over three panning rounds. Biotinylated target is immobilized on a streptavidin- or NeutrAvidin-coated surface and incubated with a population of Affimer-displaying phage. After washing away unbound phage, bound phage is eluted and propagated for subsequent rounds of panning. A typical method of Affimer selection by phage display is depicted in Fig. 2.

The described method has been used to isolate isoform-specific Affimer reagents [24, 25, 35]. The population of phage is first incubated with the homologous protein to remove binders to the isoform, prior to selection against the target protein (as depicted in Fig. 2). We have also used this method of negative selection to isolate specific binders of targets expressed as fusion proteins

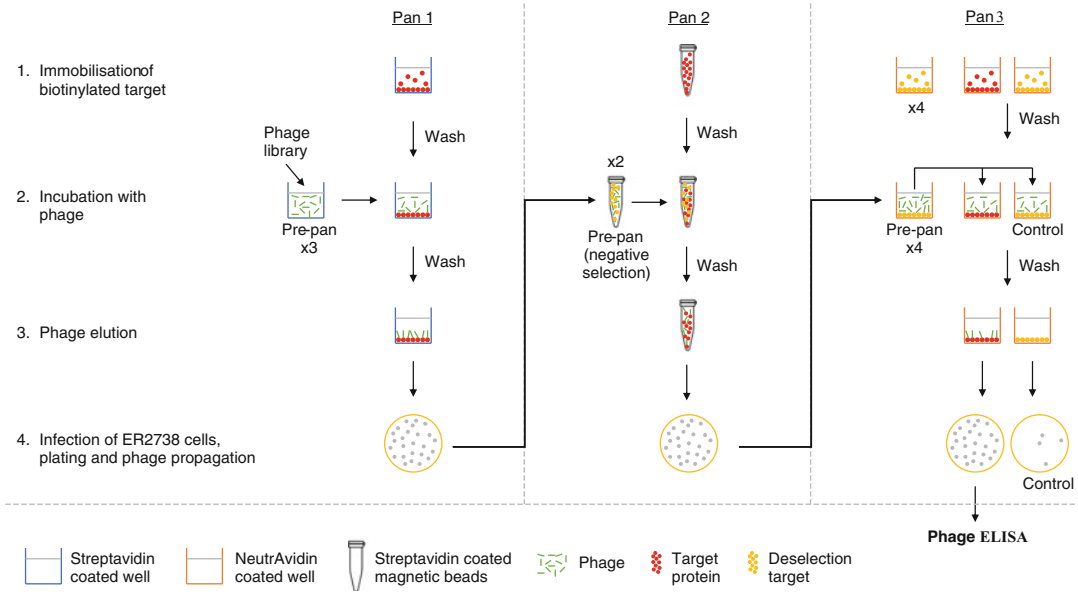


Fig. 2 Affimer selection by phage display conducted over three panning rounds. In each panning round, immobilized target (red spots) is incubated with a population of phage (green bars). After washing away any unbound phage, bound phage is eluted and propagated for subsequent rounds of panning. The population of phage can also be preincubated with a deselection target, to remove binders to the deselection target, and this is shown from Pan 2 onwards in the example depicted. After the final round of panning, selected phage clones are tested for binding to target by phage ELISA. (Figure adapted from Tang AA et al. (2017) [35])

(difficult-to-express proteins are often expressed as fusion proteins to improve solubility). The phage population can be pre-panned against the fusion tag alone, prior to incubating with the fusion protein. Alternatively, the method can also be adapted to isolate cross-reactive binders, for example, binders of highly homologous proteins or isoforms of proteins from different species. To do this, the target protein is alternated between panning rounds.

2 Materials

2.1 Affimer Selection by Phage Display

1. Affimer phage library (phagemid display on pIII).
2. Biotinylated target protein (1 μg required for each panning round).
3. Deselection target (optional—biotinylated/non-biotinylated.
4. ER2738 strain of *Escherichia coli* cells (Lucigen).
5. M13KO7 helper phage (New England BioLabs).
6. Pierce Streptavidin-coated, high binding capacity, eight-well strips (ThermoFisher Scientific, catalog no. 15501).

7. Pierce NeutrAvidin-coated, high binding capacity, eight-well strips (ThermoFisher Scientific, catalog no. 15508).
8. Dynabeads MyOne Streptavidin T1 (ThermoFisher Scientific, catalog no. 65601 or 65602).
9. Eppendorf LoBind microcentrifuge tubes (Eppendorf, catalog no. 0030 108.116).
10. Magnetic separation rack for 1.5-ml microcentrifuge tubes—for example, DynaMag™-2 Magnet (ThermoFisher Scientific, catalog no. 12321D).
11. PBS: Prepare a 10× stock by dissolving 80.0 g of NaCl (MW = 58.44), 2.0 g of KCl (MW = 74.55), 14.4 g of Na_2HPO_4 (MW = 141.96), and 2.4 g of KH_2PO_4 (MW = 136.09) in 800 ml of dH_2O. Adjust the pH to 7.4 with HCl and add dH_2O until a total volume of 1 l. Autoclave to sterilize and store at room temperature. For a 1× stock, dilute the 10× stock with dH_2O and adjust the pH to 7.4 with HCl.
12. PBST: Dilute the 10× PBS to 1× with dH_2O and 0.1% (v/v) Tween 20, and adjust to pH 7.4 with HCl.
13. 10× Casein Blocking Buffer (Sigma-Aldrich, catalog no. B6429).
14. 2× Blocking Buffer: Dilute 10× Casein Blocking Buffer five-fold in PBS.
15. 2YT medium: Dissolve 16 g of Bacto tryptone, 10 g of Bacto yeast extract, and 5 g of NaCl (MW = 58.44) in 900 ml of dH_2O. Adjust the pH to 7.0 with NaOH and add dH_2O until a total volume of 1 l. Autoclave to sterilize and store at room temperature.
16. Tetracycline hydrochloride (1000× stock): Dissolve tetracycline hydrochloride (12 mg/ml) in 70% ethanol. Store at −20 °C, protected from light. Working concentration is 12 μg/ml.
17. 0.2 M glycine, pH 2.2: Dissolve 1.5 g of glycine (MW = 75.07) in 80 ml of dH_2O. Adjust the pH to 2.2 with HCl and add dH_2O until a total volume of 100 ml. Autoclave to sterilize and store at room temperature.
18. 1 M Tris-HCl, pH 9.1: Dissolve 121.14 g of Tris base (MW = 121.14) in 800 ml of dH_2O. Adjust the pH to 9.1 with HCl and add dH_2O until a total volume of 1 l. The pH of Tris buffers varies with temperature and concentration; therefore, adjust the pH at room temperature. Autoclave to sterilize and store at room temperature.
19. Triethylamine: Dilute 14 μl of triethylamine (Sigma-Aldrich, catalog no. T0886) in 986 μl of PBS, immediately before use.

20. 1 M Tris-HCl, pH 9.1: Dissolve 121.14 g of Tris base (MW = 121.14) in 800 ml of dH_2O. Adjust the pH to 9.1 with HCl and add dH_2O until a total volume of 1 l. The pH of Tris buffers varies with temperature and concentration; therefore, adjust the pH at room temperature. Autoclave to sterilize and store at room temperature.
21. 80% glycerol: Dilute glycerol with dH_2O. Autoclave to sterilize and store at room temperature.
22. Halt protease inhibitor cocktail (100×) (ThermoFisher Scientific, catalog no. 87786).
23. Carbenicillin (500× stock): Dissolve carbenicillin (50 mg/ml) in ddH_2O. Filter-sterilize and store at −20 °C. Working concentration is 100 μg/ml.
24. LB agar plates with carbenicillin: Add 32.0 g of LB agar (Lennox L agar) (ThermoFisher Scientific, catalog no. 22700) per liter of dH_2O into an appropriately sized bottle for autoclaving and swirl to mix. Autoclave to sterilize and cool to ~50–55 °C. Add carbenicillin to a final concentration of 100 μg/ml and swirl to mix. Flame the neck of the bottle, pour into petri dishes, and allow to solidify. Store at 4 °C for up to 2 months. Alternatively, the formulation for 1 l of Lennox L agar is as follows: 10 g of Bacto tryptone, 5 g of Bacto yeast extract, 5 g of NaCl (MW = 58.44), and 12 g of agar. Swirl to mix and autoclave to sterilize.
25. Kanamycin (500× stock): Dissolve kanamycin (25 mg/ml) in ddH_2O. Filter-sterilize and store at −20 °C. Working concentration is 50 μg/ml.
26. PEG-NaCl precipitation solution: Dissolve 200 g of PEG 8000 and 146.1 g of NaCl (MW = 58.44) in dH_2O to a total volume of 1 l. Autoclave to sterilize and store at room temperature.
27. TE buffer (10 mM Tris, 1 mM EDTA), pH 8.0: Prepare stock solutions of 1 M Tris-HCl (pH 8.0) and 0.5 M EDTA (pH 8.0). To prepare 1 M Tris-HCl (pH 8.0), dissolve 121.14 g of Tris base (MW = 121.14) in 800 ml of dH_2O. Adjust the pH to 8.0 with HCl (at room temperature) and add dH_2O until a total volume of 1 l. Autoclave to sterilize and store at room temperature. To prepare 0.5 M EDTA (pH 8.0), add 93.06 g of EDTA disodium salt (MW = 372.24) to 400 ml of dH_2O. Adjust the pH to 8.0 with NaOH—the disodium salt of EDTA will not fully dissolve until the pH of the solution is adjusted to pH 8.0. Once dissolved, adjust the volume to 500 ml with dH_2O. Autoclave to sterilize and store at room temperature. For 1 l of TE buffer, mix 10 ml of 1 M Tris-HCl (pH 8.0) and 2 ml of 0.5 M EDTA (pH 8.0) with 988 ml of sterile ddH_2O. Store at room temperature.

2.2 Screening Selected Phage Clones by Phage ELISA

1. 96-well V-bottom deep well plates (Greiner Bio-One, catalog no. 786201).
2. 2YT medium: Dissolve 16 g of Bacto tryptone, 10 g of Bacto yeast extract, and 5 g of NaCl (MW = 58.44) in 900 ml of dH_2O. Adjust the pH to 7.0 with NaOH and add dH_2O until a total volume of 1 l. Autoclave to sterilize and store at room temperature.
3. Carbenicillin (500× stock): Dissolve carbenicillin (50 mg/ml) in ddH_2O. Filter-sterilize and store at −20 °C. Working concentration is 100 μg/ml.
4. M13KO7 helper phage (New England BioLabs).
5. Kanamycin (500× stock): Dissolve kanamycin (25 mg/ml) in ddH_2O. Filter-sterilize and store at −20 °C. Working concentration is 50 μg/ml.
6. Pierce Streptavidin-coated, clear, 96-well plates (ThermoFisher Scientific, catalog no. 15126).
7. PBS: Prepare a 10× stock by dissolving 80.0 g of NaCl (MW = 58.44), 2.0 g of KCl (MW = 74.55), 14.4 g of Na_2HPO_4 (MW = 141.96) and 2.4 g of KH_2PO_4 (MW = 136.09) in 800 ml of dH_2O. Adjust the pH to 7.4 with HCl and add dH_2O until a total volume of 1 l. Autoclave to sterilize and store at room temperature. For a 1× stock, dilute the 10× stock with dH_2O and adjust the pH to 7.4 with HCl.
8. PBST: Dilute the 10× PBS to 1× with dH_2O and 0.1% (v/v) Tween 20, and adjust to pH 7.4 with HCl.
9. 10× Casein Blocking Buffer (Sigma-Aldrich, catalog no. B6429).
10. 2× Blocking Buffer: Dilute 10× Casein Blocking Buffer fivefold in PBS.
11. Biotinylated target protein (recommended 0.5–1.0 μg per well).
12. Biotinylated deselection target (optional) (recommended 0.5–1.0 μg per well).
13. Anti-Fd-Bacteriophage-HRP (Seramun Diagnostica GmbH, catalog no. A-020-1-HRP).
14. TMB (SeramunBlau® fast TMB/substrate solution) (Seramun, catalog no. S-001-TMB).
15. 0.2 M H_2SO_4 (optional).

3 Methods

3.1 Affimer Selection by Phage Display

Target proteins can be biotinylated using chemical conjugation methods, such as EZ-Link NHS-Biotin (ThermoFisher Scientific, catalog no. 20217), following the manufacturer's instructions. Chemical biotinylation may block the binding sites on target proteins. Alternatively, target protein sequence can be tagged with a biotin acceptor peptide (BAP) sequence (GLNDIFEAQKIEWHE), which is enzymatically biotinylated when expressed in *E. coli* in the presence of biotin ligase.

3.1.1 Panning Round 1

Day 1

1. Block the streptavidin-coated wells with 300 μl per well of 2× Blocking Buffer. Prepare four wells per selection—three wells for pre-panning the phage and one well to immobilize the target. Cover the wells with an adhesive plate seal and incubate overnight at 37 °C.
2. Set up an overnight culture of ER2738 *E. coli* cells. Pick a single colony into 5 ml of 2YT medium with 12 μg/ml tetracycline and incubate overnight in an orbital incubator at 37 °C, 230 rpm.

Day 2

3. Wash the blocked wells 3× with 300 μl of PBST and replace with 100 μl per well of 2× Blocking Buffer.
4. Immobilization of biotinylated target and deselection target (if using) (*see* **Note 1**). Add 1 μg of biotinylated target to the selection well (*see* **Note 2**). If required, add 1 μg of biotinylated deselection target into pre-pan wells 2 and 3. Proceed immediately to **step 5**.
5. Pre-pan 1. To the first pre-pan well, add 5 μl of Affimer phage library (~10^{12} colony-forming units). Cover with a plate seal and incubate for 40 min at room temperature on a vibrating platform shaker.
6. Pre-pan 2. Remove the blocking buffer/deselection target from pre-pan well 2. If deselection target was used, wash the well 6× with 300 μl of PBST. Transfer the phage from pre-pan well 1 to pre-pan well 2. Cover with a plate seal and incubate for 40 min at room temperature on the vibrating platform shaker.
7. Pre-pan 3. Remove the blocking buffer/deselection target from pre-pan well 3. If deselection target was used, wash the well 6× with 300 μl of PBST. Transfer the phage from pre-pan well 2 to pre-pan well 3. Cover with a plate seal and incubate for 40 min at room temperature on the vibrating platform shaker.

8. Selection. Remove the target from the selection well and wash 6× with 300 µl of PBST. Transfer the phage from pre-pan well 3 to the selection well. If required, add 1 µg of non-biotinylated deselection target (*see* **Note 3**). Cover with a plate seal and incubate for 2 h at room temperature on the vibrating platform shaker.
9. Wash the selection well 27× with 300 µl of PBST, preferably using an automated plate washer.
10. Elution 1. Add 100 µl of 0.2 M glycine, pH 2.2 to the selection well and incubate for 10 min at room temperature. Neutralize by adding 15 µl of 1 M Tris-HCl, pH 9.1, and transfer the eluted phage to a microcentrifuge tube.
11. Elution 2. Dilute 14 µl of triethylamine with 986 µl of PBS. Add 100 µl of the diluted triethylamine to the selection well and incubate for 6 min at room temperature. Neutralize by adding 50 µl of 1 M Tris-HCl, pH 7, and transfer the eluted phage to the microcentrifuge tube containing elution 1.
12. Infection of ER2738 cells. Before eluting the phage, set up an 8 ml culture of ER2738 cells for each selection. Dilute the overnight cultures of ER2738 cells in 2YT medium with 12 µg/ml tetracycline to an OD600 of ~0.2 and incubate at 37 °C, 230 rpm for about 1 h until OD600 reaches ~0.6. Add the eluted phage and incubate for 1 h at 37 °C at up to 90 rpm. Mix at least once during the incubation.
13. Plating the cells. Plate 1 µl of the phage-infected ER2738 cells (diluted in 100 µl of 2YT media) onto LB agar plates with 100 µg/ml carbenicillin (the selection antibiotic for the phage library). Centrifuge the remaining culture at 3000 × *g* for 5 min. Pour off the supernatant and resuspend the cell pellet in the residual supernatant. Plate the cells on LB agar with 100 µg/ml carbenicillin. Incubate the plates overnight at 37 °C.

Day 3

14. To estimate the phage titer, count the colonies on the plate containing 1 µl of phage-infected ER2738 cells. Multiply by 8000 to determine the total number per 8 ml of cells (usually between 0.1 and 2 × 10^6).
15. Scrape the cells from the plate containing the remaining culture. To do this, add 5 ml of 2YT medium with 12 µg/ml tetracycline and scrape using a disposable plastic spreader. Transfer the scraped cells to a 50 ml centrifuge tube. Use an additional 2 ml of medium to scrape off any remaining cells and mix with the cells in the centrifuge tube. Measure the OD600 of a 1/10 dilution to determine the volume required for an 8 ml culture at OD600 0.2.

16. Dilute the appropriate amount of cells in 2YT medium with 12 μg/ml tetracycline for an 8 ml culture at OD600 0.2. Incubate for about 1 h at 37 °C, 230 rpm until an OD600 of ~0.6 is achieved.
17. Add 3.5×10^{10} M13KO7 helper phage (multiplicity of infection ~30) to the culture and incubate at 37 °C, 90 rpm for 1 h.
18. Add 16 μl of 25 mg/ml kanamycin and incubate overnight at 25 °C, 170 rpm.

Day 4

19. Centrifuge the phage-infected cultures at 3500 × *g* for 10 min and transfer the phage-containing supernatant to fresh tubes. Remove the required amount of phage-containing supernatant for panning round 2.
20. Add 2 ml of PEG-NaCl precipitation solution to the remaining supernatant and incubate for at least 1 h or overnight at 4 °C.
21. Centrifuge at 4800 × *g* for 30 min to pellet the phage. Pour off the supernatant (blotting the tube on tissue paper to remove all of the supernatant) and resuspend the pellet in 320 μl of TE buffer.
22. Transfer to a microcentrifuge tube and centrifuge at 16,000 × *g* for 10 min. Transfer the phage-containing supernatant to a fresh tube. Phage can be stored for several weeks at 4 °C. For long-term storage, add an equal volume of 80% glycerol, mix thoroughly and store at −80 °C.

3.1.2 Panning Round 2

Magnetic separation and handling using Dynabeads MyOne Streptavidin T1 magnetic beads can easily be automated on a wide variety of liquid handling platforms. Panning Round 2 has been automated on a KingFisher Flex magnetic particle processor (ThermoFisher Scientific, catalog no. 5400630) and is described by Tang et al. [35]. However, this method describes how the panning can be done manually using a magnetic separation rack.

Day 1

1. Block the streptavidin-coated magnetic beads. Resuspend the Dynabeads MyOne Streptavidin T1 magnetic beads thoroughly by vortexing and transfer 20 μl per selection to an Eppendorf LoBind microcentrifuge tube. Block in a minimum of 300 μl 2× Blocking Buffer (or 100 μl 2× Blocking Buffer per 20 μl of streptavidin beads). Incubate overnight at room temperature on a tube rotator to keep the beads in suspension.
2. Set up an overnight culture of ER2738 *E. coli* cells. Pick a single colony into 5 ml of 2YT medium with 12 μg/ml tetracycline and incubate overnight in an orbital incubator at 37 °C, 230 rpm. If performing competitive panning, set this up at the end of day 2.

Day 2

3. Centrifuge the blocked streptavidin beads at 800 × *g* for 1 min. Place the tube on the magnetic separation rack to immobilize the beads. Remove the blocking buffer and replace with fresh 2× Blocking Buffer, resuspending the beads in 100 µl per 20 µl of streptavidin beads.
4. Immobilization of biotinylated target. In an Eppendorf LoBind microcentrifuge tube, add 1 µg of biotinylated target to 200 µl of 2× Blocking Buffer and 50 µl of the blocked streptavidin beads. Incubate at room temperature on a tube rotator to keep the beads in suspension.
5. Immobilization of biotinylated deselection target (if using) for pre-panning phage (*see* **Note 1**). Add 1 µg of biotinylated deselection target to 200 µl of 2× Blocking Buffer and 50 µl of the blocked streptavidin beads. Incubate for 1 h at room temperature on the tube rotator. Centrifuge the tube at 800 × *g* for 1 min. Wash the beads 6× with 500 µl of 2× Blocking Buffer using the magnetic separation rack to immobilize the beads. Resuspend the beads in 50 µl of 2× Blocking Buffer.
6. Pre-pan 1. In an Eppendorf LoBind microcentrifuge tube, add 125 µl of phage-containing supernatant from panning round 1 to 125 µl of 2× Blocking Buffer and 25 µl of the blocked streptavidin beads (from **step 3** or **step 5**, as appropriate). Incubate for 1 h at room temperature on the tube rotator.
7. Pre-pan 2. Centrifuge the pre-pan 1 tube at 800 × *g* for 1 min and place the tube on the magnetic separation rack to immobilize the beads. Transfer the phage-containing supernatant to the tube containing the remaining 25 µl of blocked streptavidin beads (from **step 3** or **step 5**, as appropriate). Incubate for 1 h at room temperature on the tube rotator.
8. Centrifuge the tube containing the immobilized target at 800 × *g* for 1 min. Wash the beads 6× with 500 µl of 2× Blocking Buffer using the magnetic separation rack to immobilize the beads.
9. Selection. Centrifuge the pre-pan 2 tube at 800 × *g* for 1 min and place the tube on the magnetic separation rack to immobilize the beads. Transfer the phage-containing supernatant to the tube containing the immobilized target. If desired, add 1 µg of non-biotinylated deselection target (*see* **Note 3**). Incubate for 1 h at room temperature on the tube rotator.
10. Centrifuge the tube at 800 × *g* for 1 min. Wash the beads 6× with 500 µl of 2× Blocking Buffer using the magnetic separation rack to immobilize the beads. For standard panning, elute the phage immediately as described in **steps 13** and **14** (*see* **Note 4**). For competitive panning, continue with the following step (*see* **Note 5**).

11. Resuspend the washed beads in the following: 77 µl of PBS, 60 µl of 10× Blocking Buffer, 60 µl of 80% glycerol, and 3 µl of Halt protease inhibitor cocktail (100×). If desired, add 1 µg of non-biotinylated deselection target (*see* **Note 3**). Incubate for 22–24 h at room temperature on the tube rotator.

Day 2 or 3

12. Centrifuge the tube at 800 × *g* for 1 min. Wash the beads 6× with 500 µl of 2× Blocking Buffer using the magnetic separation rack to immobilize the beads.
13. Elution 1. Resuspend the beads in 100 µl of 0.2 M glycine, pH 2.2 to elute the phage. Incubate for 10 min at room temperature, placing back on the magnet for the last minute. Transfer the eluted phage to fresh tubes containing 15 µl of 1 M Tris-HCl, pH 9.1 to neutralize.
14. Elution 2. Dilute 14 µl of triethylamine with 986 µl of PBS. Elute any remaining phage by resuspending the beads in 100 µl of the diluted triethylamine. Incubate for 6 min at room temperature, placing back on the magnet for the last minute. Transfer the eluted phage to fresh tubes containing 50 µl of 1 M Tris-HCl, pH 7 to neutralize. Pool with the eluted phage from Elution 1.
15. Infection of ER2738 cells. Before eluting the phage, set up an 8 ml culture of ER2738 cells for each selection. Dilute the overnight cultures of ER2738 cells in 2YT medium with 12 µg/ml tetracycline to an OD600 of ~0.2 and incubate at 37 °C, 230 rpm for about 1 h until OD600 reaches ~0.6. Add the eluted phage and incubate for 1 h at 37 °C at up to 90 rpm. Mix at least once during the incubation.
16. Plating the cells. Centrifuge the phage-infected culture at 3000 × *g* for 5 min. Pour off the supernatant and resuspend the cell pellet in the residual supernatant. Plate the cells on LB agar with 100 µg/ml carbenicillin. Incubate the plates overnight at 37 °C.

Day 3 or 4.

17. Propagate the phage from Panning Round 2 as described for Panning Round 1 (*see* Subheading 3.1.1, **steps 15–22**).

3.1.3 Panning Round 3

Day 1

1. Block the NeutrAvidin-coated wells with 300 µl per well of 2× Blocking Buffer. Prepare six wells per selection—four wells for pre-panning the phage, one for panning against the immobilized target, and a negative control well for panning against either a deselection target or an empty well. Cover the wells with an adhesive plate seal and incubate overnight at 37 °C.

2. Set up an overnight culture of ER2738 *E. coli* cells. Pick a single colony into 5 ml of 2YT medium with 12 μg/ml tetracycline and incubate overnight in an orbital incubator at 37 °C, 230 rpm. If performing competitive panning, set this up at the end of day 2.

Day 2

3. Wash the blocked wells 3× with 300 μl of PBST and replace with 100 μl per well of 2× Blocking Buffer.
4. Immobilization of biotinylated target and deselection target (if using) (*see* **Note 6**). Add 1 μg of biotinylated target to the selection well (*see* **Note 7**). If required, add 1 μg of biotinylated deselection target to the negative control well and pre-pan wells 1, 2, 3, and 4. Cover with a plate seal and incubate for 10 min at room temperature on a vibrating platform shaker.
5. Pre-pan 1. Remove the blocking buffer/deselection target from pre-pan well 1. If deselection target was used, wash the well 6× with 300 μl of PBST. Add 200 μl of phage-containing supernatant from panning round 2 and 20 μl of 10× Blocking Buffer. Cover with a plate seal and incubate for 1 h at room temperature on the vibrating platform shaker.
6. Pre-pan 2. Remove the blocking buffer/deselection target from pre-pan well 2. If deselection target was used, wash the well 6× with 300 μl of PBST. Transfer the phage from pre-pan well 1 to pre-pan well 2. Cover with a plate seal and incubate for 1 h at room temperature on the vibrating platform shaker.
7. Pre-pan 3. Remove the blocking buffer/deselection target from pre-pan well 3. If deselection target was used, wash the well 6× with 300 μl of PBST. Transfer the phage from pre-pan well 2 to pre-pan well 3. Cover with a plate seal and incubate for 1 h at room temperature on the vibrating platform shaker.
8. Pre-pan 4. Remove the blocking buffer/deselection target from pre-pan well 4. If deselection target was used, wash the well 6× with 300 μl of PBST. Transfer the phage from pre-pan well 3 to pre-pan well 4. Cover with a plate seal and incubate for 1 h at room temperature on the vibrating platform shaker. If required, add 1 μg of non-biotinylated deselection target to the pre-panned phage prior to selection (*see* **Note 3**).
9. Selection. Wash the selection well containing the immobilized target and the negative control well 6× with 300 μl of PBST. Transfer 100 μl of pre-panned phage from pre-pan well 4 into the selection well and 100 μl into the negative control well. Cover with a plate seal and incubate for 30 min at room temperature on the vibrating platform shaker.

10. Wash the selection well and negative control well 27× with 300 µl of PBST, preferably using an automated plate washer. For standard panning, elute the phage immediately as described in **steps 13** and **14** (*see* **Note 4**). For competitive panning, continue with the following step (*see* **Note 5**).
11. Add the following to the selection well and negative control well: 80 µl of 2× Blocking Buffer, 20 µl of 80% glycerol and 1 µl of Halt protease inhibitor cocktail (100×). If desired, add 1 µg of non-biotinylated deselection target (*see* **Note 3**). Cover with a plate seal and incubate for 22–24 h at room temperature on the vibrating platform shaker.

Day 2 or 3

12. Wash the selection well and negative control well 6× with 300 µl of PBST.
13. Elution 1. Add 100 µl of 0.2 M glycine, pH 2.2 to the selection well and negative control well. Incubate for 10 min at room temperature. Neutralize by adding 15 µl of 1 M Tris-HCl, pH 9.1 and transfer the eluted phage to a microcentrifuge tube.
14. Elution 2. Dilute 14 µl of triethylamine with 986 µl of PBS. Add 100 µl of the diluted triethylamine to the selection well and negative control well. Incubate for 6 min at room temperature. Neutralize by adding 50 µl of 1 M Tris-HCl, pH 7 and transfer the eluted phage to the microcentrifuge tube containing elution 1.
15. Infection of ER2738 cells. Before eluting the phage, set up 5 ml cultures of ER2738 cells for each selection. Dilute the overnight cultures of ER2738 cells in 2YT medium with 12 µg/ml tetracycline to an OD600 of ~0.2 and incubate at 37 °C, 230 rpm for about 1 h until OD600 reaches ~0.6. Add the eluted phage and incubate for 1 h at 37 °C at up to 90 rpm. Mix at least once during the incubation.
16. Plating the cells. Plate a range of volumes (e.g., 0.01 µl, 0.1 µl, 1 µl, 10 µl, and 100 µl) onto LB agar plates with 100 µg/ml carbenicillin. Also, centrifuge and plate the remaining cells as described for Panning Round 1 (*see* Subheading 3.1.1, **step 13**). For the negative control, select one volume to plate (usually 10 µl). Incubate the plates overnight at 37 °C.

Day 3 or 4

17. Compare the number of colonies on the negative control plate with the selection plate containing the same volume of cells. A successful screen shows enrichment of target-specific phage with fewer colonies (less than half) on the negative control plate. If background is still high on the negative control plate,

further rounds of panning should be performed (e.g., fourth round of panning in streptavidin-coated wells). The target-specific phage clones can be further screened by phage ELISA (*see* Subheading 3.2).

3.2 Screening Selected Phage Clones by Phage ELISA

3.2.1 Expressing Selected Phage Clones in a 96-Well Culture Plate

1. Aliquot 200 μl per well of 2YT medium with 100 μg/ml carbenicillin into a 96-well V-bottom deep well plate using a multichannel pipette and sterile reservoir.
2. Pick individual colonies from the final panning round of phage display to inoculate into the wells—pick 48 colonies or whatever the amount you want to test.
3. Incubate overnight at 37 °C, 1050 rpm in an incubating microplate shaker (e.g., Heidolph Incubator 1000 and Titramax 1000).
4. Aliquot 200 μl per well of 2YT medium with 100 μg/ml carbenicillin into a new 96-well V-bottom deep well plate.
5. Transfer 25 μl per well of the overnight cultures into this new plate and incubate for 1 h at 37 °C, 1050 rpm in the incubating microplate shaker. It is important to keep the remaining culture plate overnight (store at 4 °C for up to a week or use it to prepare fresh cultures for glycerol stocks).
6. Dilute M13K07 helper phage (titer ~10^{14}/ml) 1/1000 in 2YT medium with 100 μg/ml carbenicillin and add 10 μl per well to the freshly grown cultures using a multichannel pipette. Incubate for 30 min at 37 °C, 450 rpm in the incubating microplate shaker.
7. Dilute the kanamycin stock (25 mg/ml) 1/20 in 2YT medium with 100 μg/ml carbenicillin and add 10 μl per well to the phage-infected cultures using a multichannel pipette. Incubate overnight at 25 °C, 750 rpm in the incubating microplate shaker.
8. Centrifuge the phage-infected culture plate at 3500 × *g* for 10 min.
9. The supernatant contains the phage and can be removed directly into the ELISA plate to test for binding to target.

3.2.2 Phage ELISA

A phage ELISA is performed to test binding of selected phage clones to target protein. The phagemid DNA from specific phage binders can then be extracted and identified by DNA sequencing.

1. Block the streptavidin-coated 96-well plate with 300 μl per well of 2× Blocking Buffer. Cover the wells with an adhesive plate seal and incubate overnight at 37 °C.
2. Wash the blocked wells 3× with 300 μl of PBST, preferably using a 96-well plate washer.

3. Immobilization of biotinylated target (*see* **Note 8**) and deselection target (if using) (*see* **Note 9**). Dilute the biotinylated target and deselection target (if using) to 10 μg/ml (or other appropriate concentration, *see* **Note 10**) with 2× Blocking Buffer. Add 50 μl per well of diluted target to the first 6 columns of the plate and 50 μl per well of deselection target (if using) or 2× Blocking Buffer to the last six columns of the plate. Incubate for 1 h at room temperature on a vibrating platform shaker.
4. Wash the wells 3× with 300 μl of PBST and add 10 μl per well of 10× Blocking Buffer into all wells using a multichannel pipette.
5. Addition of phage clones. Add 40 μl per well of phage-containing supernatant so that each one is tested against the target and a negative control well (e.g., phage clone A1 is added to wells A1 and A7). Incubate for 1 h at room temperature on the vibrating platform shaker.
6. Wash the wells 3× with 300 μl of PBST, preferably using a 96-well plate washer.
7. Detection of phage with anti-Fd-bacteriophage antibody. Dilute anti-Fd-bacteriophage-HRP 1/1000 in 2× Blocking Buffer and add 50 μl per well. Incubate for 1 h at room temperature on the vibrating platform shaker.
8. Wash the wells 10× with 300 μl of PBST, preferably using a 96-well plate washer.
9. TMB detection of HRP. Add 50 μl per well of TMB and allow to develop. (Note the amount of time the plate is allowed to develop—usually 2–3 min) Measure absorbance at 620 nm.
10. Optional: To stop the reaction, add 50 μl per well of 0.2 M H_2SO_4. Measure absorbance at 450 nm.

4 Notes

1. A biotinylated deselection target can be immobilized for pre-panning of phage.
2. The binding capacity of Pierce streptavidin-coated, high binding capacity, eight-well strips (ThermoFisher Scientific, catalog no. 15501) is ~125 pmol D-biotin per well; therefore, the biotinylated target is added in excess and can be reduced, if required.
3. A non-biotinylated deselection target can be added to the selection, if desired, for deselection of binders to highly homologous proteins or binders to fusion tags.

4. Use standard panning to isolate a more diverse repertoire of binders.
5. Use competitive panning to isolate higher affinity or more selective binders.
6. If a deselection target is used to pre-pan the phage, the negative selection in panning round 3 should be against the deselection target.
7. The binding capacity of Pierce NeutrAvidin-coated, high binding capacity, eight-well strips (ThermoFisher Scientific, catalog no. 15508) is ~60 pmol D-biotin per well; therefore, the biotinylated target is added in excess and can be reduced, if required.
8. If the method was used to isolate cross-reactive binders by alternating the target protein between panning rounds, the phage clones should be tested for binding against each of the different target proteins used.
9. If a deselection target was used during the selection process, the phage clones should be tested for binding against the deselection target as a negative control.
10. The binding capacity of Pierce Streptavidin-coated, clear, 96-well (ThermoFisher Scientific, catalog no. 15126) is ~10 pmol D-biotin per well. A concentration of 10 μg/ml of biotinylated target is usually sufficient to test in a phage ELISA, but can be optimized, if required.

Acknowledgments

We thank the Biomedical Health Research Centre, University of Leeds for funding the BioScreening Technology Group, where these methods were developed.

References

1. Bradbury A, Pluckthun A (2015) Reproducibility: standardize antibodies used in research. Nature 518(7537):27–29
2. Baker M (2015) Reproducibility crisis: blame it on the antibodies. Nature 521 (7552):274–276. https://doi.org/10.1038/521274a
3. McLeod J, Ferrigno PK (2016) Antibody alternatives. Scientist 30:38–45
4. Binz HK et al (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 332 (2):489–503
5. Koide A et al (1998) The fibronectin type III domain as a scaffold for novel binding proteins. J Mol Biol 284(4):1141–1151
6. Woodman R et al (2005) Design and validation of a neutral protein scaffold for the presentation of peptide aptamers. J Mol Biol 352 (5):1118–1133
7. Stadler LKJ et al (2011) Structure-function studies of an engineered scaffold protein derived from Stefin A. II: development and applications of the SQT variant. Protein Eng Des Sel 24:751–763
8. Tiede C et al (2014) Adhiron: a stable and versatile peptide display scaffold for molecular

recognition applications. Protein Eng Des Sel 27:145–155

9. Tiede C et al (2017) Affimer proteins are versatile and renewable affinity reagents. elife 27 (6):24903
10. Johnson A et al (2012) Sensitive affimer and antibody based impedimetric label-free assays for C-reactive protein. Anal Chem 84 (15):6553–6560
11. King R et al (2015) Inhibition of complement C3 and fibrinogen interaction: a potential novel therapeutic target to reduce cardiovascular disease in diabetes. Lancet 26 (385):60372–60375
12. Fisher MJ et al (2015) Trivalent Gd-DOTA reagents for modification of proteins. RSC Adv 5(116):96194–96200
13. Kyle HF et al (2015) Exploration of the HIF-1alpha/p300 interface using peptide and Adhiron phage display technologies. Mol BioSyst 11:2738–2749
14. Rawlings AE et al (2015) Phage display selected magnetite interacting Adhirons for shape controlled nanoparticle synthesis. Chem Sci 6:5586–5594
15. Stadler LK et al (2014) The use of a neutral peptide aptamer scaffold to anchor BH3 peptides constitutes a viable approach to studying their function. Cell Death Dis 30(5):564
16. Sharma R et al (2016) Label-free electrochemical impedance biosensor to detect human interleukin-8 in serum with sub-pg/ml sensitivity. Biosens Bioelectron 80:607–613
17. Weckman NE et al (2016) Comparison of the specificity and affinity of surface immobilised Affimer binders using the quartz crystal microbalance. Analyst 141(22):6278–6286
18. Arrata I et al (2017) Interfacing native and non-native peptides: using Affimers to recognise alpha-helix mimicking foldamers. Chem Commun 53(19):2834–2837
19. Koutsoumpeli E et al (2017) Antibody mimetics for the detection of small organic compounds using a quartz crystal microbalance. Anal Chem 89(5):3051–3058
20. Wang W et al (2017) Ultraefficient cap-exchange protocol to compact biofunctional quantum dots for sensitive ratiometric biosensing and cell imaging. ACS Appl Mater Interfaces 9(18):15232–15244
21. Xie C et al (2017) Development of an Affimer-antibody combined immunological diagnosis kit for glypican-3. Sci Rep 7(1):017–10083
22. Michel MA et al (2017) Ubiquitin linkage-specific affimers reveal insights into K6-linked ubiquitin signaling. Mol Cell 68(1):233–246
23. Gersch M et al (2017) Mechanism and regulation of the Lys6-selective deubiquitinase USP30. Nat Struct Mol Biol 24(11):920–930
24. Hughes DJ et al (2017) Generation of specific inhibitors of SUMO-1- and SUMO-2/3-mediated protein-protein interactions using Affimer (Adhiron) technology. Sci Signal 10 (505):eaaj2005
25. Robinson JI et al (2018) Affimer proteins inhibit immune complex binding to FcgammaRIIIa with high specificity through competitive and allosteric modes of action. Proc Natl Acad Sci U S A 115(1):E72–E81
26. Zhurauski P et al (2018) Sensitive and selective Affimer-functionalised interdigitated electrode-based capacitive biosensor for Her4 protein tumour biomarker detection. Biosens Bioelectron 108:1–8
27. Heidelberger JB et al (2018) Proteomic profiling of VCP substrates links VCP to K6-linked ubiquitylation and c-Myc function. EMBO Rep 19(4):21
28. Lopata A et al (2018) Affimer proteins for F-actin: novel affinity reagents that label F-actin in live and fixed cells. Sci Rep 8 (1):018–24953
29. Schlichthaerle T et al (2018) Site-specific labeling of Affimers for DNA-PAINT microscopy. Angew Chem Int Ed Engl 57 (34):11060–11063
30. Klont F et al (2018) Affimers as an alternative to antibodies in an affinity LC-MS assay for quantification of the soluble receptor of advanced glycation end-products (sRAGE) in human serum. J Proteome Res 17 (8):2892–2899
31. Hanes J, Pluckthun A (1997) In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A 94 (10):4937–4942
32. Wilson DS, Keefe AD, Szostak JW (2001) The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci U S A 98(7):3750–3755
33. Odegrip R et al (2004) CIS display: in vitro selection of peptides from libraries of protein-DNA complexes. Proc Natl Acad Sci U S A 101 (9):2806–2810
34. Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228 (4705):1315–1317
35. Tang AA et al (2017) Isolation of isoform-specific binding proteins (Affimers) by phage display using negative selection. Sci Signal 10 (505):eaan0868

Part II

Biophysical Characterization and Sample Optimization

Chapter 7

Measurements of Protein–DNA Complexes Interactions by Isothermal Titration Calorimetry (ITC) and Microscale Thermophoresis (MST)

Amandine Gontier, Paloma F. Varela, Clément Nemoz, Virginie Ropars, Magali Aumont-Nicaise, Michel Desmadril, and Jean-Baptiste Charbonnier

Abstract

Interactions between protein complexes and DNA are central regulators of the cell life. They control the activation and inactivation of a large set of nuclear processes including transcription, replication, recombination, repair, and chromosome structures. In the literature, protein–DNA interactions are characterized by highly complementary approaches including large-scale studies and analyses in cells. Biophysical approaches with purified materials help to evaluate if these interactions are direct or not. They provide quantitative information on the strength and specificity of the interactions between proteins or protein complexes and their DNA substrates. Isothermal titration calorimetry (ITC) and microscale thermophoresis (MST) are widely used and are complementary methods to characterize nucleo-protein complexes and quantitatively measure protein–DNA interactions. We present here protocols to analyze the interactions between a DNA repair complex, Ku70–Ku80 (Ku) (154 kDa), and DNA substrates. ITC is a label-free method performed with both partners in solution. It serves to determine the dissociation constant (K_d), the enthalpy (ΔH), and the stoichiometry N of an interaction. MST is used to measure the K_d between the protein or the DNA labeled with a fluorescent probe. We report the data obtained on Ku–DNA interactions with ITC and MST and discuss advantages and drawbacks of both the methods.

Key words Nucleo-protein complexes, Thermodynamic parameters, Microcalorimetry, Fluorescence, Thermophoresis, Double-Strand Break repair, NHEJ

1 Introduction

The molecular mechanisms in the nucleus are regulated by intricate networks of protein–protein and protein–DNA interactions. The word "interaction" itself is used in the literature for many purposes. It defines associations characterized by large-scale studies with

Amandine Gontier and Paloma Fernández Varela contributed equally with all other contributors.

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_7,

two-hybrid or tap-tag approaches and also in cell analyses with proximity ligation assays (PLA) or co-immunoprecipitation [1, 2]. "Interactions" include also genetic analyses as well as biophysical approaches and structural studies of complexes. This large set of approaches describe the same event of "interaction" and leads in some case to controversy. On the contrary, these multiple analyses favor in many cases an integrate view of the relevance of a given protein complex both in vitro and in the cell. For example, protein interaction measurements in the cellular context provide a characterization of the interactions in the presence of potential competitors or of post-translational modifications at different steps of the cell cycle. Biophysical approaches nicely complement these data by confirming or not the direct interaction with purified proteins and ligands. Biophysical measurements are central to characterize the strength, the specificity, and the stoichiometry of the interactions in a quantitative manner. Interactions measured in cells and not confirmed in vitro can also be informative of the role of an additional partner or the need for post-translational modifications eventually absent on purified proteins used for in vitro studies.

A large panel of biophysical approaches are used nowadays to perform quality controls and interaction analyses on macromolecular samples (some of these are presented in other chapters of this issue). Each methodology offers complementary information and comes with specific requirements on sample preparation. We present here protocols for Isothermal Titration Calorimetry (ITC) and MicroScale Thermophoresis (MST) analyses of the interactions between a protein complex and DNA substrates. The complex is formed by the heterodimer Ku70–Ku80 (or Ku). Ku is a core factor of the main DNA Double-Strand Break (DSB) repair pathway in human, called Non-Homologous End-Joining (NHEJ). Ku recognizes the DSB ends through its ring-shaped structure [3]. The heterodimer has a high affinity, in the nanomolar range, for double-strand DNA (15 bp minimum). The K_d of the Ku–DNA interaction was initially characterized by gel shift assays (EMSA) [4] and fluorescent anisotropy analyses [5]. Ku binds DNA in a sequence-independent manner and in vitro one Ku molecule can thread on DNA with one Ku bound every 15–20 bp of DNA duplex [3, 6]. Interestingly, a super resolution microscopy study showed that in cells only one or two Ku molecules bind to DSB ends, suggesting a limited threading of Ku from the DNA DSB ends in a cellular context [7]. Ku plays also a central role for the recruitment of the enzymatic activities (nucleases, polymerases, ligases) that process and ligate the DSB ends [8, 9]. Our laboratory recently described at the molecular level the mechanism of recruitment by Ku of some of the downstream NHEJ factors [10]. Here, we present the measurements by ITC of the thermodynamic parameters of the interaction between Ku and DNA substrates, and we

report the analyses by MST of the same interaction using DNA labeled with a fluorescent probe. We previously used ITC to characterize protein–protein and protein–DNA interactions involved in the NHEJ pathway [6, 11, 12] and in other DNA metabolism pathways [13–15].

Isothermal titration calorimetry measures the heat generated or consumed upon the interaction of two molecules, called hereafter the receptor and the ligand [16, 17]. The molecular interaction between the receptor and the ligand can be defined by the following equation: $\Delta G^{\circ} = -RT\ln K_a = RT\ln K_d = \Delta H - T\Delta S$, where ΔG° (kcal/M) is the standard Gibbs free energy, R (kcal/M) the gas constant, T (K) the temperature, K_a (M^{-1}) the equilibrium association constant, and K_d (M) the equilibrium dissociation constant. The second part indicates that the free energy is the sum of an enthalpy term ΔH (kcal/M) and an entropic term $-T\Delta S$ (kcal/M). Figure 1 represents a schema of the calorimeter. A constant power is applied to the reference cell. A very small temperature difference is maintained between the sample and reference cells thanks to a feedback circuit that applies a variable power to the sample cell. The injection of a small volume of ligand, loaded in the syringe, leads to a heat exchange when the ligand interacts with the receptor. Exothermic reactions generate heat and induce a decrease in the feedback power. Endothermic reactions absorb heat and lead to an increase of the feedback power. The instrument modifies the power applied to the measurement cell according to the nature of the interaction. If the reaction is exothermic, it reduces the power, and we observe a negative peak in the thermogram (Fig. 1b, top). On the contrary, if the reaction is endothermic, it increases the power, and we observe a positive peak in the thermogram (as in the Ku–DNA example presented here, Fig. 3b). The isotherm of titration (Fig. 1b, bottom) is determined by integrating each peak of the thermogram and by reporting the calculated values that correspond to the enthalpy of the reaction. The x axis of the isotherm is defined as the ratio of ligand/receptor molecules at each step of the titration. The curve that best fits these points gives, in the case of a single binding site, the ΔH, the K_d, and the stoichiometry of the interaction (Fig. 1b, bottom). ITC has several advantages: (1) it is widely used since limited optimization is needed; (2) the interaction is performed without labeling the partners; (3) both partners are in solution at opposite to methods where one partner is linked to a surface (like SPR, BLI, switchSENSE); (4) it is nondestructive, the complex receptor–ligand can be used for other application (cristallisation for example) or re-purified on a size exclusion chromatography or anion/cation exchange chromatography after measurement to separate the receptor and the ligand; (5) it has no molecular weight limitation. This method also has several limitations: (1) it consumes important quantities

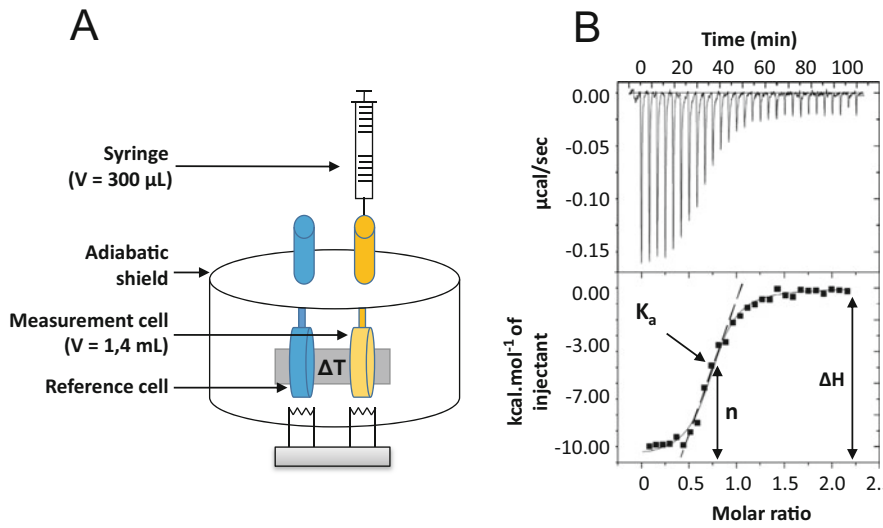


Fig. 1 (**a**) Schematic representation of an isothermal titration calorimeter (ITC). The reference and measurement cells are positioned in an adiabatic shield. A constant power is applied to the reference cell and a variable power is applied to the measurement cell to maintain a small constant temperature difference between the two cells thanks to a feedback circuit. Interactions are between a receptor deposited in the measurement cell and a small volume of ligand injected by the syringe. Most interactions generate heat (exothermic reaction). This triggers a decrease in the feedback power as represented in (**b**, top panel). Some interactions can consume heat leading to an increase in the feedback power (endothermic reaction). The integrals of the peaks from the top panel enable to calculate the enthalpy exchange at each injection. These values are reported versus the ratio of receptor and ligand. The fit of this curve is called the isotherm of titration. It gives access to the enthalpy ΔH, the association constant K_a (and thus the K_d, $K_d = 1/K_a$), and the stoichiometry N

of protein and DNA in particular for K_d in the µM range; (2) it is difficult to measure K_d below nanomolar range though under nanomolar range displacement titration can be done [18].

Microscale thermophoresis (MST) is a more recent approach for measuring the affinity constant (K_d) between a receptor and a ligand [19, 20]. This method can use either the intrinsic fluorescence of the protein (for example Monolith NT.LabelFree, Nanotemper) or the fluorescence signal (for example Monolith NT.115, Nanotemper) of a probe linked to the receptor or to the ligand. A temperature gradient is applied overtime by an IR (infra-red) laser on a capillary containing the fluorescent receptor alone or with increasing concentration of the ligand (Fig. 2a). The laser generates a local heat on the capillary inducing the diffusion of the fluorescent receptor away from the irradiated zone. The laser is turned off and the receptor diffuses again to the depleted zone. The diffusion rate of the receptor is highly sensitive to its hydration state. The formation of the receptor–ligand complex modifies locally the hydration, the size, and the charge of the receptor that can lead to detectable variation on the diffusion of the fluorescently labeled receptor in

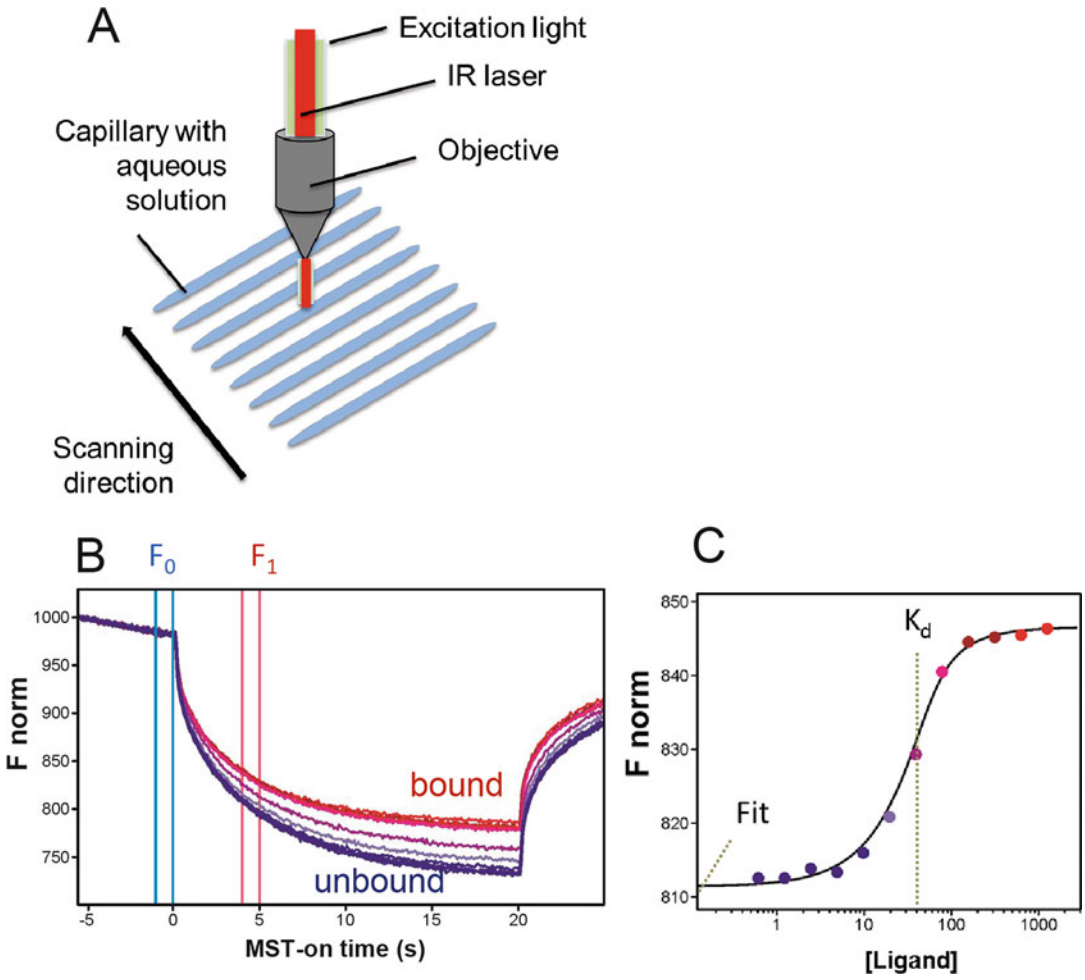


Fig. 2 (**a**) Microscale thermophoresis optic system. MST is measured in capillaries with a small sample volume (~4 μl). The fluorescence within the capillary is excited and detected through an objective. A focused IR laser is used to locally heat a defined sample volume. Thermophoresis of fluorescent molecules through the temperature gradient is detected by the objective. (**b**) Typical binding experiment. The thermophoretic movement of a fluorescent receptor (blue trace; unbound) changes upon binding to a non-fluorescent ligand (blue to red traces; bound), resulting in different traces. The change in thermophoresis is expressed as the change in the normalized fluorescence (F_{norm}), which is defined as $F1$(hot)/$F0$(cold) (F-values correspond to average fluorescence values between defined areas marked by the red and blue cursors, respectively). (**c**) Titration with a non-fluorescent ligand results in a gradual change in thermophoresis, which is plotted as F_{norm} versus ligand concentration. When relevant, the resulting binding curve can be fitted with a single binding site model and enables to define the K_d

the capillary. The fluorescence curves (called MST traces) are measured in the presence of varying ligand concentrations, ranging from 1/10 to 20-fold the expected K_d (Fig. 2b). The dissociation constant (K_d) can be estimated by fitting with a single-site model, when relevant, the normalized fluorescence against the ligand concentration (Fig. 2c). MST presents several advantages: (1) it

requires very small amounts of material—typically in the 100 ng to 1 μg range; (2) there is almost no limitation to molecular size or mass; (3) the method is easy and fast to perform; and (4) it is an immobilization free method. MST has also several limitations: (1) one of the partners should be labeled with possible steric hindrance with the interaction sites, and hydrophobic fluorescent probes may cause some nonspecific binding; and (2) the theory behind the measurement principle is less obvious than for other biophysical approaches.

2 Materials

2.1 Equipment for Isothermal Titration Calorimetry (ITC)

1. MicroCal VP-ITC: The VP-ITC microcalorimeter (Malvern) is composed of two compartments or cells: one reference cell is filled with water or buffer and one measurement cell (volume of 1.4 ml) is filled with the first partner called the receptor. Both are maintained at equal temperature and are positioned in an adiabatic environment. The ligand is introduced with small injections (5–10 μl) in the measurement cell. If the ligand interacts with the receptor, in most cases, some heat is generated or absorbed due to the interaction. This is detected by the device that compensates by respectively reducing or increasing the electric current in the measurement cell (Fig. 1). In some rare cases, at a given temperature, an interaction can happen without heat exchange ($\Delta H = 0$). In that case, changing the temperature (typically ± 10 °C) should unmask the interaction, if it exists, since the enthalpy varies with the temperature.
2. The ITC-200 microcalorimeter which is equipped with a measurement cell of 200 μl and a syringe of 40 μl requires smaller quantities of sample. In the case of the Ku–DNA interaction, presented here, the VP-ITC calorimeter provided better thermograms and a better signal/noise ratio.
3. ThermoVac™ (Malvern): This device enables to degas the buffers and samples before use and to pre-equilibrate at a temperature close to the measurement temperature. It is also used as pump for cleaning procedures.
4. VPViewer™ and Origin™ software packages: VPViewer is used to command the VP-ITC and to follow the experiment. Origin is used to analyze the raw data called thermograms.

2.2 Samples for ITC

As will be detailed below, suitable concentrations for a titration depend on the expected K_d between the ligand and the receptor and on the enthalpy associated with the formation of the complex. Quantities listed below are required to characterize the interaction between the Ku70–Ku80 (Ku) heterodimer and small DNA

duplexes (18 bp and 42 bp). The K_d expected from the literature is in the nM range [4].

1. The Ku heterodimer is produced and purified as described in [10]. The complex Ku is expressed in insect cells and purified using Ni-NTA affinity purification (Thermo Fisher Scientific) followed by an anion exchange chromatography using a Resource-Q column (GE Healthcare Life Science) [10]. Insect cell pellets from 1.2 l culture classically yield 80 mg purified Ku at the end of the protocol, which is sufficient for about 200 titrations Ku vs DNA.
2. 18 nt or 42 nt DNA oligonucleotides and their complementary strands are synthetized at the 1 μmole scale (Sigma Aldrich, HPLC 95% pure). The sequence of the DNA 18 nt is 5′GTCTCACGCTCGGATAAC3′; the sequence of the DNA 42 nt is 5′GAACGAAAACATCGGGTACGAGGACGAAG ACTGACCACGACA3′. With a command at 1 μmole scale, we obtained between 115 and 530 nmoles of oligonucleotides that is sufficient for at least 35 titrations after hybridization.
3. A dialysis membrane with a molecular weight cut-off (MWCO) of 30 kDa (Spectra/Por® molecular porous membrane tubing); 2 l of ITC buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol) prepared with deionized water (18 mΩ-cm^{-1} at 25 °C); Amicon® Ultra-4 30 kDa ultrafiltration devices.

2.3 Equipment for Microscale Thermophoresis (MST)

1. Monolith NT.115: The Monolith NT.115 Blue/Green is equipped with two LEDs (LED1: excitation 470 nm, emission 520 nm; LED2: excitation 550 nm, emission 600 nm, Nanotemper Technologies). The instrument is equipped with a sample tray that can process automatically up to 16 samples per run in 10 min. MST is measured in capillaries with a total volume of 3–5 μl. Control software guides step-by-step planning and assay setup. Binding values are obtained instantly, and results are exportable. Affinity Analysis software enables advanced data analysis like merging, grouping, or comparing data sets.
2. Capillaries: All capillaries are made of glass with infrared laser quality. Standard Treated Capillaries (Nanotemper) are physically treated to ensure high-quality surface properties. Hydrophobic (Nanotemper) and Premium-Coated (Nanotemper) Capillaries are covalently coated with a dense brush of specially designed polymers to avoid molecule adsorption.

2.4 Samples for MST

MST is based on the detection of temperature-induced changes in fluorescence of a target as a function of the concentration of a non-fluorescent ligand. In typical experiments, 5–500 nM of the fluorescently labeled molecule is used. The maximum ligand

concentration should be 10–20 times that of the K_d. Quantities listed below are required to characterize the Ku–DNA complex (the expected K_d is in the nM range).

For experiments with labeled DNA and non-labeled Ku, we use:

1. Ku heterodimer purified as described in Subheading 2.2. A total volume of 20 µl at the higher concentration is needed per run.
2. A dialysis membrane with a molecular weight cut-off (MWCO) of 30 kDa (Spectra/Por® molecular porous membrane tubing); 2 l of MST buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20).
3. Desalted 18 bp or 42 bp oligonucleotides labeled with a 5′-FAM and their complementary strands synthetized at the 1 µmole scale (*see* sequences above with a FAM fluorophore added in 5′) (Eurofins). One synthesis yielded about 100 nmoles of oligonucleotides, sufficient for 50 titrations after hybridization.

For experiments with non-labeled DNA and labeled Ku, you will need:

4. Ku heterodimer purified as described in Subheading 2.2. The buffer of Ku is changed for a buffer that does not contain primary amines, using the labeling buffer of the Monolith Protein Labeling kit BLUE-NHS (Amine Reactive) from Nanotemper (reference MO-L003), for which composition is not provided, but the optimal pH is between 8 and 8.4. A volume of 100 µl at 20 µM is required.
5. MST buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20).
6. A Nanotemper labeling kit containing fluorescence dyes that are suited for MST measurements. The protocols are optimized to ensure a good labeling, an efficient removal of free dye and about a 1:1 ratio of labeled protein to dye. Two coupling chemistries are available, one for primary amines (NHS, e.g., lysine labeling) and one for thiol groups (maleimide for cysteine labeling). Second-generation dyes are now available (Tris-NTA) but only for the instrument with red LED. Proteins with a molecular weight greater than 5 kDa can be labeled.
7. Desalted 18 bp or 42 bp non-labeled oligonucleotides and their complementary strands are synthetized at the 1 µmole scale. The DNA sequences are the same as indicated above.

3 Methods

To get an accurate K_d measurement, the receptor and the ligand concentrations should be known accurately. This can be determined with the UV spectrum of both protein and DNA samples and their extinction coefficient. The biophysical measurement of the K_d by ITC or MST (as with many other biophysical approaches) requires highly purified samples. This implies that quality control (QC) analyses are performed before measurements and that the QC indicates high purity (greater than 95%). Protein purity can be determined using several methods including UV spectrum, SDS PAGE gels, and analytical chromatography (size exclusion chromatography or anion exchange) [21]. DNA purity can be analyzed by using UV spectrum, acrylamide gels, or analytical anion exchange chromatography. The receptor is commonly used at a fixed concentration, and the ligand is used at various concentrations ranging from 1/10 to 20-fold the expected K_d. The way to define the concentrations used for the receptor in ITC and MST is described below according to the specificities of these two methods.

3.1 Methods for ITC

To get an accurate K_d measurement with ITC, the receptor concentration in the cell of the instrument and the ligand loaded in the syringe should be well chosen. The receptor concentration appropriate for an ITC experiment can be estimated using the dimensionless constant c, determined by the formula, $c = N \times K_a \times [\mathrm{Ro}]$, where N is the stoichiometry of the reaction (number of ligand molecules bound by molecule of receptor), K_a is the association constant (M^{-1}), and [Ro] the initial concentration of the receptor in the sample cell (M). To obtain a well-defined S-shaped binding isotherm with two plateau (Fig. 1b, bottom) and an accurate measurement of the thermodynamic parameters of the equilibrium, it is recommended to use an initial concentration of the receptor [Ro] that results in a c value between 20 and 200 [22]. In the absence of a priori information, typical starting concentrations are 1 μM for the receptor and 10 μM for the ligand. The ligand is usually loaded in the syringe at a concentration ten times higher than the receptor, so that at the end of the titration (typically 28 injections of 10 μl), the molar ratio of the ligand over the receptor is 2.

3.1.1 Experimental Design

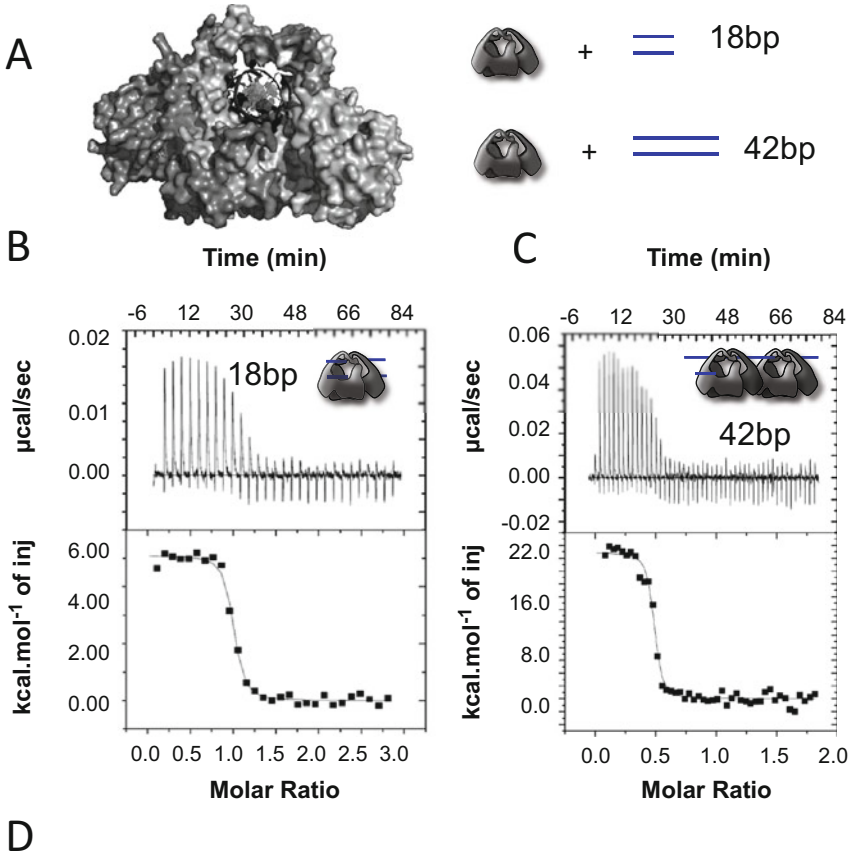
In our example Ku–DNA, with an expected K_d of about 5 nM and a stoichiometry N of 1, we started with a concentration of Ku in the cell ([Ro]) of 1 μM corresponding to a c value of 200. The ligand and the receptor are dialyzed against the same buffer to minimize heat exchanges due to buffer differences.

3.1.2 Sample Preparation for ITC

1. The Ku heterodimer (receptor) is dialyzed three times against ITC buffer for at least 2 h against a volume of buffer corresponding to at least 30-fold the volume of the receptor sample. Keep 50 ml of the last dialysis bath as reference buffer. It will be used to rinse the measurement cells and the syringe and to adjust the concentration of the receptor and ligands, if necessary.
2. Preparation of the receptor for the measurement cell: Concentrate the dialyzed protein using an Ultra-4 (Amicon), measure UV absorption at 280 nm (using for example a Nanodrop; Thermo Fisher Scientific), and adjust the concentration to 1 μM (0.15 mg/ml). A minimum volume of 2.3 ml (1.4 ml for the measurement cell plus 0.9 ml dead volume) is prepared.
3. Hybridization of DNA substrates (ligand). The oligonucleotides are classically dissolved at 500 μM in water and equal amounts of the complementary strands are mixed at a temperature of 95 °C for 5 min and cooled down overnight at room temperature. The concentration is estimated using a Nanodrop (Thermo Fisher Scientific) and UV absorption at 260 nm.
4. Verify hybridization: Inject 6 nmoles (10 μl at 0,6 mM) on a Mini-Q-1 ml (Sigma-Aldrich) column to check if the DNA is correctly hybridized and the proportion of the remaining single-strand DNA. Perform a gradient on the Mini-Q of 20 ml from 100% buffer A (Tris 20 mM, pH 8) to 100% buffer B (Tris 20 mM, pH 8, NaCl 1 M).
5. The DNA ligand is diluted at 10 μM using the last dialysis buffer to limit heat exchange due to buffer differences of the receptor and the ligand. These dilution heats may mask the desired signal. Prepare 0.6 ml of ligand to be positioned in the serynge.

3.1.3 ITC Measurements and Data Interpretation

1. The receptor (2.3 ml of Ku at 1 μM) and the ligand (0.6 ml DNA at 10 μM) are degassed and equilibrated for 10 min at the measurement temperature (here 20 °C) using the ThermoVac™.
2. The receptor is deposited in the measurement cell with a 2.5 ml glass syringe. The sample is slowly loaded in the cell to avoid bubbles. Injection is stopped when the liquid comes above a visible level.
3. The ligand is loaded on the injection syringe (in open position) using a 10-ml plastic syringe connected to the top of the ITC syringe. The syringe is then purged and filled twice to remove bubbles.
4. Start titration.



DNA	N	K_D (nM)	ΔH° (kcal.mol^{-1})	-TΔS° (kcal.mol^{-1})	ΔG° (kcal.mol^{-1})
18bp	0.97 ± 0.1	3.5 ± 0.8	5.1 ± 0.1	-16.6 ± 0.2	-11.5 ± 0.2
42bp	0.47 ± 0.1	2.4 ± 0.5	22.9 ± 0.3	-34.7 ± 0.5	-11.8 ± 0.1

Fig. 3 ITC measurements between Ku and DNA. (**a**) Crystal structure of Ku-DNA complex with a 14 bp DNA. Ku has a ring shape structure. The DNA present in the crystal structure has an additional hairpin part that has been removed for clarity (pdb code 1JEY). A schema is presented of the two interactions analyzed with Ku and either a 18 bp DNA or 42 bp DNA. (**b**) ITC measurement between Ku and a 18 bp DNA. The interaction is endothermic with positive peaks on the thermogram. The heat exchanges observed in this interaction are very weak (<0.02 µcal/s) and underline the high sensitivity of the VP-ITC. The isotherm of titration enables to define ΔH, K_d, and N of the reaction, reported in Table (**d**). One Ku molecule binds to an 18 bp DNA. (**c**) ITC measurement between Ku and a 42 bp DNA hairpin. ΔH, K_d, and N of the reaction are reported in Table (**d**). Two Ku molecules bind on the 42 bp DNA leading to a stoichiometry N close to 0.5

We analyzed the interactions of Ku with a 18 bp and a 42 bp DNA duplex (Fig. 3b, c) and observed two endothermic reactions (positive peaks) in both the cases. The interaction between Ku and the 18 bp DNA shows a stoichiometry of 0.97 ± 0.1 molecules of 18 bp DNA bound to one Ku, a K_d of 3.5 ± 0.8 nM, and an

entropy-driven reaction with a positive unfavorable enthalpy. The interaction with the 42 bp DNA gives a stoichiometry of 0.47 ± 0.1 molecule of 42 bp DNA bound to one Ku in good agreement with the expected positioning of two Ku molecules on such DNA. The interaction between and the 42 bp DNA shows a K_d of 2.4 ± 0.5 nM and a reaction also driven by the entropy. The heat exchanges measured for the two interactions are very small (maximum value of heat exchange ~0.02 μcal/s for Ku–DNA (18 bp)) and highlight the high sensitivity of the VP-ITC.

3.2 Methods for MST

We describe here general guidelines to measure interactions using MST and we illustrate two experimental setups with a labeled DNA and non-labeled Ku or a labeled Ku and non-labeled DNA. We will call receptor the fluorescently labeled molecules and ligand the non-labeled one hereafter. The first step of an MST experiment is to find a suitable combination of the receptor concentration (10–100 nM are usually sufficient) and excitation LED power. The LED can be adjusted to 1–100%. For the binding experiment itself, you will need ~200 μl at twice the concentration you have established at this step (for one replicate measurement). To get an accurate, well-defined K_d, it is important to use a high enough concentration of the ligand to reach saturation. Saturation is a hallmark of ligand-specific molecular interaction: if the interaction cannot be saturated, i.e., the signal does not change anymore by increasing the ligand concentration, it is not a specific interaction, but a nonspecific adsorption effect. To reach saturation, it is recommended that the highest concentration of ligand is equal or higher than 20-fold the expected K_d value.

In a preliminary assay, we compare the fluorescence intensity of the receptor after switching on the IR laser LED (MST power) in the absence and in the presence of a high enough concentration of the ligand (respective to the expected K_d). We perform this experiment with at least four capillaries with and without ligand. The signal is defined as the response amplitude F_{norm}. It is the ratio between fluorescence intensity after turning on the IR laser LED (usually measured at the 2-second or 5-second mark, called "MST on time") divided by the fluorescence intensity measured before turning on the IR laser LED. It is expressed as ‰. The noise floor is the error (standard deviation) between replicates without ligand. If it is above 8‰, it may reflect some protein aggregation that needs optimization. This also determines what should be considered binding (or not) of the ligand: binding is only confirmed, if the change observed upon ligand binding is threefold larger than your noise level established at this step. For example, for a typical noise

level of 2‰ at "medium" MST power, the signal change upon ligand binding should be at least 6‰ (e.g., a change from an F_{norm} value of 857‰ to 863‰). Both the noise and the amplitude are displayed when reviewing the dose–response panel. In order to measure the K_d, we incubate our (fluorescently labeled) receptor, at a concentration below the K_d, with increasing concentrations of non-labeled ligand. An easy way to do this is to perform a dilution series, in the form of 1:1 dilutions. In the Monolith instruments, we can use 16 concentrations; this will cover a concentration range of 3×10^4 (2^{15}).

3.2.1 Sample Preparation

Non-labeled Ku and DNA are prepared as described in Subheading 3.1.2. First, always spin down the stocks for 5 min at 15,000 × *g* to remove large aggregates, which is the main source of noise. Always prepare volumes of at least 20 μl of sample in the smallest micro-reaction tubes to avoid surface absorption.

3.2.2 Fluorescence Pretests

Before starting binding assays, several pretests are required to optimize. Check the optimal concentration of the fluorescence molecules. Ideal fluorescent intensities are between 200 and 1500 counts.

Check the capillary type that is optimal for our study (standards, hydrophobic, and premium). The aim is to obtain narrow, shoulder-free, regular, and symmetrical fluorescent peaks.

Check the optimal buffer composition. BSA and detergents can be used to improve sample homogeneity. pH and/or ionic strength are also parameters that can be adjusted if needed. Reducing agents freshly prepared can also be used when needed.

3.2.3 Preliminary Binding Assays

In these preliminary assays, we compare the fluorescent signals of the receptor in the absence and in the presence of the higher concentration used for the ligand. The DNA solution (10 μl of [DNA-FAM5] at 10 nM) is mixed with 10 μl buffer or with 10 μl of Ku at 266 nM concentration. We analyzed fluorescence intensity and homogeneity, and the signal to noise ratio by repeating the measurement four times. MST traces should be smooth, which will indicate the absence of aggregation. The signal is defined as the response amplitude (difference of fluorescence amplitude between receptor alone and receptor–ligand complex) and the noise is the standard deviation of errors between replicates. A signal-to-noise ratio less than 5 suggests a signal too weak to measure a K_d, a ratio of 5 or more indicates a quite favorable case for a K_d measurement and a ratio higher than 12 reflects optimal conditions to measure a K_d with MST.

3.2.4 K_d Determination

Prepare 16 small micro-tubes and transfer 20 µl of the highest concentration of the ligand (non-labeled molecule) to the first one. The Concentration Finder software tool can be used to simulate the interaction with different concentrations of ligand and receptor. Fill 10 µl of the optimal assay buffers into the tubes 2–16. Use the same pipette tip and transfer 10 µl of tube 1 to tube 2 and mix well by pipetting up and down several times. Do the same with tube 3 to tube 16 to obtain a serial dilution. Remove 10 µl from tube 16 after mixing. Add 10 µl of receptor (labeled molecule) to each tube and mix well by pipetting up and down several times. Here change tip for every well to ensure the addition of the same quantity in each tube and no contamination of the labeled molecule that may be due to the pipetting. Incubate if needed before loading into the capillaries. Most interactions reach equilibrium after a few minutes, so additional incubation is often not needed. When larger conformational changes are required from both partners, slow kinetics can be observed, and longer incubations might be necessary. Fill 16 capillaries of the correct type and place them in the instrument. Start the measurements.

3.2.5 Assays with Labeled Ku (and Non-labeled DNA)

Ku is labeled with NT495 final concentration 20 nM. Titrant DNA 18 bp is prepared at concentrations ranging from 123 nM to 3.9 pM by serial dilution. The buffer used is MST buffer and the capillaries are premium capillaries. Measurements are performed at 22 °C, at 40% LED excitation and 40% or 60% MST power. The signal to noise ratio is 3 in these conditions and does not enable to measure a reliable K_d (Fig. 4a). The excitation power of 40% was enough to have around 350 fluorescence counts. Increasing the MST power increases the signal, but also the noise of the data.

3.2.6 Assays with Labeled DNA (and Non-labeled Ku)

DNA is purchased with a FAM fluorophore in 5′ of one 18 nt single-strand DNA. The fluorescent probe is separated to the 18 nt by three nucleotides to avoid interference with the binding site (Sigma Aldrich, HPLC 95% pure). The stock DNA is prepared at 230 µM with the same hybridization protocol as for ITC (*see* above). The final concentration of DNA in MST assays is 10 nM. Ku is prepared by serial dilution from 1330 nM to 80 pM. The buffer used is the MST buffer. Standard capillaries are used as well as low binding tubes and tips. Measurements are performed at 60% (data not shown), 80%, and 100% LED excitation, 40% MST power, and temperature of 22 °C (Fig. 4b, d). The signal to noise ratio at 60%, 80%, and 100% are 5.8, 8.3, and 17.7, respectively. The K_d is poorly defined at 60% excitation power with a K_d estimated of 100 ± 500 nM (data not shown). We measured a more precise K_d of 16.0 ± 7.7 nM at 80% (standard error of regression of 1.04 and a reduced χ^2 of 0.95) (Fig. 4b). At 100% excitation power, we measured the K_d with the smallest standard error, K_d of 10.8 ± 2.6 nM (standard error of regression is 0.37 and reduced χ^2 is 0.19) (Fig. 4c, d).

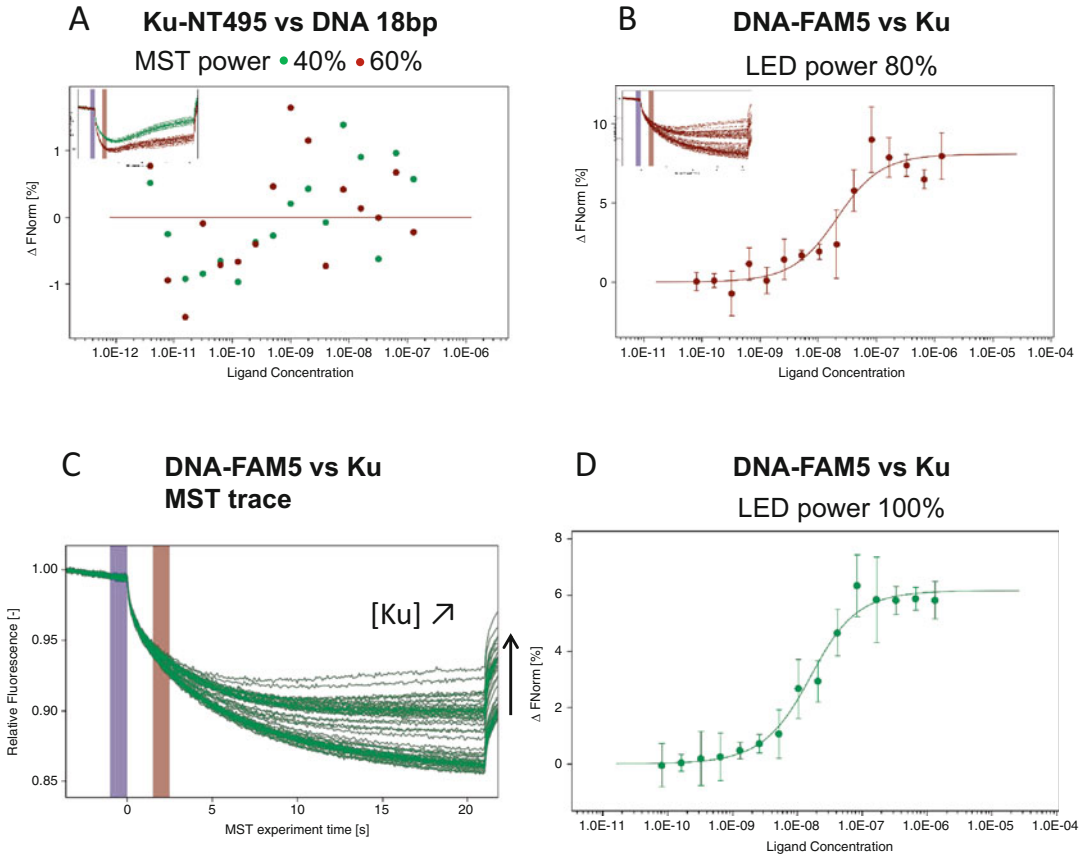


Fig. 4 MST measurements between Ku70–Ku80 and an 18 bp DNA. (**a**) Ku is labeled with NT-495 fluorescent probe. The signal to noise ratio is too weak, and no K_d can be measured at 40% or 60% MST power. (**b**) The 18 bp DNA is labeled in 5′ with a FAM fluorophore on one strand of the duplex. The measurements are performed at 60%, 80%, and 100% excitation power. The signal to noise ratio and the fits are better defined at 100% excitation (**d**) than at 80% excitation. The MST traces observed at 100% excitation are shown in (**c**)

4 Additional Remarks

1. Before each ITC experiment, make sure that the measurement cell and the syringe are clean. Clean the sample cell and the syringe with 14% v/v decon and rinse them with, respectively, 1 l and 250 ml of water. A control ITC measurement is then performed with 20 injections of 10 μl of water from the syringe into the cell filled with water. The signal should be small (DP < 0.01 μcal/s) and regular. The next user can then safely use the device for his measurements. When needed, stronger cleaning procedures could be used (*see* recommendations of the manufacturer (Malvern)). The Hastelloy alloy that covers the

ITC cells is compatible with most buffers except strong acids. The ITC instrument should be preferentially in a room with a control temperature (about 20 °C).

2. The buffer composition can influence the ITC signals. Reducing agents like DTT cause erratic baseline variations and TCEP or β-mercaptoethanol are preferred to maintain reducing conditions, when needed. The ionization properties of the buffer can have an influence on the values measured by ITC in particular when the interface involves charge exchange. In that case, it is recommended to compare the values obtained with different buffers and at different ionic strengths (e.g., [23]).
3. If no heat exchange is observed during an interaction, it can happen that the reaction exists but has a null enthalpy at the temperature chosen. To evaluate this possibility, the ITC measurements can be performed at different temperatures (e.g., 10 °C lower or higher).
4. The ITC gives access to the stoichiometry (N) of the interaction. In many reactions, a value of 1 for N is expected, one molecule of ligand bound by molecule of receptor. Even, in this case, some deviation can be observed for several reasons: (1) the presence of inactive receptors (unfolded, aggregated, etc.) in the sample cell will lead to a lower N value, proportional to the fraction of inactive receptor. The presence of inactive ligands in the syringe will lead to a higher N value, proportional to the fraction of inactive ligand. Similarly, errors on receptor or ligand concentrations will have a direct impact on the N value.
5. In parallel to the measurement of the receptor–ligand interaction, it is important to perform the ITC reaction with the ligand in the syringe and buffer in the cell. This control will evaluate if the signals observed are really due to the interaction and not to other event like the dissociation of the ligand when it is introduced in the sample cell. This can happen in particular when the ligand is a protein that can form oligomers and dissociate when injected in the cell.
6. Kinetic information can be recovered from ITC measurement using a method called kinITC that can be applied for some interactions presenting asymmetric peaks in the thermogram [24].
7. The range of K_d accessible by ITC is classically between 0.1 nM and 10 μM. For lower K_d, the slope of the isotherm will be sharp, preventing a precise measurement of the K_d. However, the ITC can be used in these conditions to determine ΔH and N of the interaction. At higher K_d, the measurement will often need a large amount of receptors and ligands to have a correct fit of the K_d.

8. For MST experiments, it is recommended to use low binding tubes and tips. For the serial dilution, better results are obtained without changing the tips from tube to tube. However, when adding the labeling molecule, it is recommended to change the tip at each tube. Avoid making bubbles by pipetting carefully and never vortex these small volumes to avoid denaturation of samples. Handle capillaries with care and never touch at the center where the measurement is performed.
9. An SDS denaturation test (SD-test) has to be performed to distinguish between fluorescence changes caused by an interaction and those caused by nonspecific effects, e.g., loss of protein due to aggregation or adsorption to labware. This test consists in the denaturation of all proteins contained in the sample using a mix of SDS and DTT and heating to 95 °C. In this context, the receptor–ligand interaction is disrupted. After denaturation, the change of fluorescence should not be observed in the absence or in the presence of increasing concentration of ligands. If addition of ligands induces fluorescence change with denatured protein, this suggests some interference of the ligand and some new experimental conditions should be researched.
10. If not enough fluorescence is observed, one can increase the concentration of the labeled receptor or increase the LED power. A different strategy of labeling can be used. In cases of absorption to the capillary, adding detergents or other amphiphilic polymer additives will help.
11. Under certain conditions, MST can also be used to obtain the stoichiometry of the interaction. For that, we need to know precisely the concentration of receptor and ligand as well as the K_d of their interaction. In this case, we need a much higher concentration of the receptor at least 20-fold than the K_d.
12. Thermodynamic can be extracted thanks to the Van't Hoff equation: $\ln K_D = \frac{\Delta H}{-RT} + \frac{\Delta S}{R}$. When plotting $\ln(K_D) = f(1/T)$, one can obtain a straight line which can be fitted as a linear equation where the slope is given by $\Delta H/-R$ and the intercept is $\Delta S/R$.
13. It can be worth to label the other molecule and repeat the experiment in the reverse way to confirm the result.
14. A positive control and a negative control are advisable to confirm that there is no unspecific interaction or that the conditions of the assay are optimal.

Acknowledgments

J.B.C. is supported by ARC program (SLS220120605310), ANR (ANR-12-SVSE8-012), INCA DomRep (PLBIO 2012-280), and CEFIPRA grant 5203C and by the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INBS-05). A.G. is supported by a CIFRE PhD fellowship with Sanofi. We thank Pierre Soule from Nanotemper Technologies for his availability and all the fruitful discussion. The experiments were performed on the Platform PIM (Platform for measurements of Interactions of Macromolecules) (https://www.i2bc.paris-saclay.fr/spip.php?article280).

References

1. Gavin AC, Bosche M, Krause R et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415(6868):141–147
2. Uetz P, Giot L, Cagney G et al (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403 (6770):623–627
3. Walker JR, Corpina RA, Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature 412(6847):607–614
4. Blier PR, Griffith AJ, Craft J, Hardin JA (1993) Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. J Biol Chem 268(10):7594–7601
5. Arosio D, Costantini S, Kong Y, Vindigni A (2004) Fluorescence anisotropy studies on the Ku-DNA interaction: anion and cation effects. J Biol Chem 279(41):42826–42835
6. Tadi SK, Tellier-Lebegue C, Nemoz C et al (2016) PAXX is an accessory c-NHEJ factor that associates with Ku70 and has overlapping functions with XLF. Cell Rep 17(2):541–555
7. Britton S, Coates J, Jackson SP (2013) A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. J Cell Biol 202 (3):579–595
8. Chang HHY, Pannunzio NR, Adachi N, Lieber MR (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol 18 (8):495–506
9. Frit P, Ropars V, Modesti M, Charbonnier JB, Calsou P (2019) Plugged into the Ku-DNA hub: the NHEJ network. Prog Biophys Mol Biol 147:62–76
10. Nemoz C, Ropars V, Frit P et al (2018) XLF and APLF bind Ku80 at two remote sites to ensure DNA repair by non-homologous end joining. Nat Struct Mol Biol 25(10):971–980
11. Ropars V, Drevet P, Legrand P et al (2011) Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. Proc Natl Acad Sci U S A 108 (31):12663–12668
12. Malivert L, Ropars V, Nunez M et al (2010) Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. J Biol Chem 285(34):26475–26483
13. Bacquin A, Pouvelle C, Siaud N et al (2013) The helicase FBH1 is tightly regulated by PCNA via CRL4(Cdt2)-mediated proteolysis in human cells. Nucleic Acids Res 41 (13):6501–6513
14. Dherin C, Gueneau E, Francin M et al (2009) Characterization of a highly conserved binding site of Mlh1 required for exonuclease I-dependent mismatch repair. Mol Cell Biol 29(3):907–918
15. Liberti SE, Andersen SD, Wang J et al (2011) Bi-directional routing of DNA mismatch repair protein human exonuclease 1 to replication foci and DNA double strand breaks. DNA Repair (Amst) 10(1):73–86
16. Holdgate GA (2001) Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. BioTechniques 31 (1):164–170
17. Krell T (2008) Microcalorimetry: a response to challenges in modern biotechnology. Microb Biotechnol 1(2):126–136
18. Velazquez-Campoy A, Freire E (2006) Isothermal titration calorimetry to determine

association constants for high-affinity ligands. Nat Protoc 1(1):186–191

19. Asmari M, Ratih R, Alhazmi HA, El Deeb S (2018) Thermophoresis for characterizing biomolecular interaction. Methods 146:107–119
20. Jerabek-Willemsen M, Wienken CJ, Braun D, Baaske P, Duhr S (2011) Molecular interaction studies using microscale thermophoresis. Assay Drug Dev Technol 9(4):342–353
21. Raynal B, Lenormand P, Baron B, Hoos S, England P (2014) Quality assessment and optimization of purified protein samples: why and how? Microb Cell Factories 13:180
22. Myszka DG, Abdiche YN, Arisaka F et al (2003) The ABRF-MIRG'02 study: assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction. J Biomol Tech 14(4):247–269
23. Czarny B, Stura EA, Devel L et al (2013) Molecular determinants of a selective matrix metalloprotease-12 inhibitor: insights from crystallography and thermodynamic studies. J Med Chem 56(3):1149–1159
24. Dumas P, Ennifar E, Da Veiga C et al (2016) Extending ITC to kinetics with kinITC. Methods Enzymol 567:157–180

Chapter 8

switchSENSE Technology for Analysis of DNA Polymerase Kinetics

Guillaume Bec and Eric Ennifar

Abstract

The switchSENSE technology is a recent approach based on surface sensor chips for the analysis of interactions of macromolecules. The technology relies on electro-switchable DNA nanolevers tethered at one end on a gold surface via a sulfur linker and labeled with a Cy3 dye on the other end. The switchSENSE approach is effective in the determination of a large panel of biophysical parameters such as binding kinetics, dissociation constant, hydrodynamic radius, or melting temperature. In addition, it can also give access to some enzymatic data such as nuclease or polymerase activity. Here we describe a DNA polymerase assay that allows retrieving, in a single experimental set, association and dissociation rates, as well as the catalytic rate of the enzyme.

Key words HIV reverse transcriptase, switchSENSE technology, Kinetics, Biosensor, DNA polymerase

1 Introduction

Immobilization of molecules on solid supports is a fairly well-known method that allows for very sensitive measurement of molecular interaction, as for the very popular surface plasmon resonance (SPR) approach [1]. Usually, biophysical properties such as molecular size or hydrodynamic radius are inferred from the molecular mobility under the influence of external forces. switchSENSE technology is a combination of these two approaches with the measurement of analyte adsorption on a layer of actuated surface-bound fluorescent probe [2, 3]. Two main measurement modes are accessible: (1) a static mode, where analyte binding is measured thanks to fluorescence intensity variation, and (2) a dynamic mode, where binding is detected through the change of the oscillation rate of the probe actuated by an AC electric field. These two alternative measurement modes give access to a wide range of biophysical parameters as k_{ON}, k_{OFF}, K_D, hydrodynamic diameter, or conformational changes (*see* for instance refs. 4–11).

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_8,

Master piece of the switchSENSE technology, the probe is made by a single-stranded DNA bearing a fluorescent dye at one extremity and attached on its opposite end to a gold-quenching surface through a sulfur bond. This single-stranded DNA is annealed to a complementary strand (RNA or DNA) that could be used alone or conjugated coupled to a protein. Hybridization of the complementary strand generates a rigid negatively charged electro-switchable biosensor, also referred as nanolever (Fig. 1). Making use of the DNA plasticity and of biochemical tools, a wide variety of nanolevers could be adapted to the biological context of the experiment: size and type of DNA may vary, even allowing the use of complex DNA origami constructs [12], RNA motifs can be easily adapted [13], various types of modified extremity can be used (protein, cofactor, NTA, biotin, etc.) [2], as well as various dyes (Cy3, Cy5, etc.).

Here we describe a specific setup adapted from Langer and colleagues [14] for polymerase activity characterization, using an overhang DNA as nanolever. With this kind of substrate, we show that it is possible to follow the entire biological reaction including DNA polymerase association, polymerization, and dissociation rate in one set of experiment.

2 Materials

2.1 Instrument and Accessories

1. DRX 2400 instrument (Dynamic Biosensors).
2. Nanodrop ND-1000 spectrophotometer (Thermo Fisher) or equivalent.
3. switchCONTROL, switchBUILD, and switchANALYSER software (Dynamic Biosensors).
4. Biochip for enzymatic studies (ENZ-80-1-Y1-S, Dynamic Biosensors) bearing an 80-mer DNA probe attached at the 3′ extremity.
5. Passivation and regeneration solutions (Dynamic Biosensors, SOL-PAS-1-5 and SOL-REG-12-1).
6. The P36T80 oligonucleotide, a 36-mer DNA complementary to 3′ end region of the 80-mer DNA probe attached to the biochip (*see* **Note 1**).
7. Heparin sodium salt, physico-chemical grade (Sigma-Aldrich).
8. 10× Auxilary buffer (100 mM sodium phosphate, pH 7.0, 400 mM NaCl, 0.5% Tween 20, 0.5 mM EDTA).
9. Deoxynucleotide, 100 mM solutions (dATP, dGTP, dCGT, and dTTP).
10. Ultrapure RNase-free water.

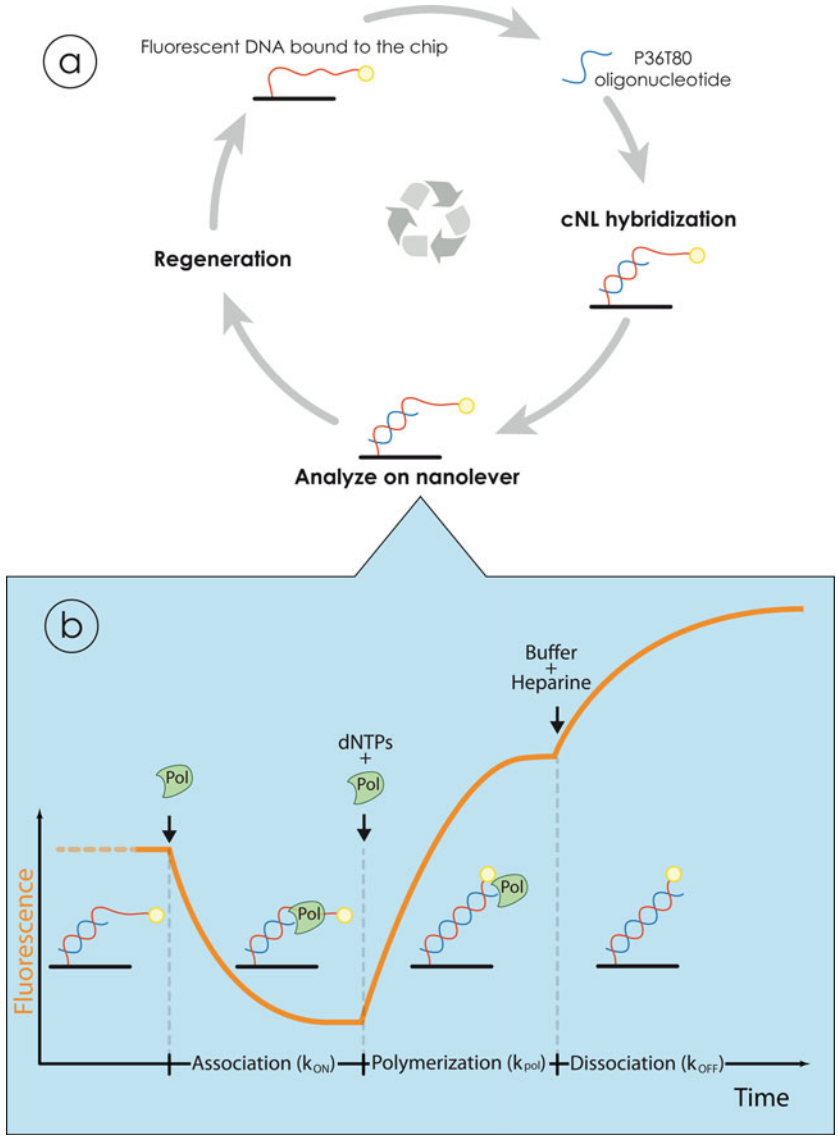


Fig. 1 Schematic representation of the HIV-1 DNA polymerase switchSENSE experiment. (**a**) Illustration showing the building of the DNA nanolever following by regeneration of the biosensor chip. (**b**) Scheme of the switchSENSE measurement described here, containing three main steps: association, polymerization, and dissociation

11. Ethanol absolute 99.9%.
12. 1.5 and 10 mL autosampler vials with Septa caps and insets.

2.2 Enzyme Preparation

1. Reverse transcriptase HIV-1, BH10 isolate (RT) was expressed and purified as described previously [15] and stored as a suspension in a 2 M $(NH_4)_2SO_4$ solution.

2. Heparin solution stock (10 mg/mL) is diluted in ultrapure RNase-free water.
3. 10× Running Buffer (100 mM potassium phosphate, pH 8.0, 800 mM NaCl, 60 mM $MgCl_2$, 0.5% Tween 20).
4. Amicon Ultra-4 centrifugation filter unit 30,000 MWCO (Millipore, Billerica, MA, USA).

3 Methods

3.1 Polymerase Preparation for switchSENSE Experiment

1. Aliquots of HIV-1 reverse transcriptase (ammonium sulfate stock solution) are centrifuged for 20 min at 15,000 × *g*, 4 °C. Protein pellets are resuspended for 1 h on ice in the 1× Running Buffer at final concentration of ~1 mg/mL.
2. The protein is then dialyzed against 1× Running Buffer using an Amicon Ultra-4 by three successive 10-min centrifugations at 7,500 × *g*, 4 °C.
3. After the last centrifugation step, the polymerase is concentrated at a final concentration of ~1 mg/mL.
4. Precise concentration of polymerase is obtained using two independent measures with a nanodrop spectrophotometer ($\varepsilon 280 = 231{,}480\ mol^{-1}\ cm^{-1}$ for pure HIV-1 reverse transcriptase).

3.2 Solutions Preparation

1. 300 µL of RT diluted with 1× Running Buffer (ranging from 0.03 to 4 nM, based on concentration estimated in the previous step).
2. 250 mL of freshly diluted 1× Auxiliary Buffer and 1× Running Buffer are filtrated on a 2-µm nitrocellulose filter, inserted in the fluidic system of the DRX 2400 and primed using the switchCONTROL command of the system software.
3. 300 µL of p36T80 oligonucleotide diluted in ultrapure water at a final concentration of 500 nM.
4. 1 mL of regeneration solution.
5. 10 mL of freshly diluted 1× passivation solution.
6. All the solutions and buffers are arranged as shown in the switch CONTROL software window.

3.3 Schematic Overview of the Experiment

1. Building of the DNA nanolever is the first step of the experiment. This reaction is achieved by the hybridization of the P36T80 oligonucleotide on the 80-mer DNA probe attached to the chip (3′ end), to generate a DNA duplex with a 44 bases 5′ overhang (Fig. 1a, *see* **Note 2**).

2. One concentration of polymerase in the 1× Reaction Buffer is then injected into the fluidic system and the fluorescence variation recorded to retrieve the k_{ON} (Fig. 1b—association).
3. Without regeneration of the chip, a second injection is directly performed with a 1× Reaction Buffer containing the same concentration of polymerase (*see* **Note 3**) and an oversaturated concentration of the four dNTP mixed nucleotides. The increase in fluorescence signal is recorded over time to retrieve the polymerization rate signal (Fig. 1b—polymerization, *see* **Note 4**).
4. A last injection of 1× Reaction Buffer completed with heparin (*see* **Note 5**) on the non-regenerated chip is used to retrieve the k_{OFF} (Fig. 1b—dissociation).

3.4 Experimental Design

1. The experimental workflow is designed using switchBUILD software by making successive program blocks.
   (a) In the first block, channel number, buffer type (P40), and biochip (ENZ-80-1) are designed.
   (b) A "Passivation" step is added as a second block (*see* **Note 6**).
   (c) The third block is a "Kinetics" step, where the ligand is the P36T80 oligonucleotide, and the analyte is a defined concentration of polymerase (*see* **Note 7**), the measurement mode set as "Static Mode" (*see* **Note 8**). Following parameters are also defined in this block: temperature, electrode number (one of the six available in each channel), association time (*see* **Note 9**), flow rate (*see* **Note 10**), uncheck option for "regeneration." If it is desired to change any condition in the experiment (temperature, polymerase concentration, etc.), a new block of this type, with the corresponding modified parameter, should be inserted next. It is also possible to add a blank-run block, where the analyte is replaced with 1× Reaction Buffer alone: resulting curves could be used for baseline correction if a significant drift of the fluorescence signal is observed.
   (d) A last "Standby" block is used at the end of the experiment (*see* **Note 11**).
   (e) Once completed, the switchBUILD file is saved, and sample vials are disposed as requested in the "Autosampler" window of the software.
2. The previous file is loaded in the switchCONTROL software. If no other modifications are needed (*see* **Note 12**), the experiment could then start (typically 1 h 30 min for a complete cycle with one polymerase concentration). Examples of data generated are provided on Fig. 2.

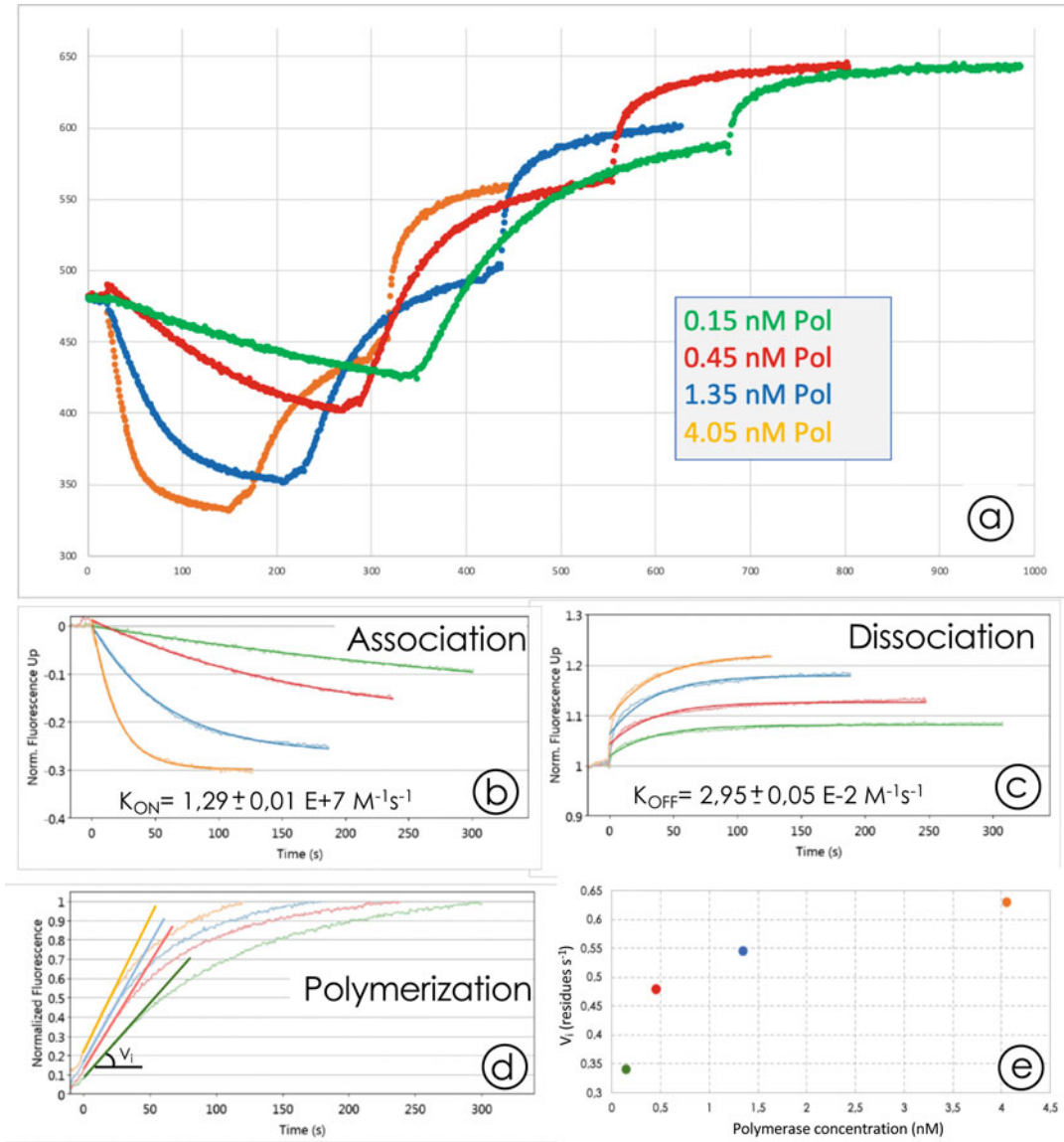


Fig. 2 Example of DNA polymerization experiment using four HIV-1 RT concentrations. Fluorescence signal observed upon addition of HIV-1 RT onto DNA nanolevers (**a**). on- (**b**) and off (**c**) rates obtained using the switchANALYSIS software package. Determination of the k_{pol} using switchANALYSIS

3. Once completed, datasets generated during the "Kinetics" step are loaded in the switchANALYSIS software package. For each concentration of polymerase, three distinct kinetics could be successively analyzed: k_{ON}, k_{POL}, and k_{OFF}. If several concentrations of polymerase have been used, a global fit can be used in order to significantly improve the accuracy of each fitted value. Polymerization rates (in residue s^{-1}) are calculated considering a 44 base elongated DNA.

4. Once all datasets are analyzed in the switchANALYSIS software, a publication-quality image could be copied in the clipboard. Raw datasets could also be exported in various file formats (ASCII, Excel, etc.) compatible with other software tools.

3.5 Consideration About Obtained Kinetics Parameters

Polymerization rate is only fitted during the first 30 s of the reaction. As for classical studies in enzymology, this time window corresponds to the initial linear part allowing to measure initial rate of the reaction, before the substrate becomes limiting.

In this experiment datasets, please note that the determined k_{ON} value stands for the association of polymerase on a 5′ protruding DNA strand and that the k_{OFF} value stands for dissociation of the polymerase from a fully extended blunt end DNA. Therefore, no proper K_D value (usually deduced from the ratio $\frac{k_{OFF}}{k_{ON}}$) could be calculated, because of the change in the nature of the substrate that occurs during polymerization. K_D of the polymerase for both forms of the DNA could be calculated using a different experimental set using the polymerase as an analyte and two complementary DNA nanolevers of 36 or 80 nucleotides long as ligands.

4 Notes

1. P36T80 oligonucleotide (GAGTAACCTGAAGTGTGAGCATAAAGAGTGATGTAA) was chemically synthetized at a 250 nmoles scale, followed by a standard desalting.
2. This construct constitutes the substrate of the polymerase where P36T80 is the primer strand oligonucleotide, and template strand is the 80-mer DNA probe attached on the chip on its 3′ end.
3. The enzyme concentration is kept constant in the fluidic system during polymerization to maintain the balance of associated/dissociated polymerase on the DNA substrate.
4. During this step, the fluorescence increases significantly because the dye moves away from the fluorescence-quenching surface as the polymerase converts the upper DNA part from a floppy single- to a rigid double strand.
5. As for many polymerase-involved reactions, heparin is used as a trap to ensure that the dissociated polymerase do not re-associate on the DNA nanolever. The use of heparin allows the same fluorescent signal amplitude to be recovered for dissociation as for the association step.

6. This automatized standard procedure (using a commercially available kit solution) is used to passivate the chip, avoiding any unspecific binding of analytes of the biosurface.
7. The procedure to analyze three successive steps for polymerase reaction is not usual and deviates from the pre-programmed assay in the switchBUILD software. For this reason, we have chosen to insert three successive injections, each starting directly after the previous one, without regenerating the chip. For this purpose, three successive dilutions have to be defined manually, and the "Regeneration" option should be unchecked. Instead of filling the tubes with three different dilutions of polymerase (as requested in the predefined parameters), the first dilution tube will contain one concentration of polymerase diluted with the 1× Reaction Buffer, the second will contain the same concentration of polymerase and 250 μM of each dNTP, and the third dilution tube will contain 1× Reaction Buffer supplemented with 0.1 mg/mL heparin.
8. Static mode means that a constant voltage (−0.1 V in our conditions) is applied on the chip, maintaining the nanolever at a constant angle in the 1× Reaction Buffer. Any event affecting this angle and/or the distance of the fluorophore with the quenching surface of the biochip (protein binding, polymerization, dissociation) will be an interpretable signal for the experiment.
9. Completion of association, polymerization, and dissociation steps are dependent of kinetics parameters of the considered polymerase, concentration, and temperature. Typically, for the HIV-1 RT polymerase at 25 °C, these three steps are set to 5, 4, 3, and 2 min for the concentrations of 0.15, 0.45, 1.35, and 4.05 nM, respectively.
10. Flow rate is automatically set from the estimated k_{ON} value set in the "predicted interaction value" tool implemented in the switchSENSE software. For HIV-1 RT polymerase (k_{ON} = 1.3 E+7 M^{-1} s^{-1} at 25 °C), flow rate is automatically set to 50 μL/min for all the steps.
11. This automatized standard procedure removes analytes and ligands from the surface and defines conditions suitable for chip storage.
12. The switchCONTROL software takes the simplicity of switchBUILD scripts and converts them into advanced commands for the unattended operation of the DRX instrument.

References

1. Rich RL, Myszka DG (2011) Survey of the 2009 commercial optical biosensor literature. J Mol Recognit 24:892–914
2. Knezevic J, Langer A, Hampel PA et al (2012) Quantitation of affinity, avidity, and binding kinetics of protein analytes with a dynamically switchable biosurface. J Am Chem Soc 134:15225–15228
3. Langer A, Hampel PA, Kaiser W et al (2013) Protein analysis by time-resolved measurements with an electro-switchable DNA chip. Nat Commun 4:2099
4. Alt A, Dang HQ, Wells OS et al (2017) Specialized interfaces of Smc5/6 control hinge stability and DNA association. Nat Commun 8:14011
5. Blocquel D, Li S, Wei N et al (2017) Alternative stable conformation capable of protein misinteraction links tRNA synthetase to peripheral neuropathy. Nucleic Acids Res 45:8091–8104
6. Denichenko P, Mogilevsky M, Clery A et al (2019) Specific inhibition of splicing factor activity by decoy RNA oligonucleotides. Nat Commun 10:1590
7. Nemoz C, Ropars V, Frit P et al (2018) XLF and APLF bind Ku80 at two remote sites to ensure DNA repair by non-homologous end joining. Nat Struct Mol Biol 25:971–980
8. Ploschik D, Ronicke F, Beike H et al (2018) DNA primer extension with cyclopropenylated 7-Deaza-2′-deoxyadenosine and efficient bioorthogonal labeling in vitro and in living cells. Chembiochem 19:1949–1953
9. Rueda FO, Bista M, Newton MD et al (2017) Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. Nat Commun 8:1610
10. Webster MW, Chen YH, Stowell JAW et al (2018) mRNA deadenylation is coupled to translation rates by the differential activities of Ccr4-not nucleases. Mol Cell 70:1089–1100. e8
11. Webster MW, Stowell JA, Passmore LA (2019) RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. elife 8:e40670
12. Kroener F, Heerwig A, Kaiser W et al (2017) Electrical actuation of a DNA origami nanolever on an electrode. J Am Chem Soc 139:16510–16513
13. Clery A, Sohier TJM, Welte T et al (2017) switchSENSE: a new technology to study protein-RNA interactions. Methods 118-119:137–145
14. Langer A, Schraml M, Strasser R et al (2015) Polymerase/DNA interactions and enzymatic activity: multi-parameter analysis with electro-switchable biosurfaces. Sci Rep 5:12066
15. Freisz S, Bec G, Radi M et al (2010) Crystal structure of HIV-1 reverse transcriptase bound to a non-nucleoside inhibitor with a novel mechanism of action. Angew Chem Int Ed Engl 49:1805–1808

Chapter 9

Sedimentation Velocity Methods for the Characterization of Protein Heterogeneity and Protein Affinity Interactions

Christine Ebel and Catherine Birck

Abstract

Sedimentation velocity analytical ultracentrifugation is a powerful and versatile tool for the characterization of proteins and macromolecular complexes in solution. The direct modeling of the sedimentation process using modern computational strategies allows among others to assess the homogeneity/heterogeneity state of protein samples and to characterize protein associations. In this chapter, we will provide theoretical backgrounds and protocols to analyze the size distribution of protein samples and to determine the affinity of protein–protein hetero-associations.

Key words Sedimentation velocity, Analytical ultracentrifugation, Protein heterogeneity, Sedimentation coefficient, Hetero-association, Protein interaction

1 Introduction

The characterization of biological samples to assess their homogeneity and to study the formation of protein complexes is of interest in many areas of biological research. Getting a homogeneous sample is among others an essential requirement for the accuracy and reproducibility of the experimental results, the correct interpretation of biophysical data, and high-resolution structural work. It will also benefit protein interaction studies, as the presence of aggregates may decrease the fraction of protein competent for binding or lead to improper modeling of the interaction. Few methods allow the characterization of protein homogeneity and protein binding properties. It is the case of sedimentation velocity (SV) analytical ultracentrifugation (AUC) which is a powerful biophysical technique commonly used to determine the size and shape of macromolecules in solution and to study protein interactions. It is an absolute method based on first principles that benefits from constant developments in analysis software. In SV, macromolecules loaded in sector-shaped cells are separated at high centrifugal force and their sedimentation profiles, the evolution of the protein

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_9,

concentration with time and radial position, are monitored in real time using the optical system of the analytical ultracentrifuge. Three optics are currently available, absorbance, interference, and fluorescence optics which allow the detection of a variety of macromolecules in a broad range of concentrations. Compared to other hydrodynamic techniques like dynamic light scattering (DLS) and size exclusion chromatography, SV presents a higher hydrodynamic resolution and a wider range of protein molecular weights to be studied, from a few hundred daltons to several hundred-million daltons, corresponding to peptides and large macromolecular complexes, respectively. Furthermore, the presence of contaminants or aggregates below 1% can be quantified using SV [1, 2]. The analysis of the sedimentation profiles allows the determination of the sedimentation and the diffusion coefficients of the macromolecules, from the motion and shape of the sedimenting boundaries, and the calculation of their molecular mass using the Svedberg equation (*see* Subheading 1.1). For non-interacting species, these hydrodynamic parameters are directly related to the size and shape of macromolecules. SV can also contribute to our understanding of macromolecular assemblies with the determination of the equilibrium constant governing the interaction and the size of the macromolecular complex, and a robust approach to obtain these parameters is the isotherm analysis. Binding isotherms are derived from sedimentation coefficients distributions of a concentration series and fit with a hetero-association model to obtain the dissociation constant and sedimentation coefficients of the macromolecular complexes. Concerning the stoichiometry of the complex, it is deduced from the quality of the fit using different association models but may require a more complicated analysis using multi-signal SV approach, especially in the case of multi-component systems. The dynamic range in binding affinities that can be explored in SV is notably high from picomolar to millimolar K_d values due to the availability of the three optics and their different sensitivity. It is noteworthy that the fluorescence detection system that was recently developed allows now the study of protein interactions at low picomolar concentrations [3].

The present chapter aims to provide a theoretical background of SV, to detail the SV protocols to characterize the heterogeneity of a protein, and to measure a binary protein–protein interaction. Two case studies will be presented. As SV technology and expertise are accessible to researchers via many research infrastructures (*see* **Note 1**), an additional goal of this chapter is to contribute to make known the SV applications and raise interest for this technique. More deep descriptions of the theory and applications of SV can be found in recent reviews [2, 4–7].

1.1 Sedimentation Velocity Equations

Sedimentation velocity measures in a rotor spinning at high angular velocity, ω, in the centrifuge, the evolution of the weight concentration, c, with time, t, and radial position, r. For each homogeneous ideal solute, and given the sector shaped cells used in analytical ultracentrifugation, the transport is described by the Lamm equation:

$$(\partial c/\partial t) = -1/r\ \partial/\partial r\ \left[r\left(cs\omega^2 r - D\ \partial c/\partial r\right)\right], \tag{1}$$

where s and D are the sedimentation and diffusion coefficients of the macromolecule. s is defined as the ratio of the macromolecule velocity (cm s^{-1}) to the centrifugal field ($\omega^2 r$ in cm s^{-2}). s is expressed in Svedberg unit S (1 S = 10^{-13} s). s and D are functions of the molar mass M, the hydrodynamic radius R_H (also referred to as the Stokes radius R_S), and the partial specific volume $\overline{\nu}$ of the macromolecule. s and D also depend on the solvent density ρ and viscosity η. The Svedberg equation relates s to R_H (or D), M, and $\overline{\nu}$:

$$s = M(1 - \rho\overline{\nu})/(N_A 6\pi\eta R_H) = M(1 - \rho\overline{\nu})D/RT \tag{2}$$

N_A is Avogadro's number and T the absolute temperature.

The Stokes–Einstein equation relates D to R_H:

$$D = RT/(N_A 6\pi\eta R_H) \tag{3}$$

The ratio of R_H to the minimum theoretical hydrodynamic radius R_{min} of non-hydrated volume, V, of the particle defines the frictional ratio f/f_{min}:

$$V = (4/3)\pi {R_{min}}^3 = M\overline{\nu}/N_A \tag{4}$$

$$R_H = (\ f/f_{min})\ R_{min} \tag{5}$$

The value of f/f_{min} depends on shape and hydration, but not, for compact macromolecules, on the size or molar mass [8]. A typical value is 1.25 for globular compact particles, 1.5 for a moderately anisotropic shape and 1.8 for a significative anisotropic shape (*see* e.g. [9]). Glycosylated proteins are characterized by larger values, such 1.5–1.8, for a moderate asymmetrical shape (*see* e.g. [10]). f/f_{min} can also be calculated from pdb files [11].

Because of its dependence on buffer viscosity and density, the experimental s value (s_{exp}) is often normalized to standard solution conditions of water at 20 °C ($s_{20,w}$):

$$s_{20,w} = s\left[\left(1 - \rho_{20,w}\overline{\nu}\right)/(1 - \rho\overline{\nu})\right]\left(\eta/\eta_{20,w}\right) \tag{6}$$

D is also often expressed as $D_{20,w}$:

$$D_{20,w} = D\ (T_{20}/T)\ \left(\eta/\eta_{20,w}\right) \tag{7}$$

Non-ideality effects in concentrated samples influences s and D [12]. The sedimentation and diffusion at infinite dilution, s_0 and D_0, can be derived from the linear approximations:

$$s^{-1} = {s_0}^{-1} (1 + k_s c) \quad (8)$$

$$D = D_0 (1 + k_D c) \quad (9)$$

1.2 The c(s) Analysis

During centrifugation, domination of the sedimentation term gives rise to a migrating boundary that spreads with time because of an opposing diffusional flow in response to the resulting concentration gradient [2, 13]. The time evolution of the radial concentration distribution during sedimentation is given by the Lamm equation (Eq. 1). Numerical solutions to the Lamm equation are used in many SV analysis programs to extract s and D from SV experimental data.

The $c(s)$ method implemented in the SV program SEDFIT [14, 15] allows to obtain a high-resolution distribution of particle according to their s-values. The $c(s)$ analysis considers a large number of types of particles. It deconvolutes the effect of diffusion broadening, assuming that all proteins have the same shape and density (same $\overline{\nu}, f/f_{min}$), which gives a relationship between s and D, or, in an equivalent way, s and M, or s and R_H (Eqs. 2–5). The program simulates for each type of particle (each s-value), the set of SV profiles, and allows to refine the value of f/f_{min}. It then determines the best combination of simulated SV profiles that fits the experimental data. The resulting $c(s)$ distribution shows peaks for all sedimenting species in solution, assuming a constant shape. Peaks of the $c(s)$ plot can be easily integrated, providing their s-value, their signal, and the percentage of each species.

1.3 Characterization of Protein Heterogeneity Using SV

1.3.1 Non-interacting Versus Interacting System

It is noteworthy that SV data are highly sensitive to the presence of trace aggregates, and a wide range of s-values can be covered in the $c(s)$ distribution from a single experiment. The $c(s)$ analysis allows to determine if samples are heterogeneous. However, a $c(s)$ plot with only one peak does not mean that the sample is homogeneous. In general, there are fewer peaks than species when particles are interacting. For this reason, SV profiles should be acquired at different concentrations. The nice superposition of the normalized $c(s)$ plots for protein samples at different concentrations indicates that the protein does not self-associate. A shift in the peak position indicates that the protein is in an equilibrium of association. When working at high protein concentrations, above the mg/mL range, a slight decrease of s-value is expected due to non-ideality effect (Eq. 8). This is a complication in the analysis of weakly interacting systems [12].

1.3.2 Non-interacting System Analysis

When *s*-values are invariant with concentration, the sample may be considered as composed of non-interacting species, a peak in the $c(s)$ representing a species. The non-interacting species analysis then allows the determination of independent values for s and D, thus M and R_H, if the system consists of a limited number of species (typically one or two).

1.3.3 Example of a Non-interacting Protein Species

Bovine serum albumin (BSA) will be used as an example of non-interacting protein species. BSA is a 66-kDa protein, well-known to form irreversible oligomers. The distribution of monomeric, dimeric, and trimeric species is commonly studied in AUC to determine the instrument performance [16]. The SV data of two BSA samples taken before or after a gel filtration step will be analyzed to illustrate the behavior of heterogeneous versus homogeneous samples.

1.4 Characterization of Protein Hetero-association Using SV

SV offers two significant advantages for the study of interacting macromolecules: a relatively high hydrodynamic size-dependent resolution allowing the separation of free and bound species and the fact that during the experiment, the complexes will remain in a bath of the free species. Thereby, dissociating complexes can re-associate during the sedimentation process in a way that will reflect their equilibrium and kinetic properties [17]. The standard methods of analyzing SV data to obtain association constants between interacting macromolecules are Lamm equation fitting and $c(s)$-based isotherm analysis. Practically, the determination of equilibrium and kinetic binding constants requires to run multiple experiments at concentrations that cover a range of about 1/10 to 10 K_d. It is advisable to make dilution or titration series to prepare samples. In this way, these values can be refined with added constraints linking the concentrations of different experiments. In dilution series of stock mixtures, the molar ratio is constant and can be fixed or refined as a single global parameter. In titration series, one protein is kept at low constant concentration in each sample while the concentration of the binding partner varies in a wide range.

1.4.1 Lamm Equation Fitting to Characterize Protein Hetero-association

Current computational strategies allow for direct boundary modeling of SV data for reacting systems with a set of coupled Lamm equations describing all the species participating in the interaction combined with information on the equilibrium association constant and reaction kinetics [18, 19]. It is noteworthy that only this approach allows to estimate the kinetic off-rate constant k_{off} of interactions (*see* **Note 2**). This direct boundary fitting method has been implemented in SEDPHAT with different binding models available. Compared to the SV isotherm analysis, this analysis can be conducted when only a few SV data sets are available (minimum

two) and yield more precise parameter estimates about K_d, k_{off}, and $s_{complex}$. This approach has been demonstrated to be efficient in a number of cases but has several limitations including the fact that individual species need to be reasonably described as discrete species with the correct molecular weights. This is not always the case, for example, when impurities contribute to boundary broadening and thus preclude a good model fitting and accurate results. Furthermore, as this method will take some time to converge, it is not the first method of choice, but may be conducted using parameter values obtained with the isotherm analysis.

1.4.2 Isotherm Analysis to Characterize Protein Hetero-association

For the construction of an isotherm, the first step is the $c(s)$ analysis of all SV data corresponding to the concentration series. In fact, the $c(s)$ analysis can also be applied to slowly as well as rapidly interacting systems. For slowly interacting systems, the boundary pattern directly reflects the population of different species, which can often be hydrodynamically resolved. In the case of a bimolecular interaction of the type A + B forming an AB complex with s-values such that $s_A < s_B < s_{AB}$, three boundaries will be observed corresponding to free A, free B, and complex AB. For rapidly interacting systems, the association-dissociation events during sedimentation lead to at most two boundaries, one always sedimenting at the s-value of either free A or free B, termed undisturbed boundary, and the other at a composition-dependent s-value between s_B and s_{AB}, termed the reaction boundary. When the integration limits encompass all sedimenting species, the result is termed signal-weighted average sedimentation coefficient, s_w, which is directly related to the overall transport of the interacting system and independent of the kinetics of the interactions. When the integration limits encompass only the reaction boundary, the result is termed s_{fast} referring to the sedimentation coefficient of the fast-sedimenting species. The resulting s_w and s_{fast} isotherms can be combined and fitted with mass action law models implemented in program SEDPHAT to determine the binding constants and species size [19]. In addition to these two s_w and s_{fast} isotherms, a population isotherm can be constructed by integrating each $c(s)$ peak and tabulating the reported amplitudes, in signal units, and combined in the analysis.

1.4.3 Example of Protein Hetero-association

We will use the association between the BTG2 factor and the poly (A)-binding protein PABP C1 as model system to illustrate protein hetero-association. These proteins were shown to interact directly and their association sufficient in vitro and necessary in cellulo for BTG2 to stimulate mRNA deadenylation. The BTG2/PABP interaction is also required for BTG2 to exert its antiproliferative function [20].

2 Materials

2.1 Protein Samples

1. BSA (SIGMA) was dissolved at 5 mg/mL in 20 mM Na-K Phos 150 mM NaCl pH 7 and filtered at 45 μm. BSA1 sample was obtained by five times dilution to 1 mg/mL. BSA2 sample at 0.8 mg/mL was obtained from the main elution peak of 500 μL BSA stock sample injected at 0.5 mL/min on a Superose 12 10/300 GL (GE Healthcare) size exclusion chromatography column.
2. Recombinant BTG2 fused to glutathione S transferase (GST-BTG2) and PAPB proteins were purified by affinity and size exclusion chromatography [20]. Stocks of GST-BTG2 protein at 1.1 mg/mL and PABP at 5.1 mg/mL in buffer 10 mM K phosphate, pH 7.5, 100 mM NaCl, 0.5 mM TCEP were used in this study (*see* **Note 3**). In the dilution series, mixtures contain a fixed concentration of GST-BTG2 (5 μM) and a range of seven PABP concentrations to give molar ratio (PABP/GST-BTG2) of 1:2, 1:1, 2:1, 4:1, 8:1, 16:1, and 32:1.

2.2 Instrument and Sample Cells

1. An analytical ultracentrifuge (Optima XLI Beckman) and rotor (8-hole AnTi-50, Beckman). Analytical ultracentrifugation cell assemblies equipped with sapphire windows and two-channel Titanium, Aluminum or Epon centerpieces of 1.5, 3, and/or 12 mm optical path length (Nanolytics or Beckman).
2. Programs: The program SEDNTERP created by D. Hayes, T. Laue, J. Philo and available free (http://bitcwiki.sr.unh.edu/index.php/Downloads) for calculating the parameters relevant to SV analysis (*see* **Note 4**). The program SEDFIT created by P. Schuck and available free (http://sedfitsedphat.nibib.nih.gov), for the $c(s)$ analysis of SV experiments [14]. The program SEDPHAT created by P. Schuck and available free (http://sedfitsedphat.nibib.nih.gov), for the global analysis of different AUC [21] and other biophysical experiments [22, 23], for the analysis of isotherms [17] and to generate simulated data sets. The program GUSSI created by C. Brautigam and available free (http://biophysics.swmed.edu/MBR/software.html), for calculations (e.g., the integration of the $c(s)$ peaks, the output of isotherm data files), and publication quality illustrations of AUC experiments [24].

3 Methods

3.1 Data Acquisition

1. Prepare the analytical ultracentrifugation cells assemblies. Fill the sample compartments with 400 μL (12 mm path length centerpiece) or 100 μL (3 mm path length centerpiece) of protein solutions and the solvent compartments with 420 or

110 μL of buffer (*see* **Note 5**). Close the cells and check that cells that will be placed in opposite positions in the rotor have the same weights (tolerance of 0.5 g). Place and align the cells in the rotor holes.

2. Prepare the instrument: Put the rotor then the optical arm into the analytical ultracentrifuge, and initiate vacuum and temperature equilibration. Wait at least 2 h at the requested temperature before starting the SV experiment.
3. Check the absorbance signal in each cell. During temperature equilibration, start the program ProteomeLab XL-I. Give a title for each cell and a common folder for the collected data. Start the AUC at 3000 rpm and acquire for each cell a radial scan at 280 nm to check that the absorbance signal corresponds to expectation.
4. Start the SV run: In the method window, enter 150 for the number of scans. In the options window, enter 3 for the last scans overlay and select stop XL after the last scan. Start the centrifuge at the chosen speed (*see* **Note 6**) then start the method scan. After the stop of the ultracentrifuge, cells are disassembled, cleaned, and reassembled. Raw data are copied for analysis. A detailed protocol can be found in [8].

3.2 Non-interacting System Data Analysis with BSA Data Sets

We perform heterogeneity analysis of two BSA samples using both $c(s)$ and non-interacting species models to illustrate the potential and limits of these approaches. The first sample, BSA1, which contains aging aggregates is heterogeneous while the second, BSA2, freshly prepared from size exclusion chromatography and composed of monomeric BSA, is homogenous.

1. The $c(s)$ analysis: Open the program SEDFIT and load a set of SV scans corresponding to the whole sedimentation process (Fig. 1). Set the meniscus with the red line (air-sample interface) and the radial limits for the fit with the green lines. Choose the continuous $c(s)$ distribution model. In the parameter window, enter the experimental values for partial specific volume, density, and viscosity, change the confidence level to 0.68 (*see* **Note 7**) and mark the frictional ratio and meniscus check boxes. Use the default values for all other parameters (100 for resolution, 0–20 S for sedimentation coefficient range, 1.2 for frictional ratio value) and press the run command to perform a simulation of the sedimentation process using the given parameters. The simulated data are displayed as lines with the loaded scans, and a first $c(s)$ distribution plot appears. Then press the fit command to optimize the checked parameters (frictional ratio and meniscus position). After convergence, simulated data should fit very well to the experimental data

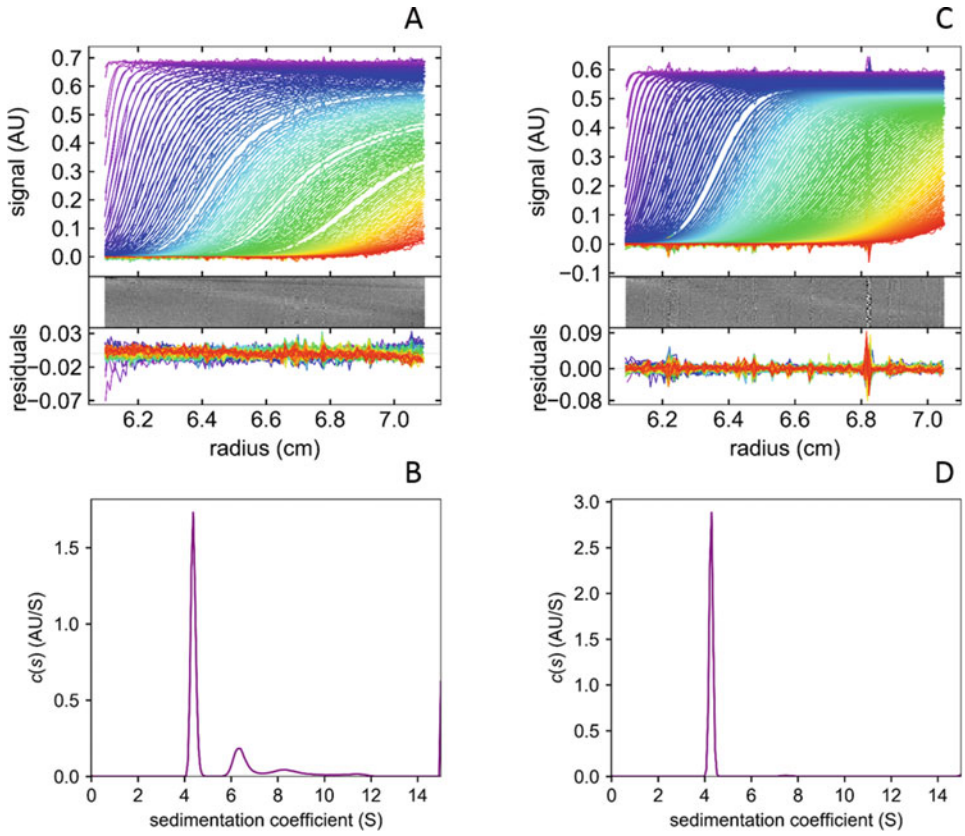


Fig. 1 SV data and *c*(*s*) analysis of BSA. Superposition of the experimental and fitted sedimentation velocity profiles for BSA before (**a**) and after (**b**) size exclusion chromatography—samples BSA1 and BSA2, respectively. Residuals of the fit are displayed in gray or absorbance scales. Corresponding *c*(*s*) distributions are shown below each SV data. Sedimentation velocity profiles were obtained at 280 nm, at 20 °C, and 130,000 × *g* (42,000 rpm) using double-sector centerpieces with 12 mm optical path length. Analysis was performed with the first 100 profiles, collected during 350 min. The *c*(*s*) analysis was obtained considering 200 particles with s-values in the 0–15 S range and a partial specific volume of 0.73 mL/g solvent density of 1.005 g/mL and a viscosity of 1.02 cp. A confidence level of 0.68 was used for the regularization procedure. The rmsd were 0.006 and 0.009 for the BSA1 and BSA2 samples, respectively. The fitted frictional ratio were 1.25 and 1.33, respectively

(criteria for the quality of the fit are low rmsd value, less than 0.01 signal units under most cases, and randomly distributed residuals around zero) to get the true best-fit distribution *c*(*s*). Integration of the peaks in the *c*(*s*) distribution will give the values of the different molecular parameters. Detailed steps of the *c*(*s*) analysis can be found in SEDFIT help website (www.analyticalultracentrifugation.com/sedfit_help.htm). Figure 1, panels A and B, shows a nice fit for the two BSA samples. When BSA is diluted, the sample consist of a mixture of monomers, dimers, trimers, and traces of higher oligomers which can

be easily quantified by integration of the peaks. After size exclusion chromatography, the sample is homogeneous, with only one species in solution at 4.3 S, corresponding to the globular compact monomer. Note that to ascertain the fact that this peak corresponds to a non-interacting species, we should have measured the samples at different concentrations. Here we obtain this conclusion because the main species in the heterogeneous samples have the same *s*-value.

2. The non-interacting species analysis: in SEDFIT, choose the non-interacting discrete species model. In the parameter window, enter estimate values for the total concentration, the molar mass, and *s*-value of the species and mark the corresponding check boxes. Press the run command then the Fit command. Figure 2 shows the result of an analysis in terms of one non-interacting species for the two samples. This choice is done for pedagogical purpose, since the $c(s)$ analysis clearly showed that diluted BSA (without purification) is heterogeneous. Figure 2 shows that the fit for the homogeneous sample is excellent: the residuals are randomly null. The calculated molar mass is 65.9 kDa, in very good agreement with the theoretical—and checked by mass spectrometry—value of 66.4 kDa. On the other hand, a non-cautious analysis could describe a nearly good superposition of the experimental and fitted SV profiles for the heterogeneous sample in Fig. 2a. But the residuals show clearly non-random deviations. And,

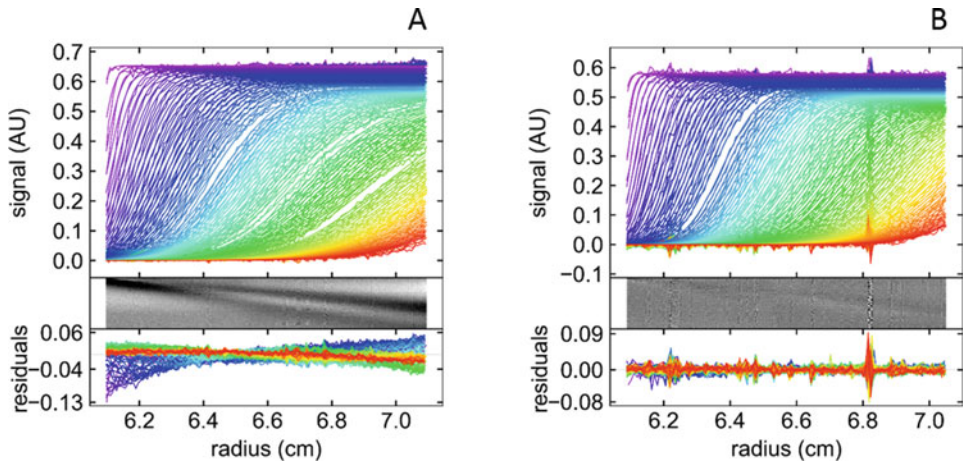


Fig. 2 SV data and analysis of BSA using a non-interacting species model. Superposition of the experimental and fitted SV profiles for BSA, before (**a**) and after (**b**) size exclusion chromatography—samples BSA1 and BSA2, respectively—using one non-interacting species model. Residuals of the fit are displayed in gray or absorbance scales. Experimental conditions are detailed in the legend of Fig. 1. Sedimentation coefficients of 4.7 and 4.3 S with the corresponding molar masses of 27.4 and 65.9 kDa and rmsd on fits of 0.018 and 0.009 were obtained for BSA1 and BSA2 samples, respectively

because the model is inappropriate, the calculated molar mass of 27.4 kDa is definitively wrong, half the theoretical value for the monomer. This is because boundary spreading is interpreted by the program—in the framework of the incorrect model—as caused by diffusion, while heterogeneity affects more predominantly the shape of the boundaries. In conclusion, the non-interacting species model provides valuable results, but only when this model is appropriate.

3.3 Hetero-association Data Analysis with BTG2-PABP Mixtures Data Sets

We illustrate here $c(s)$ analysis in SEDFIT and integration of the sedimentation coefficient distributions in GUSSI to determine s_w and s_{fast} values, reflecting the properties of the interacting system. Isotherm analysis in SEDPHAT.

1. Perform a $c(s)$ analysis in SEDFIT with SV data for each mixture of the titration series and save the corresponding $c(s)$ distribution in GUSSI format. Open the program SEDFIT and perform the $c(s)$ analysis with SV data for the first mixture. Excellent fit of the raw data is essential for the accuracy of parameters determined in next steps. The $c(s)$ distribution is exported to GUSSI using Plot/Gussi $c(s)$ plot, then accept the GUSSI terms in the GUSSI window that appears, select File/Save data only, choose the correct directory, give a name to the file and save. A file *.gcofs is created. Then File/quit. Proceed the same way for all SV data.
2. Superimpose all the $c(s)$ distributions corresponding to the titration series and transform experimental $c(s)$ in corrected $c(s)$ distributions (Fig. 3). Open GUSSI, accept terms and press OK in the GUSSI window with default $c(s)$ selection. In the GUSSI $c(s)$ module window, select File/Add Distribution and load the GUSSI file *.gcofs corresponding to the first $c(s)$ distribution (corresponding to the lowest concentration of A). Repeat the loading procedure in ascending order of concentration for all the $c(s)$ distributions. Select Axis/Standardization/Standardization, then Axis/Standardization/Modify Standardization Parameters, and enter experimental conditions in Standardization window for density, viscosity, partial specific volume (values at experimental temperature and 20 °C), and temperature. Then press propagate to all distributions and finally press commit. The experimental $c(s)$ distributions are transformed to corrected $c(s)$ distributions. Then File/Save GUSSI State to save the corrected $c(s)$ superimposition file. This *.gussi file can be recalled if needed.
3. Integrate all the superimposed corrected $c(s)$ distributions to generate the signal-weighted average s_w isotherm file (Fig. 4). In GUSSI, select Integration/Make Isotherm/Hetero/sw, press no for the exclusion zone selection (*see* **Note 8**), and

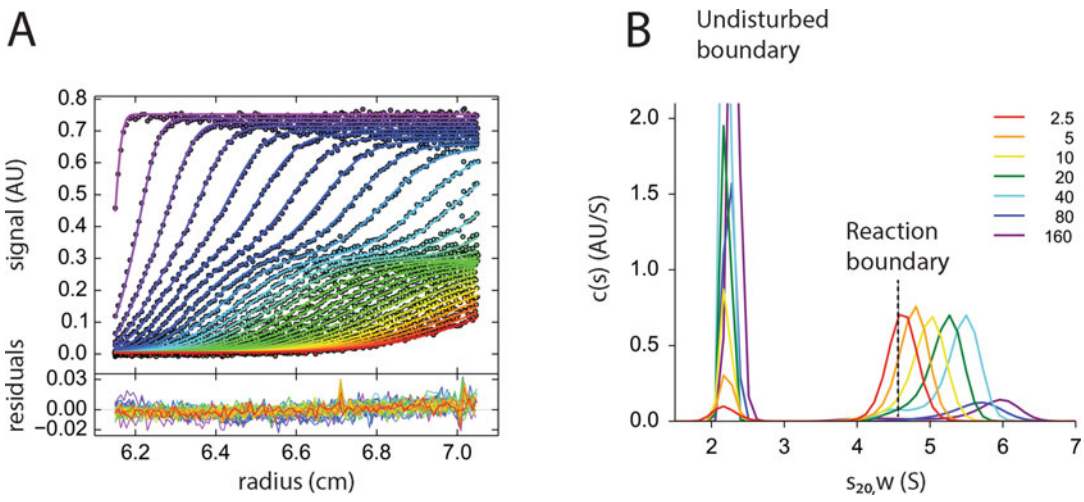


Fig. 3 SV data and *c*(*s*) analysis of GST-BTG2: PABP mixtures. (**a**) Experimental SV data, fit, and residuals for a mixture of 5 μM GST-BTG2 and 20 μM PABP. Individual data points are shown as circles and fits to experimental data as lines. Residuals of the fit are displayed in absorbance scales. Sedimentation velocity profiles were obtained at 280 nm, at 4 °C, and 182,000 × *g* (50,000 rpm) (**b**) Sedimentation coefficient distributions *c*(*s*) derived from SV data of a titration series of 5 μM GST-BTG2 with 2.5, 5, 10, 20, 40, 80, and 160 μM PABP. The vertical dotted line indicates the *s*-value of GST-BTG2 determined in a separate experiment ($s = 4.49$ S). For better clarity of the reaction boundary, the *y*-axis was truncated at 2.1 value. To optimize absorbance measurements, centerpieces with a 12 mm optical path length were used to analyze mixtures with 2.5, 5, 10, 20, and 40 μM PABP, and centerpieces with a 3 mm optical path length for ones with 80 and 160 μM PABP. The s_w isotherm was generated using integration of the *c*(*s*) overlays from 1 to 7 S and s_{fast} isotherm using integration from 3.2 to 7 S corresponding to the reaction boundary

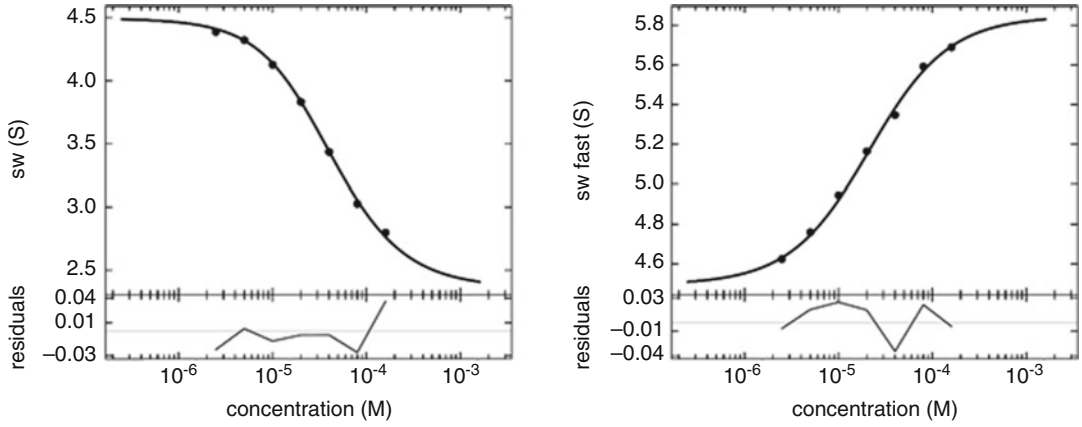


Fig. 4 Binding isotherms of sedimentation coefficients s_w and s_{fast}. The two isotherms were extracted from the sedimentation coefficient distributions shown in Fig. 3. Isotherm of s_w is derived from the integration of all *c*(*s*) peaks while s_{fast} is from only the fast *c*(*s*) peak, corresponding to the reaction boundary. The global analysis of both isotherms was performed in SEDPHAT using a hetero-association model with: $M = 80{,}860$ Da, $s = 4.48$ S, and $\varepsilon_{280nm} = 122{,}620\ M^{-1}\ cm^{-1}$ for species A (GST-BTG2) and $M = 22{,}387$ Da, $s = 2.20$ S, and $\varepsilon_{280nm} = 15{,}930\ M^{-1}\ cm^{-1}$ for species B (PABP). Starting values of $\log K_a = 5$ and $s_{AB} = 5.5$ S yielded a best-fit for K_d of 20 μM (95% confidence interval: 14–29 μM) with refined values for $s_A = 4.49$ S, $s_B = 2.36$ S, and $s_{AB} = 5.85$ S. These values are in agreement with initial fitting values and observed peak positions in the *c*(*s*) overlays (*see* **Note 11**)

OK to the sw picking mode then pick twice to define the *s*-values of the integration. In the sw isotherm constructor window, select save xp, save SEDPHAT, and fill the values of A and B concentrations for each corrected $c(s)$ distribution, the molar extinction coefficient (xt component) of A and B, and values for s_A, s_B, s_{AB}, and $\log(K_a)$. Press Save, give a name to the file, and save. In fact, 3 files with the same name but different extensions will be created, the s_w isotherm file (sw*.isotherm), the SEDPHAT experiment file (sw*.xp), and the SEDPHAT configuration file (sw*. sedphat).

4. Integrate all the superimposed corrected $c(s)$ distributions to generate EPT swfast isotherm file (isotherm of signal-weighted s-value of the reaction boundary) (Fig. 4). In GUSSI, select Integration/Make Isotherm/Hetero/EPT swfast, press OK to the EPT swfast picking mode, then pick twice to define the *s*-values of the integration of the reaction boundary. In the EPT swfast isotherm constructor window, follow the same procedure as for the sw isotherm constructor window to create swfast* files.

5. Isotherm analysis in SEDPHAT (Fig. 4). Open SEDPHAT, in Data/Load Experiment, open sw*.xp and swfast*.xp files, enter the vbar value at 20 °C, OK, then Data/Read Configuration from file, open sw*.sedphat. In the SEDPHAT A+B <-> AB model window that pops up, enter the molar masses for A and B, and select the parameters s_{AB} and $\log(K_a)$, OK. Choose Run/Global Run, if the initial values in the model are not too fare from the experimental data, Fit/Global Fit (*see* **Note 9**). Then, in Global Parameters, select s_A and s_B, then Fit/Global Fit. Evaluate the quality of the fit and the refined values for s_A, s_B, s_{AB}, and $\log K_a$ (*see* **Note 10**). Choose Statistics/Automatic confidence interval, and search w projection method to calculate the confidence intervals of $\log K_a$. Press OK in the two confidence search windows that pop up, then uncheck the $\log K_a$ parameter. Press CANCEL in the edit experimental parameters window and OK in the constrained parameter window. In the confidence search window that pops up, enter the *P* value of 0.95 and accept all the default parameters. At the end of the calculations, a message box appears with the confidence limits of $\log K_a$.

4 Notes

1. For example, the European Instruct-Eric (https://instruct-eric.eu/) and French FRISBI (http://frisbi.eu/) infrastructures provide open access to AUC instruments and expertise.

2. The effective particle theory (EPT) explains the behavior of sedimentation boundary patterns arising in rapidly interacting systems [25]. Zhao and colleagues [17] have showed that the regime (slow or fast) of the association equilibrium in SV depends on the kinetic off-rate constant (*see* Fig. 7 from 17). An equilibrium with $\log(k_{off})$ of −1 can be considered infinitely fast in the context of SV. When $\log(k_{off}) = -3$, the equilibrium is still fast but on the limit where kinetics should significantly influence the boundary pattern. A value of −3.4 corresponds to an intermediate regime, a value of −4 corresponds to the slow regime but still influenced by kinetics, and a value of −6 means an essentially infinitely slow regime on the SV time-scale.
3. On Day D—1 before SV experiment, GST-BTG2 was loaded on a gel filtration column equilibrated in SV buffer to eliminate traces of aggregates, and PABP was dialyzed overnight against the same buffer. As TCEP is not very stable in phosphate buffers, the SV buffer was prepared immediately before use. Data showed that TCEP is 100% oxidized within 72 h in PBS pH 7 and 50% oxidized within 72 h in PBS pH 8 compared to only 20% oxidized within 3 weeks in Tris or Hepes buffers in a broad range of pH.
4. The program SEDNTERP allows to calculate the molar mass, partial specific volume, and extinction coefficient of proteins from their amino acid compositions, and the solvent density and viscosity from the solvent composition using tabulated data. For unusual solvents, these latter parameters should be measured experimentally using a density-meter and a viscosity-meter.
5. Filling the sample compartment with a larger volume of sample will allow to collect a larger number of scans and therefore extract more information from the SV data. Up to 450 μL of sample can be loaded in the cell compartment with a 12 mm centerpiece. The volume of buffer has to match the volume of sample in case of interference data acquisition and be equal or higher in the case of absorbance data acquisition. There is no need of a blank buffer sector in the case of fluorescence data acquisition which allows both sectors of a double-sector centerpiece to be loaded with different samples.
6. Using the AN-50 Ti rotor, the maximum rotor speed is 50,000 rpm (201,600 × *g*). For proteins with $M < 300$ kDa, SV runs at 50,000 rpm are recommended to optimize hydrodynamic resolution. For $M > 300$ kDa, SV runs at lower speed (40,000 rpm or less) are performed to maximize the number of scans that will be collected before complete pelleting of the macromolecule [21, 26].

7. Use $P = 0.5$ for the analysis without regularization procedure. Use P value of 0.68 or 0.95 (confidence level of one or two standard deviations, respectively) for analysis with regularization. Increasing the P value will decrease the resolution of the distribution but increase the accuracy of its information. Sedimenting species that will disappear in the $c(s)$ distribution at higher P-value should be interpreted with care.
8. In the presence of a contaminant species that does not participate in the interaction, it is possible with the program GUSSI to define an s-value range that will be excluded from the calculation of the s_w values [24].
9. A poor fit may indicate that the interaction model is incorrect. Before drawing such conclusion, carefully verify there is no error in the entered parameters (molar mass, molar extinction coefficients, s_A and s_B values).
10. The refined values for s_A, s_B, and s_{AB} should be in agreement with values observed in the $c(s)$ distributions of individual components and titration series. If parameters are not well restrained by experimental data, refined values may be unrealistic. In this case, s_A and s_B should be fixed in the refinement [19].
11. It is noteworthy that GST-BTG2 (species A) is a dimer in solution, and we can expect two independent binding sites for PABP (species B). However, due to the low affinity of the complex and the concentrations used, it is not possible to determine the sedimentation coefficient and K_d for the ABB species.

Acknowledgments

The authors acknowledge the support and the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC: the platforms of the Grenoble Instruct-ERIC center (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB) and the Strasbourg Instruct-ERIC center (Centre de Biologie Intégrative, CBI) within IGBMC (CNRS UMR 7104-Inserm U 1258-Université de Strasbourg). CE also acknowledges the support of the COST action CA 15126 MOBIEU and GRAL financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).

References

1. Arthur KK, Kendrick BS, Gabrielson JP (2015) Guidance to achieve accurate aggregate quantitation in biopharmaceuticals by SV-AUC. Methods Enzymol 562:477–500
2. Correia JJ, Stafford WF (2015) Sedimentation velocity: a classical perspective. Methods Enzymol 562:49–80
3. Zhao H, Mayer ML, Schuck P (2014) Analysis of protein interactions with picomolar binding affinity by fluorescence-detected sedimentation velocity. Anal Chem 86:3181–3187
4. Kingsbury JS, Laue TM (2011) Fluorescence-detected sedimentation in dilute and highly concentrated solutions. Methods Enzymol 492:283–304
5. Schuck P (2013) Analytical ultracentrifugation as a tool for studying protein interactions. Biophys Rev 5:159–171
6. Le Roy A, Wang K, Schaack B, Schuck P, Breyton C, Ebel C (2015) AUC and small-angle scattering for membrane proteins. Methods Enzymol 562:257–286
7. Unzai S (2018) Analytical ultracentrifugation in structural biology. Biophys Rev 10:229–233
8. Salvay AG, Communie G, Ebel C (2012) Sedimentation velocity analytical ultracentrifugation for intrinsically disordered proteins. Methods Mol Biol 896:91–105
9. Tabarani G, Reina JJ, Ebel C, Vivès C, Lortat-Jacob H, Rojo J, Fieschi F (2006) Mannose hyperbranched dendritic polymers interact with clustered organization of DC-SIGN and inhibit gp120 binding. FEBS Lett 580:2402–2408
10. Dach I, Olesen C, Signor L, Nissen P, le Maire M, Møller JV, Ebel C (2012) Active detergent-solubilized H+,K+-ATPase is a monomer. J Biol Chem 287:41963–41978
11. Rocco M, Byron O (2015) Computing translational diffusion and sedimentation coefficients: an evaluation of experimental data and programs. Eur Biophys J 44:417–431
12. Chaturvedi SK, Ma J, Brown PH, Zhao H, Schuck P (2018) Measuring macromolecular size distributions and interactions at high concentrations by sedimentation velocity. Nat Commun 9:4415
13. Lebowitz J, Lewis MS, Schuck P (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci 11:2067–2079
14. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78:1606–1619
15. Dam J, Schuck P (2004) Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. Methods Enzymol 384:185–212
16. Zhao H, Ghirlando R, Alfonso C, Arisaka F, Attali I, Bain DL, Bakhtina MM, Becker DF, Bedwell GJ, Bekdemir A, Besong TMD, Birck C, Brautigam CA, Brennerman W, Byron O, Bzowska A, Chaires JB, Chaton CT, Cölfen H, Connaghan KD, Crowley KA, Curth U, Daviter T, Dean WL, Díez AI, Ebel C, Eckert DM, Eisele LE, Eisenstein E, England P, Escalante C, Fagan JA, Fairman R, Finn RM, Fischle W, de la Torre JG, Gor J, Gustafsson H, Hall D, Harding SE, Cifre JGH, Herr AB, Howell EE, Isaac RS, Jao S-C, Jose D, Kim S-J, Kokona B, Kornblatt JA, Kosek D, Krayukhina E, Krzizike D, Kusznir EA, Kwon H, Larson A, Laue TM, Le Roy A, Leech AP, Lilie H, Luger K, Luque-Ortega JR, Ma J, May CA, Maynard EL, Modrak-Wojcik A, Mok Y-F, Mücke N, Nagel-Steger L, Narlikar GJ, Noda M, Nourse A, Obsil T, Park CK, Park J-K, Pawelek PD, Perdue EE, Perkins SJ, Perugini MA, Peterson CL, Peverelli MG, Piszczek G, Prag G, Prevelige PE, Raynal BDE, Rezabkova L, Richter K, Ringel AE, Rosenberg R, Rowe AJ, Rufer AC, Scott DJ, Seravalli JG, Solovyova AS, Song R, Staunton D, Stoddard C, Stott K, Strauss HM, Streicher WW, Sumida JP, Swygert SG, Szczepanowski RH, Tessmer I, Toth RT, Tripathy A, Uchiyama S, Uebel SFW, Unzai S, Gruber AV, von Hippel PH, Wandrey C, Wang S-H, Weitzel SE, Wielgus-Kutrowska B, Wolberger C, Wolff M, Wright E, Wu Y-S, Wubben JM, Schuck P (2015) A multilaboratory comparison of calibration accuracy and the performance of external references in analytical ultracentrifugation. PLoS One 10: e0126420
17. Zhao H, Balbo A, Brown PH, Schuck P (2011) The boundary structure in the analysis of reversibly interacting systems by sedimentation velocity. Methods 54:16–30
18. Brautigam CA (2011) Using Lamm-equation modeling of sedimentation velocity data to determine the kinetic and thermodynamic properties of macromolecular interactions. Methods 54:4–15
19. Zhao H, Brautigam CA, Ghirlando R, Schuck P (2013) Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. Curr Protoc Protein Sci. Chapter 20, Unit20.12

20. Stupfler B, Birck C, Seraphin B, Mauxion F (2016) BTG2 bridges PABPC1 RNA-binding domains and CAF1 deadenylase to control cell proliferation. Nat Commun 7:10811
21. Balbo A, Minor KH, Velikovsky CA, Mariuzza RA, Peterson CB, Schuck P (2005) Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. Proc Natl Acad Sci U S A 102:81–86
22. Zhao H, Piszczek G, Schuck P (2015) SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. Methods 76:137–148
23. Zhao H, Schuck P (2015) Combining biophysical methods for the analysis of protein complex stoichiometry and affinity in SEDPHAT. Acta Crystallogr D Biol Crystallogr 71:3–14
24. Brautigam CA (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. Methods Enzymol 562:109–133
25. Schuck P (2010) Sedimentation patterns of rapidly reversible protein interactions. Biophys J 98:2005–2013
26. Demeler B (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Curr Protoc Protein Sci. Chapter 7, Unit 7.13

Chapter 10

Hands on Native Mass Spectrometry Analysis of Multi-protein Complexes

Stéphane Erb, Sarah Cianférani, and Julien Marcoux

Abstract

By maintaining intact multi-protein complexes in the gas-phase, native mass spectrometry provides their molecular weight with very good accuracy compared to other methods (typically native PAGE or SEC-MALS) (Marcoux and Robinson, Structure 21:1541–1550, 2013). Besides, heterogeneous samples, in terms of both oligomeric states and ligand-bound species can be fully characterized. Here we thoroughly describe the analysis of several oligomeric protein complexes ranging from a 16 = kDa dimer to a 801-kDa tetradecameric complex on different instrumental setups.

Key words Structural mass spectrometry, Noncovalent mass spectrometry, Oligomeric states, Stoichiometry, Subcomplexes

1 Introduction

Mass spectrometry (MS) plays a pivotal role for the characterization of multi-protein complexes. Routinely coupled with reversed-phase liquid chromatography (LC-MS), it allows purity and homogeneity assessment of biological complexes. Classical LC-MS workflow requires organic solvents (usually H_2O/acetonitrile) and acidic conditions (e.g., acetic or formic acid) to achieve best chromatographic resolutions along with optimal MS detection. However, those experimental conditions induce dissociation of multi-protein noncovalent complexes. In the early 1990s, MS approaches performed in ammonium acetate buffer maintaining quaternary structures of multi-protein complexes have been first described [1–3], opening the way for a new field of application called noncovalent or "native" MS [4, 5]. By transferring intact noncovalent assemblies into the gas phase of the mass spectrometer, native MS enables multi-protein complex stoichiometry assessment thanks to accurate mass measurements. Mostly performed on TOF and Q-TOF [6] instruments in its early years, it has recently benefited from technological improvements to reach high-

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_10,

resolution native MS [7]. Native MS is currently implemented in most structural biology departments to complement conventional structural biophysical techniques like crystallography, electron microscopy, NMR, or SAXS. The methodology is versatile as any type of multi-protein complex can be analyzed, ranging from homo-oligomeric proteins [4] to hetero-oligomers [8, 9] or even protein/nucleic assemblies [10, 11] and membrane complexes [12–14]. Latest methodological breakthroughs in the field include the direct analysis of complexes from cell lysates [15] or from native membranes [13] and the release of a new instrument enabling the complete analysis of very large complexes with unprecedented resolution [16].

In this chapter, we illustrate the possibilities of native MS on different types of multi-protein complexes and different MS platforms to assess stoichiometries. Data interpretation is usually performed by comparing the masses of the individual subunits of the complexes obtained by classical MS analysis in denaturing conditions and the mass of the intact noncovalent multi-protein assembly measured in native conditions.

2 Materials

Prepare all solutions using deionized water and analytical grade reagents. Prepare and store all reagents at 4 °C.

2.1 Desalting

Ammonium acetate: 200 mM solution in water (*see* **Note 1**). Weigh 1.54 g ammonium acetate (>98%) and transfer to a 100-mL cylinder. Add water to a volume of 95 mL. Mix and adjust pH to 7.4 with ammonium hydroxide (*see* **Note 2**). Make up to 100 mL with water.

2.2 Cesium Iodide Calibration Mix

Resuspend cesium iodide powder in 50/50 isopropanol/water (v/v) to a final concentration of 2 mg/mL.

2.3 Ammonium Hexafluorophosphate Calibration Mix

Resuspend ammonium hexafluorophosphate powder in 50/50 isopropanol/water (v/v) to a final concentration of 1 mg/mL.

2.4 GroEL Reconstitution

GroEL must be freshly prepared.

1. Prepare 9.445 mL of 1 M Tris-acetate by weighting 1.21 g of Tris-base and top up to 10 mL with acetic acid.
2. Prepare 10 mL of reconstitution buffer, composed as follows: 200 µL Tris-acetate at 1 M previously prepared, 50 µL EDTA at 100 mM, 10.7 mg magnesium acetate tetrahydrate, 5.5 mg ATP, 37.3 mg potassium chloride to reach final concentrations of 20 mM, 0.5 mM, 5 mM, 1 mM and 50 mM, respectively.

3. Prepare a 200 mM ammonium acetate solution pH 6.9 by weighting 770.8 mg in 50 mL.
4. Reconstitute GroEL (Chaperonin 60 from *Escherichia coli*, Sigma-Aldrich reference no. C7688) powder (1 mg) by adding 80 μL of refolding buffer and 20 μL of methanol directly into the vial.
5. Shake for 2 h at room temperature.
6. Transfer the entire volume into an Eppendorf, add 100 μL acetone, and agitate gently.
7. After 2 min, centrifuge for 1 min at 11,000 × *g*: a white precipitate should be present. Delete the supernatant. Then, add 200 μL of refolding buffer and agitate gently until complete dissolution.
8. Shake for 1 h at room temperature. Centrifuge for 1 min at 1,500 × *g* and take the supernatant to perform the first desalting step as explained in Subheading 3.1 (on Zeba™ spin column).
9. The second desalting step is performed as explained in Subheading 3.2 with 10 cycles (each one takes between 8 and 15 min) at 1.9 × *g* with a 50-kD MWCO membrane. Finally, check the protein concentration by UV absorbance at 280 nm. Typically, approximately 10 μM of 14-mer should be obtained ($\varepsilon = 10{,}430\ M^{-1}\ cm^{-1}$ for GroEL monomer at 280 nm).

3 Methods

3.1 Benchtop Size Exclusion Chromatography Desalting

1. Operate at +4 °C.
2. You can use 6 K MWCO Micro Bio-Spin™ 6 (Biorad) or 0.5 mL 7 K MWCO Zeba™ (Thermo Fisher Scientific) spin columns.
3. Remove the storage solution by centrifugation for 2 min at 1,000 × *g* for Micro Bio-Spin™ columns and 1 min at 1,500 × *g* for Zeba™ columns.
4. Equilibrate the columns by adding 500 μL (Micro Bio-Spin™) or 300 μL (Zeba™) of 200 mM ammonium acetate solution, centrifuge for 1 min at 1,000 × *g* (Micro Bio-Spin™) or at 1,500 × *g* (Zeba™) and discard the eluate.
5. Repeat **step 3** three times.
6. Load between 20 and 75 μL (Micro Bio-Spin™) or 30 and 130 μL (Zeba™) of sample. Loading smaller volume may affect recovery (*see* **Note 3**).

7. Place the column in a 1-mL Eppendorf tube and elute the desalted proteins by centrifuging for 4 min at 1,000 × *g* (Micro Bio-Spin™) or 2 min at 1,500 × *g* (Zeba™).

3.2 Ultrafiltration Desalting

1. Operate at +4 °C.
2. After selecting the appropriate MWCO of the ultrafiltration device, equilibrate its membrane with 500 µL of 200 mM ammonium acetate. Typical devices include Amicon™ Ultra 0.5 mL or Vivaspin™ 500 centrifugal concentrators (Merck). Centrifuge at 10,000 × *g* for 30 s, discard the eluate, and repeat three times.
3. Load your sample (up to 500 µL).
4. Concentrate down to 50 µL at 10,000 × *g*. Make up to 500 µL with 200 mM ammonium acetate and repeat five times this dilution/concentration cycle.

3.3 Direct Infusion of the Sample with the TriVersa Nanomate™ (Advion)

1. In the interface setting panel, set the controller power On and the temperature to 4 °C (*see* **Note 4**).
2. Load 10 µL of desalted protein sample to the 96- or 384-well plate.
3. In the Spray Optimization panel, set the Sample Volume to 5 µL, the gas pressure to 0.35 psi, the Voltage to apply to 1.60 kV and press "Deliver Sample" (*see* **Note 5**).

3.4 Calibration

3.4.1 TOF LCT (Upgraded for High Mass Detection by MS Vision)

1. Using the TriVersa Nanomate™ (*see* Subheading 3.3), deliver 5 µL of cesium iodide calibration mix (*see* Subheading 2.2).
2. In MassLynx MS Tune window, manually set the following parameters: Operate the instrument in positive mode with an *m/z* range from 1,000 to 10,000. Apply 120 V and 5 V for the sampling cone and extraction cone, respectively. The source temperature is set at 90 °C, but the capillary voltage and the gas flows are not used in the TriVersa Nanomate™ interface setup.
3. Adjust the backing pressure to 6 mbar (*see* **Note 6**).
4. Start a 2-min acquisition with 4-s scans.
5. Average the signal over the 2 min in the Chromatogram panel and in the Spectrum panel, select "Tools," then "Make Calibration," and select "CsI_esi.ref" profile. Adjust calibration points for the *m/z* range targeted and click on File\Save Spectrum, then click OK and finally accept the calibration after closing the window.
6. Do not forget to load this calibration file (e.g., named CsI_-date) before starting the first acquisition of the analyzed sample. To do so, click on "Acquire," then select "calibration," and load the calibration file. Finally perform the acquisition.

3.4.2 Q-TOF Synapt G2 and G2Si

1. Using the TriVersa Nanomate™ (*see* Subheading 3.3), deliver 5 μL of cesium iodide calibration mix (*see* Subheading 2.2).
2. In MassLynx MS Tune windows, manually set the following parameters: Operate the instrument in positive and sensitivity modes with an *m/z* range from 1,000 to 8,000.
3. Nanoflow+ panel: Sampling Cone: 150 V. Source Offset: 30 V. Source Temperature: 80 °C. The Capillary Voltage and the gas flows are not used in the TriVersa Nanomate™ interface setup.
4. Instrument: Trap and Transfer Collision Energy are turned off (4 V and 2 V, respectively, by default).
5. Save these parameters as an .ipr file.
6. Press "acquire" to start a 2-min TOF-MS acquisition in the 1,000–8,000 *m/z* range.
7. Average the signal over the 2 min in the Chromatogram panel and in the Spectrum panel, select Process\Automatic Peak Detection with the "Set Peak Detection Parameters Automatically" option activated.
8. Click on File\Save Spectrum and click OK.
9. In the Acquity UPLC Console, click on Intellistart, select "Create Calibration," and click Start.
10. In the calibration profile editor, select the right Mass Calibration Profile (here CsI in the 1,000–8,000 *m/z* range), right click, and reset it.
11. Right Click and edit the Mass Calibration Profile. Below the Positive Polarity chart, click on edit. Select the right Reference Compound (here CsI_1,000–8,000 (positive) with reference masses ranging from 1,172.1450 Da to 7,927.2031 Da) (*see* **Note 7**).
12. Browse to select the previous acquisition file. Click on History and select the last AccMass2 line.
13. Close the four windows by clicking OK and close the Mass Calibration Profile Editor.
14. Click Next, Next, and Start to launch the calibration.

3.4.3 Q-TOF Maxis II

1. Using the HESI source and the 500 μL Hamilton™ syringe, deliver cesium iodide calibration mix (*see* Subheading 2.2) with a flow rate of 3 μL/min.
2. In the MS Tune window, manually set the following parameters:

   Operate the instrument in positive mode with an *m/z* range from 1,000 to 8,000. Apply 20 eV and 5 eV for is CID and CE parameters, respectively. Furthermore, native experiments involve adapting transfer and prepulse storage times (*see* **Note 8**): set to 320 μs and 15 μs, respectively. Concerning the

source parameters, set the capillary voltage at 4 kV, the nebulizer at 0.4 bar, the dry gas at 4.0 L/min, and the temperature at 220 °C.

3. To perform the calibration, go to the "Calibration" tab and select the CsI reference file, then check that the calibration mode is on "Enhanced Quadratic Mode." Finally, click to "Calibrate" and accept the calibration.

3.4.4 *Exactive™ Plus EMR Orbitrap*

1. Using the HESI source and the 500 µL Hamilton™ syringe, deliver cesium iodide calibration mix (*see* Subheading 2.2) with a flow rate of 10 µL/min.
2. In the MS Tune page, manually set the following parameters:
   Operate the instrument in positive mode with an m/z range from 1,000 to 20,000. Apply 25 eV and 100 eV for the CID and CE parameters, respectively. The source temperature is set at 250 °C, the capillary voltage is set to 4 kV, and the gas flow is set to 10 u.a. The trapping gas pressure is set to 7 u.a. Set the ion optics (injection, inter, and bent flatapoles) at 4 V, the nominal resolution at 17,500 and activate the EMR mode.
3. To perform the calibration, wait for a stable Total Ion Chromatogram (<12%), then select "Calibrate," and check "EMR MS Mass Calibration (pos)".

3.4.5 *Q-Exactive™ BioPharma Orbitrap*

1. Using the syringe pusher to deliver the ammonium hexafluorophosphate calibration mix at 5 µL/min.
2. In the Calibration Panel, select "HMR Mode calibration (pos).".

3.5 *Native MS Analysis of* E. coli *RNA Polymerase on a Modified TOF LCT*

High mass detection on a modified ESI-TOF is illustrated by the analysis of *E. coli* RNA polymerase, a ~400 kDa hexameric enzyme consisting of a dimer of α subunits bound to single copies of β, β', ω, and a σ-subunits [17]. Here, a complex purified using an established procedure [18] is characterized using a simple TOF analyzer, previously optimized for the transfer of high-molecular weight species (*see* **Note 9**).

1. Desalt approximatively 100 µL of *E. coli* RNA polymerase at 30 µM in 500 mM ammonium acetate at pH 6.9 as explained in Subheading 3.2 using a Vivaspin™ 100 kDa cut-off membrane at 15,000 × *g* and +4 °C.
2. Infuse the sample at 5 µM with the TriVersa Nanomate™ (*see* Subheading 3.3).
3. Start with the same parameters as the ones used for CsI calibration but set the cone voltage to 150 V and the extraction cone to 50 V. Increase the backing pressure up to 7 mbar (*see* **Note 6**). Adjust the pressure inside the first hexapole ion guide to

0.11 mbar using argon (available for the upgraded LCT) in order to optimize the transmission of high molecular species for native MS application.

4. Start a 2-min acquisition in the 1,000–20,000 *m/z* range.
5. The corresponding smoothed full mass spectrum is presented in Fig. 1a and the corresponding deconvoluted spectrum in Fig. 1b. In order to perform this deconvolution with UniDec [19, 20] (http://unidec.chem.ox.ac.uk/), first export the spectrum list (Edit\Copy spectrum list), then paste it in a text editor and save it as a .txt file. Import the .txt file in the UniDec software. Process the data with the following parameters: *m/z* range 1,000–12,000 Th; Subtract Curved: 0; Gaussian Smoothing: 10; Bin Every: 10. Deconvolute the processed spectrum with the following parameters: Charge Range: 20–50; Mass Range: 20,000–500,000 Da; Sample Mass Every 1 Da; Peak FWHM 15 Th; Peak Shape Function: Gaussian. Detect and label the deconvoluted species with the following parameters: Peak Detection Range: 50; Peak Detection Threshold: 0.1; Peak Normalization: Max.

   As shown in Fig. 1, the pentameric $\alpha_2\beta\beta'\omega$ *E. coli* RNA polymerase is detected at the expected molecular weight (390,315 Da), together with some partially dissociated α subunit (36,509 Da).

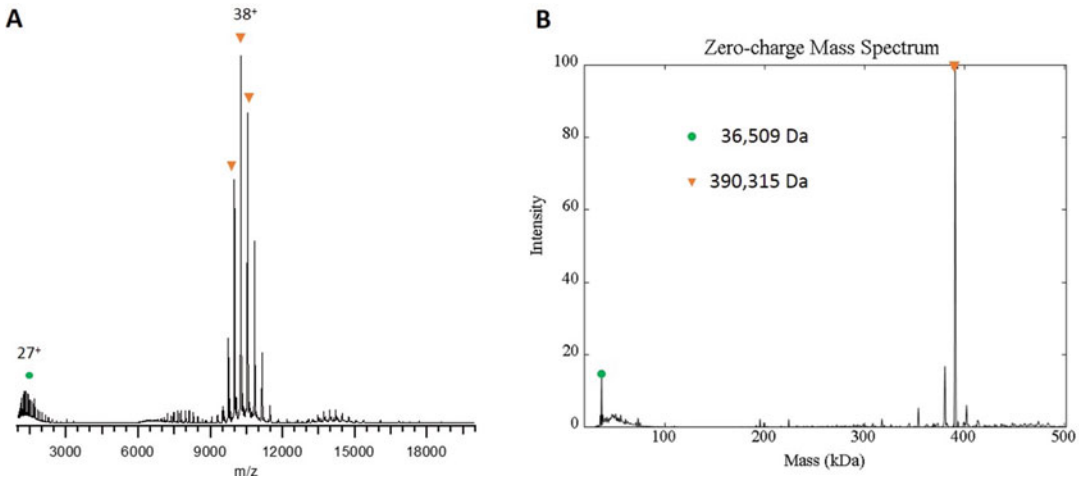


Fig. 1 Native MS analysis of the *E. coli* RNA polymerase complex. (**a**) Full MS spectrum of *E. coli* RNA polymerase complex showing charged states 35+ to 41+ in the 9,000–12,000 *m/z* range corresponding to the hetero-oligomer composed of five subunits (orange triangle) and 16+ to 33+ under *m/z* 2,500 corresponding to the α-subunit (green circle). (**b**) Nondenaturing mass spectrum of *E. coli* RNA polymerase after deconvolution with UniDec highlights the presence of the hetero-oligomeric protein (390,315 Da) and the monomeric α-subunit (36,509 Da)

3.6 Native MS Analysis of the CDK-Activating Kinase Complex on a Q-TOF Synapt G2

The use of a Q-TOF Synapt G2 platform is illustrated by the characterization of the CDK-activating kinase complex (CAK), a sub-complex of the transcription/DNA repair factor TFIIH [21, 22]. CAK is composed of a cyclin-dependent kinase (Cdk7), its associated cyclin (Cyclin H), and a third polypeptide known as MAT1, which stabilizes the Cdk7/cyclin H pair and bridges the kinase module of TFIIH to the core-complex composed of seven additional subunits. Recombinant CAK was produced in insect cells using the baculovirus expression system as described [23].

1. Desalt 20 µL of recombinant CAK at 20 µM in 200 mM ammonium acetate supplemented with 0.02% *n*-dodecyl β-D-maltoside (*see* **Note 10**), as explained in Subheading 3.1.
2. Connect the TriVersa NanomateTM as explained in Subheading 3.3.
3. In MassLynx MS Tune window manually set the following parameters:

   Operate the instrument in positive and sensitivity modes with an *m/z* range from 1,000 to 10,000 (*see* **Note 11**).

   Nanoflow+ panel: Sampling Cone: 175 V (*see* **Note 12**). Extraction cone: 5 V. Source Temperature: 90 °C.

   Instrument: Increase the backing pressure up to 6 mbar (*see* **Note 6**). Trap Collision Energy is turned on (75 V) and Transfer Collision Energy is turned off (2 V by default) (*see* **Note 13**). Trap gas flow (argon) is set to 4.5 mL/min.

   Save these parameters as an .ipr file.
4. Press "acquire" to start a 2-min TOF-MS acquisition in the 1,000–10,000 *m/z* range.
5. Analyze data.

   The corresponding averaged and smoothed (20 times) raw data is represented in Fig. 2a. Manual or software-driven (*see* Subheading 3.5, **step 5**) deconvolution of the data identifies

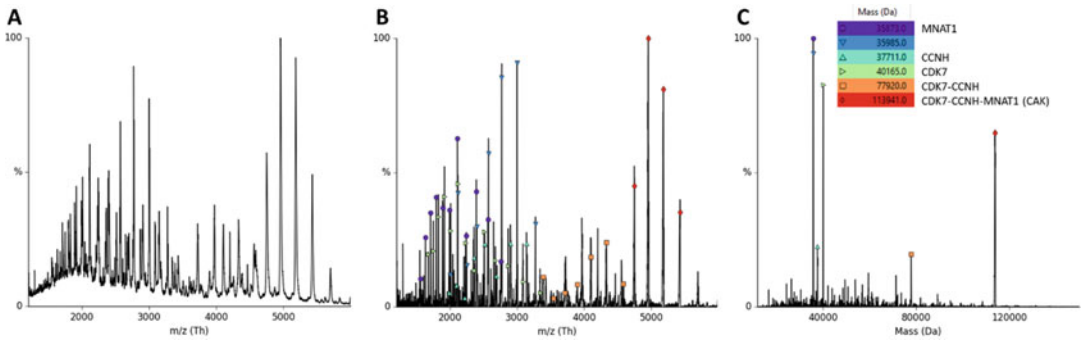


Fig. 2 Native MS spectrum of the recombinant CDK-activating kinase complex desalted on a Micro Bio-SpinTM 6 column. MS spectrum (**a**) before and (**b**) after data processing; and (**c**) after deconvolution with UniDec

different species corresponding to the three monomers of the CAK complex: namely, Cdk7 (MW: 40.1 kDa), Cyclin-H (MW: 37.7 kDa), and MNAT1 (MW: 35.9 kDa) (Fig. 2b, c). A species corresponding to the heterodimer Cdk7-Cyclin-H at 77.9 kDa is visible, confirming their direct interaction. Finally, the physiologically active heterotrimeric CAK is identified at 113.9 kDa.

3.7 Protein/DNA Complex Analysis on a Q-TOF Synapt G2

The noncovalent interaction between the Estrogen-Related Receptor DNA binding domain (ERR-DBD) and its DNA response element [24] can be monitored by direct infusion on a Synapt G2 platform. This transcription factor belongs to the steroid hormone nuclear receptor family and shares strong similarity in its DNA-binding domain (DBD) with that of the estrogen receptor (ER). In vitro, ERR binds with high affinity inverted repeat REs with a 3-bp spacing (IR3), but in vivo, it preferentially binds to single half-site response elements extended at the 5′-end by 3 bp.

1. Desalt 50 μL of ERRα DBD-BE26PSIR3 complex [24] at 28 μM on a Zeba™ spin column equilibrated with 200 mM ammonium acetate at pH 7.1 as explained in Subheading 3.1.
2. Use the TriVersa Nanomate™ (*see* Subheading 3.3) coupled to the mass spectrometer and infuse the protein complex at 5 μM.
3. In the Tune page, start with the parameters used for the CsI calibration and adapt the cone voltage at 140 V, the extraction cone at 5 V and the backing pressure at 6 mbar (*see* **Note 6**).
4. Start a 2-min acquisition in the 1,000–6,000 m/z range.
5. The average full MS spectrum obtained is presented in Fig. 3a. The deconvoluted spectrum of the protein/DNA complex (*see* Subheading 3.5, **step 5**) highlights a 1:1 stoichiometry for the protein/DNA complex (Fig. 3). The ERR-DBD binds two zinc atoms and one DNA duplex (27,892.5 Da for full-length DNA). Other species corresponding to truncated DNA associated with the protein with a 1:1 stoichiometry are also detected (27,740.6 Da and 27,586.3 Da).

3.8 SEC Native MS Coupling (ADH) on a Q-TOF Synapt G2Si

Size-exclusion chromatography (SEC) is commonly used both as a purification and for analytical purposes to monitor sample heterogeneity. SEC is often coupled with UV Absorbance, Refractive Index (RI), or Multi-Angle Laser Light Scattering (MALLS) to estimate in particular the average molecular weight and the concentration of each component of the sample. SEC can also be coupled with mass spectrometry (MS) to precisely measure the molecular weight of each separated species. This approach is illustrated here by the characterization of yeast ADH oligomers on an ACQUITY UPLC H-Class Bio coupled to a Q-TOF Synapt G2Si.

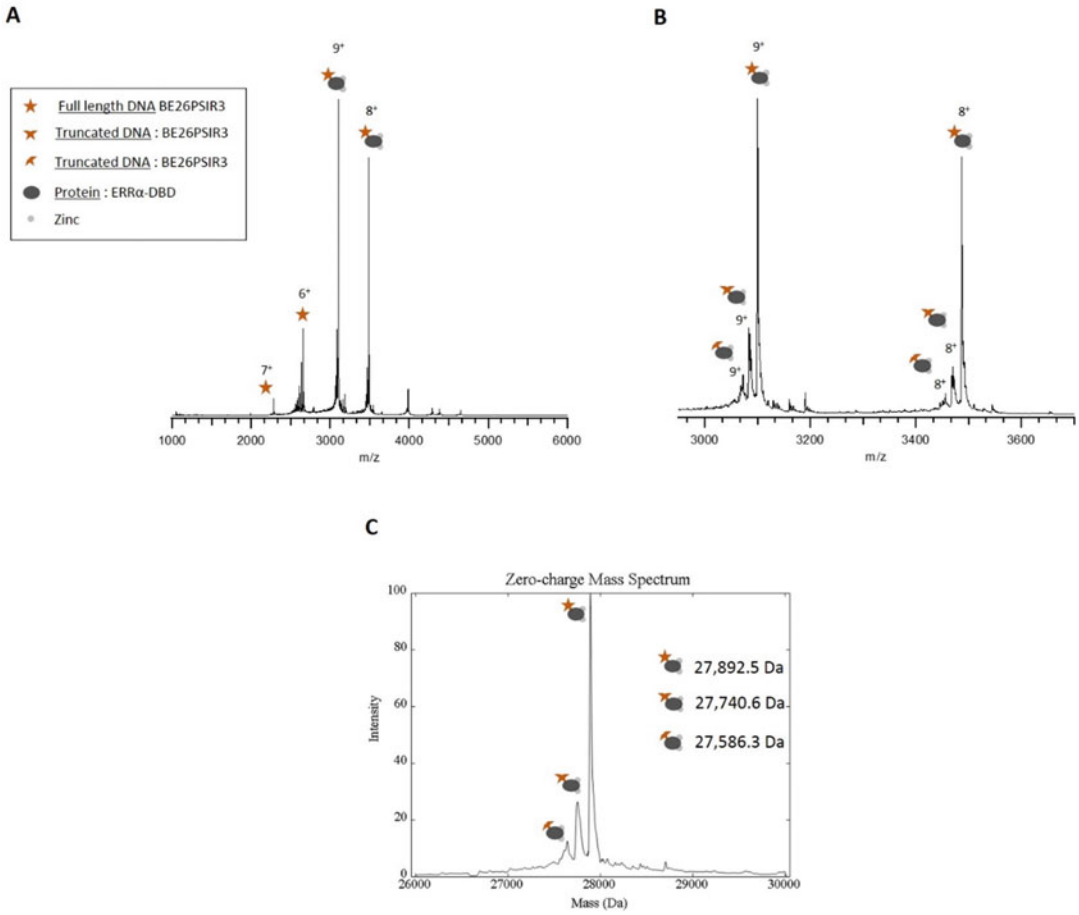


Fig. 3 Native MS analysis of the ERRα DBD-BE26PSIR3 complex. (**a**) Full MS spectrum of protein/DNA complex. Two charge states distributions in the 2,000–4,000 *m/z* range are present, first corresponding to the 6+ and 7+ charge states of DNA alone in the 2,000–4,000 *m/z* range and second to the 9+ and 8+ charge states of the protein/DNA complex in the 3,000–3,800 *m/z* range. (**b**) Zoom in the 3,000–3,600 *m/z* range showing the protein/DNA complex 9+ and 8+ charge states revealing the presence of two truncated DNA species. (**c**) The associated deconvolution of native mass spectrum with UniDec software highlights a 1:1 stoichiometry for the protein/DNA complex (27,892.5 Da) and confirms the presence of two truncated DNA species which are also able to interact with the protein (27,740.6 Da and 27,586.3 Da)

1. Connect the ACQUITY UPLC H-Class Bio to the Synapt G2Si.
2. Connect the ACQUITY UPLC Protein BEH SEC Column (200 Å, 1.7 μm, 4.6 mm × 150 mm) and equilibrate with 200 mM ammonium acetate at 300 μL/min.
3. Weigh 1.6 mg of alcohol dehydrogenase from *Saccharomyces Cerevisiae* (Sigma-Aldrich). Resuspend in 280 μL of 200 mM ammonium acetate to reach a tetrameric concentration of 40 μM.

4. In MassLynx, set up an MS file with the following parameters:
   Start Time: 0 min; End Time: 10 min; Polarity: Positive; Analyser Mode: Sensitivity; Dynamic Range: Normal; Sensitivity: Normal; Low Mass: 1,000 Da; High Mass: 10,000 Da; Scan Time: 1 s (Continuum); No Trap or Transfer CE.
5. In MassLynx, set up an Inlet file with the following parameters: Flow rate 200 μL/min with 100% solvent A (200 mM ammonium acetate) for 10 min.
6. In MassLynx MS Tune windows, manually set the following parameters:
   Nanoflow+ panel: Capillary: 3 kV; Sampling Cone: 150 V. Source Offset: 80 V. Source Temperature: 100 °C. Cone Gas: 50 L/h; Nano Flow Gas: 0 Bar; Purge Gas: 100 L/h.
   Instrument: Trap and Transfer Collision Energy are turned off (4 V and 2 V, respectively by default).
   Save these parameters as an MS tune file (.ipr file).
7. In MassLynx, create a sample list with the corresponding MS file, Inlet file, and MS tune file. Specify the sample position (Bottle) and the volume to be injected (10 μL).
8. Start the acquisition.
   The chromatogram shows two main peaks at 6.14 min and 7.32 min (Fig. 4a), corresponding to the tetramer (Fig. 4c) and monomer (Fig. 4b) of the ADH, respectively.

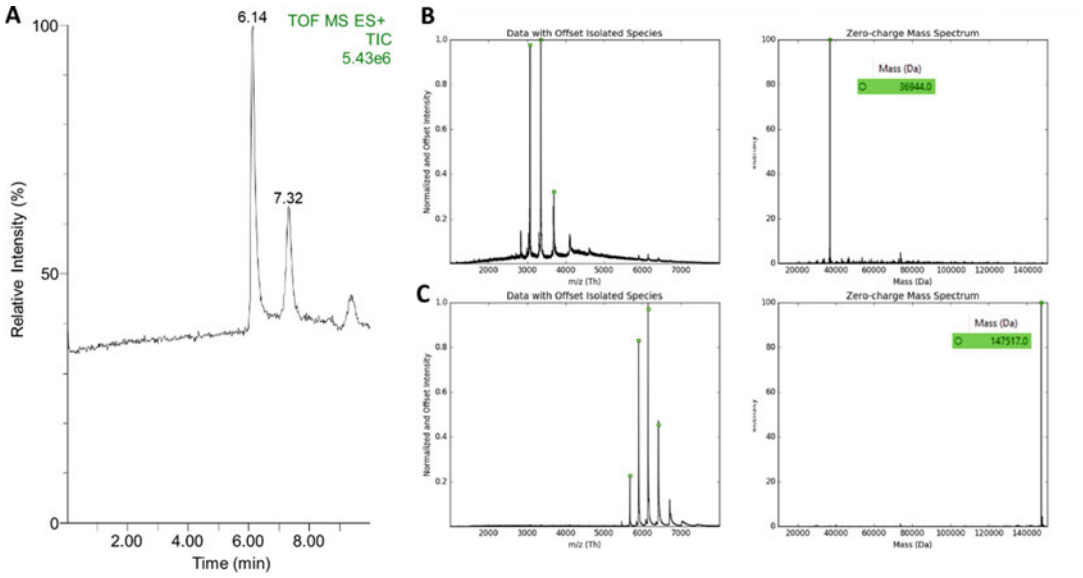


Fig. 4 SEC-Native MS analysis of *Saccharomyces cerevisiae* ADH. (**a**) Total Ion Chromatogram (TIC) corresponding to the SEC separation of the monomer and tetramer of the ADH. (**b**) Native MS spectrum of the monomer (36.9 kDa) before (left) and after (right) deconvolution with UniDec. (**c**) Native MS spectrum of the tetramer (147.5 kDa) before (left) and after (right) deconvolution with UniDec

3.9 Homo-Oligomeric Concanavalin A Analysis on a High-Resolution Q-TOF Maxis II

Concanavalin A from Jack bean is a lectin commonly used as a standard in Native MS experiments. Its physiological oligomeric state is tetrameric although higher-order oligomers can often be observed at high concentrations. We use it here to describe the parameters allowing its quaternary structure assessment with a Q-TOF Maxis II platform.

1. Resuspend 1.5 mg of Concanavalin A from Canavalia ensiformis (Jack bean, Sigma-Aldrich reference no. C5275) in 200 µL of 10 mM ammonium acetate at pH 6.8 to reach a concentration around 50 µM for the tetrameric species.
2. Desalt 100 µL of the protein (*see* Subheading 3.2) on a Vivaspin™ 10 kDa cut-off membrane (6–8 cycles at 15,000 × *g* and +4 °C during 8–10 min).
3. Connect the TriVersa Nanomate™ to the mass spectrometer. Infuse the protein at 10 µM.
4. In the tune page, use the same parameters used for CsI calibration except for the isCID that should be set at 35 eV.
5. Start a 2-min acquisition in the 1,000–8,000 *m/z* range.
6. The average smoothed mass spectrum is represented in Fig. 5a. The corresponding deconvoluted spectrum is obtained after UniDec processing (Fig. 5b, *see* Subheading 3.5, **step 5**) and allows the identification of Concanavalin A tetrameric species at 102,707 Da.

3.10 High Molecular Weight Homo-oligomeric Protein on a High-Resolution Exactive™ Plus EMR Orbitrap

The use of an Exactive Plus platform is illustrated by the characterization of the GroEL chaperone. This 57-kDa protein assembles as a 14-mer to form a double toroidal complex, which together with the GroES co-chaperonin facilitates protein folding in an ATP-dependent manner [25].

1. Use the freshly prepared and desalted GroEL protein (*see* Subheading 2.4).
2. Connect the TriVersa Nanomate™ (*see* Subheading 3.3) to the mass spectrometer. Infuse the sample concentrated at 5 µM of oligomer.
3. In the Tune page, set the following parameters: Positive mode with a 1,000–20,000 *m/z* range. Activate the EMR mode and set the trapping gas pressure at 7 u.a., the ion optics at 4 V (inter, injection, and bent flatapoles), the S-lens RF voltage at 200 V, the temperature at 250 °C, the fragmentation parameters at 25 eV and 150 eV for CID and CE values, respectively. Select a nominal resolution of 17,500.
4. Start a 2-min acquisition with a stable TIC and analyze the MS spectrum.

   The resulting averaged mass spectrum is presented in Fig. 6a (full mass range) and Fig. 6b (zoom between 11,000

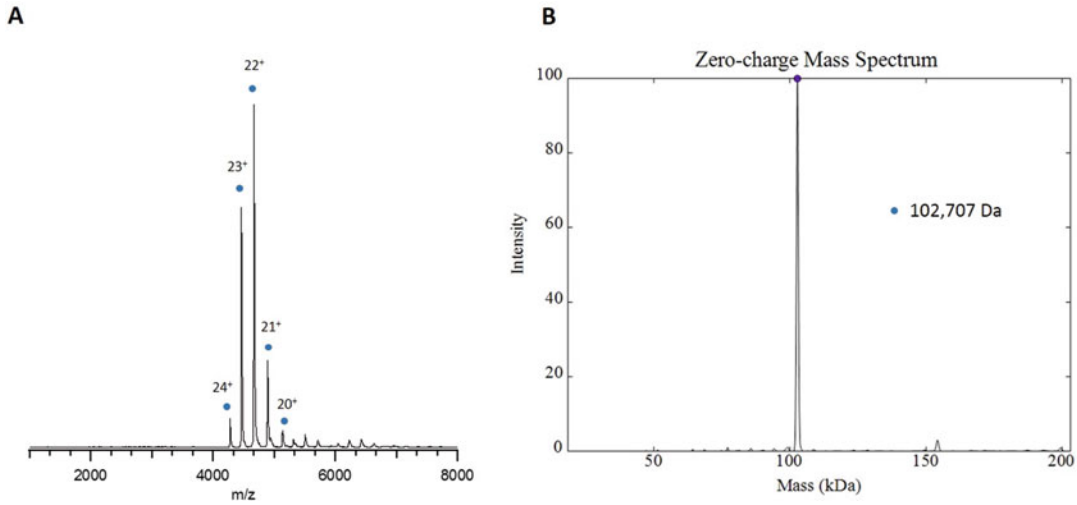


Fig. 5 Native MS analysis of Concanavalin A from *Canavalia ensiformis* protein. (**a**) Full MS spectrum of Concanavalin A showing charged states 24+ to 20+ in the *m/z* range from 4,000 to 5,500 corresponding to the homo-oligomeric protein (blue circle). (**b**) Non-denaturing mass spectrum of Concanavaline A after deconvolution with UniDec highlights the tetrameric state of the protein complex (102,707 Da)

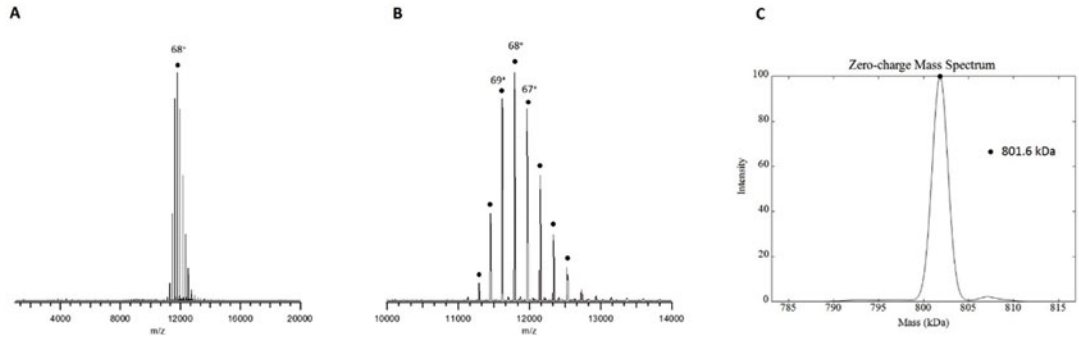


Fig. 6 Native MS analysis of Chaperonin 60 (GroEL) complex. (**a**) Full MS spectrum of GroEL complex showing a unique charge states distribution around *m/z* 12,000. (**b**) Zoom in the 10,000–14,000 *m/z* range where charge states 64+ to 71+ of the protein complex are detected. (**c**) Associated deconvolution obtained after data processing using UniDec software in the 785–815 kDa mass range, showing one species corresponding to the 14-mer homo-complex at 801.6 kDa

and 13,000 *m/z*). After data processing (*see* Subheading 3.5, **step 5**), a unique species corresponding to the 14-mer complex is identified (Fig. 6c) at 801.6 kDa.

3.11 Protein Oligomeric State Determination on a High-Resolution Q-Exactive™ Biopharma Orbitrap

The use of a Q-Exactive™ Biopharma Orbitrap platform for the determination of protein oligomeric states is exemplified by the analysis of the p8/TTD-A TFIIH subunit and yeast alcohol dehydrogenase (ADH). p8 is a 72-amino acid protein shown to form stable homodimers in solution [26]. The ADH is an enzyme required for the reduction of acetaldehyde to ethanol and is active as a tetramer [27].

3.11.1 The Trichothiodystrophy Group A Protein (TTD-A or p8)

1. Desalt 20 µL of recombinant p8 [28] at 24 µM as explained in Subheading 3.1.
2. Connect the TriVersa Nanomate™ to the Q-Exactive™ Biopharma Orbitrap as explained in Subheading 3.3.
3. In Excalibur Tune window, manually set the following parameters:

   Operate the instrument in positive mode with an m/z range from 200 to 4,000.
4. Intact protein mode: off; HRM mode: off; resolution: 120,000; Micro Scan Count: 1; Max. Ion Time: 100 ms; AGC target: 1e6, S-Lens RF Level: 40, Capillary Temperature: 275 °C, Capillary Voltage: 1.5 kV.
5. Save these parameters as an .mstune file.

   Start a 2 min acquisition in the 200–4000 m/z range and analyze data.

   The corresponding averaged raw data is represented in Fig. 7a. Manual or software-driven (*see* Subheading 3.5, **step 5**) deconvolution of the data identifies different species corresponding to the monomeric (MW = 8,272 Da) and dimeric (MW = 16,544 Da) p8 protein (Fig. 7b). A close-up view of the peak at 2,069 m/z (42,000 resolution) shows isotopic distributions corresponding to both the 4+ charged state of the monomer and the 8+ charged state of the dimer (Fig. 7a, inset). The effective resolution ranges from 35,000 (m/z 2,758) to 56,000 (m/z 1,035).

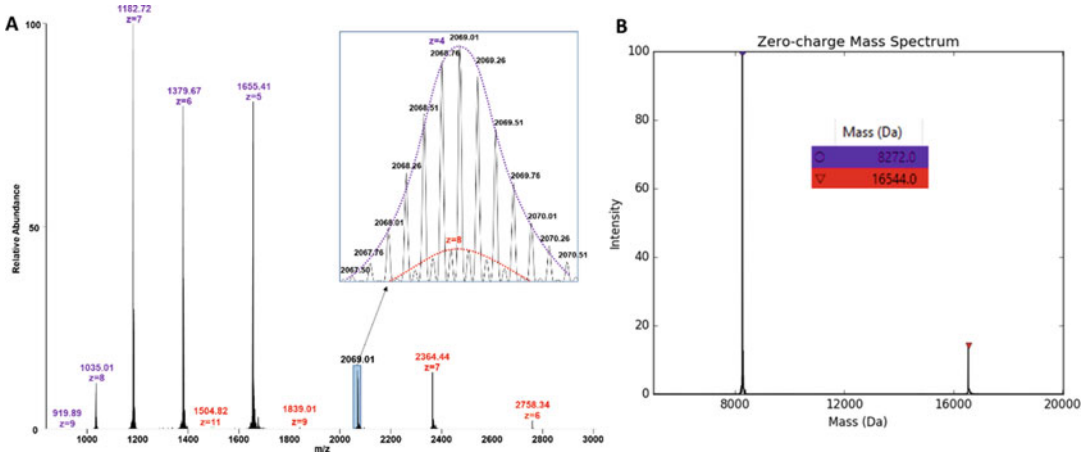


Fig. 7 Native MS analysis of the p8 subunit from the human TFIIH complex (**a**) Raw MS spectrum of p8 showing charged states 4+ to 9+ corresponding to the monomer (purple) and 6+ to 11+ corresponding to the dimer (red). The inset shows a close-up view of the 2,069.01 peak, revealing an overlap of the monomeric 4+ and dimeric 8+ charged states. (**b**) Native MS spectrum of p8 after deconvolution with UniDec, showing the presence of monomeric (8,272 Da) and dimeric (16,544 Da) species

3.11.2 The Yeast Alcohol Dehydrogenase (ADH)

1. Weigh 1.6 mg of alcohol dehydrogenase from *Saccharomyces Cerevisiae* (Sigma-Aldrich). Resuspend in 280 µL of 200 mM ammonium acetate to reach a tetrameric concentration of 40 µM.
2. Connect the TriVersa Nanomate™ to the Q-Exactive™ Biopharma Orbitrap as explained in Subheading 3.4.
3. In Excalibur Tune window, manually set the following parameters:

   Operate the instrument in positive mode with an m/z range from 2,500 to 8,000.

   Intact protein mode: off; HRM mode: on; resolution: 15,000; Micro Scan Count: 10; Max. Ion Time: 200 ms; AGC target: 3e6, S-Lens RF Level: 200, Capillary Temperature: 320 °C, Capillary Voltage: 4.4 kV.

   Save these parameters as an .mstune file.
4. Start a 2-min acquisition in the 2,500–8,000 m/z range and analyze data.

   The corresponding averaged raw data is represented in Fig. 8a. Manual or software-driven (*see* Subheading 3.5, **step 5**) deconvolution of the data identifies different species corresponding to the monomeric (MW = 36,879 Da) and tetrameric (MW = 147,516 Da) ADH protein (Fig. 8b). The effective resolution ranges from 2,000 (m/z 5,674) to 4,600 (m/z 2,844).

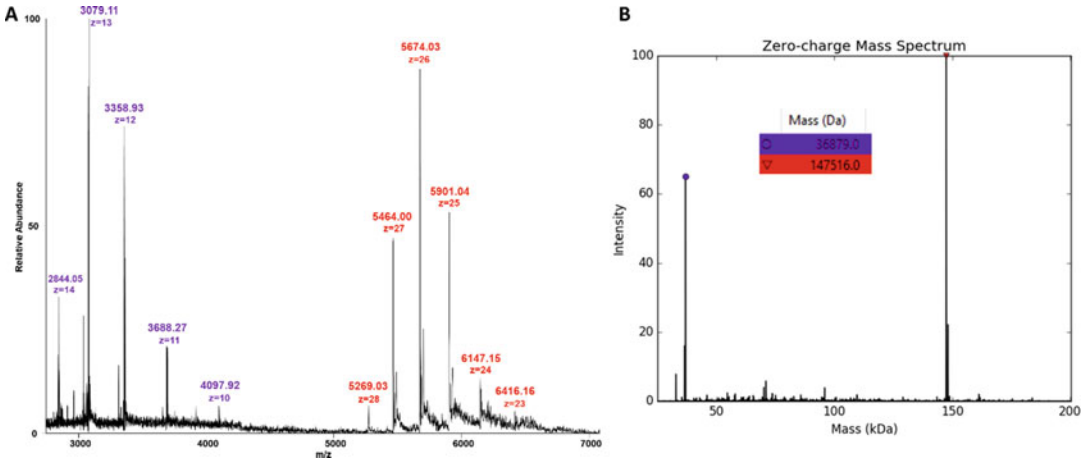


Fig. 8 Native MS analysis of *Saccharomyces cerevisiae* ADH. (**a**) Raw MS spectrum of ADH showing charged states 10+ to 14+ corresponding to the monomer (purple) and 23+ to 28+ corresponding to the tetramer (red). (**b**) Native MS spectrum of ADH after deconvolution with UniDec, showing the presence of monomeric (36,879 Da) and tetrameric (147,516 Da) species

4 Notes

1. The optimum ammonium acetate concentration is protein-dependent and can be optimized by testing concentrations ranging from 10 mM up to 1 M. If needed, volatile additives can be added, such as dithiothreitol or tris(2-carboxyethyl) phosphine for disulfide bond reduction, imidazole, some detergents (*see* **Note 10**), or metallic ions in a volatile buffer (calcium acetate, zinc acetate, etc.).
2. Volatility of the buffer is key for native MS. Beware to avoid any trace of nonvolatile salts such as sodium chloride. Ammonium acetate resuspended at 200 mM in water is usually at pH 6.9. If you want to adjust pH below 6.9, do not use HCl but formic acid or acetic acid instead. Alternatively, you can dilute a commercial 7.5 M ammonium acetate solution (Sigma). Conversely, to increase the pH, use liquid NH_3 and not NaOH.
3. For smaller volumes, 7 K MWCO Zeba™ Micro Spin columns (2–12 μL, Thermo Fisher Scientific) can be used.
4. Do not forget to turn off the cooling of the plate after use of the nanomate. Otherwise, condensation will occur, water will fill the wells, and the plate will have to be discarded the next day.
5. The gas pressure and voltage to apply are the main two parameters to adjust. The stability of the spray can be visualized in the Spray Optimization window, by pressing "Graph." If the intensity of the electrospray is dropping, incrementally increasing the gas pressure (no more than 0.5 psi) or the Voltage to apply (no more than 1.80 kV) might help. If the spray is not coming back, use another nozzle by pressing "Next Nozzle." Reducing the gas pressure and Voltage to apply while spraying usually results in spray loss. The optimum Voltage to apply is protein-dependent and has to be carefully optimized, as higher values usually increase the ion transmission (more intense signal) and buffer removal (better resolved signal) but might result in complex dissociation. Ticking "Return unused sample to tray" enables to recover most of the sample after data acquisition.
6. The backing pressure is a key parameter on Waters Instruments (LCT, Synapt), which needs to be optimized for each protein complex. Generally, this value is set around 6 mbar and can be adjusted if necessary. This is possible by adding a homemade system with two external valves that reduce the suction of the primary pump.

7. Reference files for higher *m/z* ranges can be custom made. In the case of CsI, increments of +259.81 Da can be directly added to the initial CsI calibration file.
8. In order to optimally accumulate and transfer heavy ions, transfer time and prepulse storage time parameters should be increased compared to small ions analysis. These two parameters are linked, but the transfer time has to be higher than the prepusle storage time.
9. Waters TOF or Q-TOF mass spectrometers can be upgraded (MS Vision, the Netherlands) by modifying electronics and gas pressure in order to improve the transmission of high molecular mass species in non-denaturing conditions. On our upgraded LCT, the combination of an additional gas arrival (argon) in the first hexapole housing through a needle valve and the modification of different electronic parts (e.g., higher voltages applied to sample and extraction cones) results in an improvement of high molecular mass transmission. Typically, the sensitivity for GroEL14 analysis is increased sixfold by argon admission if the pressure is set around 4×10^{-2} mbar. A micro-Pirani gauge is installed to control the pressure applied.
10. *n*-Dodecyl β-D-maltoside can be used at two times the critical micellar concentration (0.02%) in order to stabilize very labile soluble complexes. It can also be used at the same concentration to spray membrane proteins [29, 30].
11. The *m/z* range on Waters Q-TOF depends on the type of Quadrupole that is installed: up to 16,000 *m/z* for a 4,000 *m/z* quadrupole, up to 32,000 *m/z* for an 8,000 *m/z* quadrupole and up to 128,000 *m/z* for a 32,000 *m/z* quadrupole.
12. The sample cone is a key parameter on Waters Instruments (LCT, Synapt, etc.) that needs to be optimized for each protein complex. Similarly, to the voltage applied to the nozzle of the TriVersa Nanomate™, higher values usually increase the intensity of the signal and provide better resolved peaks by removing residual buffer. However, high sample cone values might result in complex dissociation. Here again, an optimum value has to be found that provides good signal without undermining the multiprotein complex integrity. A good strategy is to start with a maximum Sampling Cone voltage (150 V) and progressively reduce it by 10-V steps.
13. The trap collision energy is another key parameter on Waters Q-TOF. In the case of membrane proteins or very large multiprotein complexes (>500 kDa), increasing the trap collision energy might be helpful or even mandatory in order to thermodynamically cool down the ions (large complexes) or

remove the remaining detergent molecules. Simultaneously and incrementally increasing the Trap gas flow (from 2 mL/min up to 8 mL/min) might improve the transmission of these particular large or membrane protein complexes.

Acknowledgments

The authors would like to thank Arnaud Poterszman (CAK complex), Albert Weixlbaumer (*E. coli* RNA polymerase), and Isabelle Billas Massobrio (ERR complexes) from the IGBMC (Strasbourg, France) and Virginie Gervais and Alain Milon (p8 protein) from IPBS (Toulouse, France) for kindly providing samples. This work was supported by the French Ministry of Research (Investissements d'Avenir Program, Proteomics French Infrastructure, ANR-10-INBS-08) and the FEDER (Fonds Européens de Développement Régional), Toulouse Métropole, and the Région Midi-Pyrénées.

References

1. Ganem B (1991) Detection of noncovalent receptor-ligand complexes by mass spectrometry. J Am Chem Soc 113:6294–6296
2. Ganem B (1991) Observation of noncovalent enzyme-substrate and enzyme-product complexes by ion-spray mass spectrometry. J Am Chem Soc 113:7818–7819
3. Katta V, Chait B (1991) Observation of the hemoglobin complex in native myoglobin by electrospray-ionization mass spectrometry. J Am Chem Soc 113:8534–8535
4. Nettleton EJ, Sunde M, Lai Z, Kelly JW, Dobson CM, Robinson CV (1998) Protein subunit interactions and structural integrity of amyloidogenic transthyretins: evidence from electrospray mass spectrometry. J Mol Biol 281 (3):553–564
5. Pinkse MWH, Heck AJR, Rumpel K, Pullen F (2004) Probing noncovalent protein-ligand interactions of the cGMP-dependent protein kinase using electrospray ionization time of flight mass spectrometry. J Am Soc Mass Spectrom 15(10):1392–1399
6. Sobott F, Hernández H, McCammon MG, Tito MA, Robinson CV (2002) A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. Anal Chem 74(6):1402–1407
7. Rose RJ, Damoc E, Denisov E, Makarov A, Heck AJ (2012) High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. Nat Methods 9(11):1084–1086
8. Andersson FI, Tryggvesson A, Sharon M, Diemand AV, Classen M, Best C, Schmidt R, Schelin J, Stanne TM, Bukau B, Robinson CV, Witt S, Mogk A, Clarke AK (2009) Structure and function of a novel type of ATP-dependent Clp protease. J Biol Chem 284(20):13519–13532
9. Sharon M, Mao H, Boeri Erba E, Stephens E, Zheng N, Robinson CV (2009) Symmetrical modularity of the COP9 signalosome complex suggests its multifunctionality. Structure 17 (1):31–40
10. Martinez-Zapien D, Saliou JM, Han X, Atmanene C, Proux F, Cianférani S, Dock-Bregeon AC (2015) Intermolecular recognition of the non-coding RNA 7SK and HEXIM protein in perspective. Biochimie 117:63–71
11. Saliou JM, Manival X, Tillault AS, Atmanene C, Bobo C, Branlant C, Van Dorsselaer A, Charpentier B, Cianférani S (2015) Combining native MS approaches to decipher archaeal box H/ACA ribonucleoprotein particle structure and activity. Proteomics 15(16):2851–2861
12. Barrera NP, Di Bartolo N, Booth PJ, Robinson CV (2008) Micelles protect membrane complexes from solution to vacuum. Science 321 (5886):243–246
13. Chorev DS, Baker LA, Wu D, Beilsten-Edmands V, Rouse SL, Zeev-Ben-Mordehai T, Jiko C, Samsudin F, Gerle C,

Khalid S, Stewart AG, Matthews SJ, Grünewald K, Robinson CV (2018) Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. Science 362(6416):829–834

14. Marcoux J, Robinson CV (2013) Twenty years of gas phase structural biology. Structure 21 (9):1541–1550
15. Gan J, Ben-Nissan G, Arkind G, Tarnavsky M, Trudeau D, Noda Garcia L, Tawfik DS, Sharon M (2017) Native mass spectrometry of recombinant proteins from crude cell lysates. Anal Chem 89(8):4398–4404
16. van de Waterbeemd M, Tamara S, Fort KL, Damoc E, Franc V, Bieri P, Itten M, Makarov A, Ban N, Heck AJR (2018) Dissecting ribosomal particles throughout the kingdoms of life using advanced hybrid mass spectrometry methods. Nat Commun 9 (1):2493
17. Opalka N, Brown J, Lane WJ, Twist KA, Landick R, Asturias FJ, Darst SA (2010) Complete structural model of *Escherichia coli* RNA polymerase from a hybrid approach. PLoS Biol 8(9):e1000483
18. Twist KA, Husnain SI, Franke JD, Jain D, Campbell EA, Nickels BE, Thomas MS, Darst SA, Westblade LF (2011) A novel method for the production of in vivo-assembled, recombinant *Escherichia coli* RNA polymerase lacking the α C-terminal domain. Protein Sci 20 (6):986–995
19. Marty MT, Baldwin AJ, Marklund EG, Hochberg GK, Benesch JL, Robinson CV (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem 87 (8):4370–4376
20. Reid DJ, Diesing JM, Miller MA, Perry SM, Wales JA, Montfort WR, Marty MT (2019) MetaUniDec: high-throughput deconvolution of native mass spectra. J Am Soc Mass Spectrom 30(1):118–127
21. Radu L, Schönwetter E, Braun C, Marcoux J, Kölmel W, Schmitt DR, Kuper J, Cianferani S, Egly JM, Poterszman A, Kisker C (2017) The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH. Nucleic Acids Res 45(18):10872–10883
22. Kolesnikova O, Radu L, Poterszman A (2019) TFIIH: a multi-subunit complex at the crossroads of transcription and DNA repair. Adv Protein Chem Struct Biol 115:21–67
23. Fouillen L, Abdulrhaman W, Moras D, van Dorsselaer A, Poterszman A, Sanglier-Cianférani S (2010) Analysis of recombinant phosphoprotein complexes with complementary mass spectrometry approaches. Anal Biochem 407 (1):34–43
24. Mohideen-Abdul K, Tazibt K, Bourguet M, Hazemann I, Lebars I, Takacs M, Cianférani S, Klaholz BP, Moras D, Billas IML (2017) Importance of the sequence-directed DNA shape for specific binding site recognition by the estrogen-related receptor. Front Endocrinol (Lausanne) 8:140
25. Horwich AL, Fenton WA, Chapman E, Farr GW (2007) Two families of chaperonin: physiology and mechanism. Annu Rev Cell Dev Biol 23:115
26. Vitorino M, Coin F, Zlobinskaya O, Atkinson A, Moras D, Egly JM, Poterszman A, Kieffer B (2007) Solution structure and self-association properties of the p8 TFIIH subunit responsible for Trichothiodystrophy. J Mol Biol 368(2):473–480
27. Raj SB, Ramaswamy S, Plapp BV (2014) Yeast alcohol dehydrogenase structure and catalysis. Biochemistry 53(36):5791–5803
28. Gervais V, Muller I, Mari PO, Mourcet A, Movellan K, Ramos P, Marcoux J, Guillet V, Javaid S, Burlet-Schiltz O, Czaplicki G, Milon A, Giglia-Mari G (2018) Small molecule-based targeting of TTD-A dimerization to control TFIIH transcriptional activity represents a potential strategy for anticancer therapy. J Biol Chem 293(39):14974–14988
29. Marcoux J, Wang S, Politis A, Reading E, Ma J, Biggins P, Zhou M, Tao H, Zhang Q, Chang G, Morgner N, Robinson CV (2013) Mass spectrometry reveals synergistic binding of nucleotides, lipids and drugs to a multidrug resistance efflux pump. Proc Natl Acad Sci 110 (24):9704–9709
30. Laganowsky A, Reading E, Hopper JT, Robinson CV (2013) Mass spectrometry of intact membrane protein complexes. Nat Protoc 8 (4):639–651

Chapter 11

Studying Protein–DNA Interactions by Hydrogen/Deuterium Exchange Mass Spectrometry

Ruzena Filandrova, Daniel Kavan, Alan Kadek, Petr Novak, and Petr Man

Abstract

Protein hydrogen/deuterium exchange (HDX) coupled to mass spectrometry (MS) can be used to study interactions of proteins with various ligands, to describe the effects of mutations, or to reveal structural responses of proteins to different experimental conditions. It is often described as a method with virtually no limitations in terms of protein size or sample composition. While this is generally true, there are, however, ligands or buffer components that can significantly complicate the analysis. One such compound, that can make HDX-MS troublesome, is DNA. In this chapter, we will focus on the analysis of protein–DNA interactions, describe the detailed protocol, and point out ways to overcome the complications arising from the presence of DNA.

Key words DNA, Hydrogen/deuterium exchange, Protein–DNA binding, Structural mass spectrometry, Transcription factor

1 Introduction

Structural proteomics is a rapidly developing field focusing specifically on protein structural dynamics and characterization of the architecture of protein macromolecular assemblies. Hydrogen/deuterium exchange (HDX) ranks as one of the most versatile and prominent methods in this area. Its biggest advantages are not being limited by protein size and the ability to study proteins in their truly native-like environment, at virtually any pH, buffer composition, protein concentration, or temperature (*see* **Note 1**). In a typical setup, an HDX experiment starts by the dilution of a protein into a deuterated buffer. At selected time points, aliquots are taken, and the exchange is quenched (nearly stopped) by rapid acidification to pH 2.5 and by lowering the temperature to 0 °C. Subsequently, each sample is digested by acidic, nonspecific proteases, and deuteration of the generated peptides is measured by mass spectrometry (MS) [1–3].

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247,
https://doi.org/10.1007/978-1-0716-1126-5_11,

Albeit very powerful and usually straightforward, HDX-MS has also some disadvantages, e.g., rather limited spatial resolution and need for sample dilution as the experiment often starts with ten-fold dilution into a deuterated buffer. However, due to the coupling to sensitive mass spectrometric detection, micro- and submicromolar concentrations are easily managed. Although HDX in combination with mass spectrometry was introduced more than two decades ago and there are hundreds of research articles describing its uses to study protein folding/unfolding, protein–protein and protein–ligand interactions [4, 5], the number of reports utilizing HDX for protein–nucleic acid interaction is rather limited. This is most likely due to the nature of nucleic acids that complicate HDX-MS analyses. During the quenching step (lowering the pH) in an HDX-MS workflow, the DNA backbone becomes protonated, which often results in its poor solubility and precipitation [6]. The protein interacting with DNA then tends to precipitate as well. The presence of DNA can also have adverse effects on chromatographic separation. Such behavior, however, depends on the DNA composition and length. Therefore, reports dealing with larger DNA stretches always utilize technical tricks to deal with the troublesome DNA while those using very short oligonucleotides may sometimes be spared these adverse effects.

The first ever attempt to use HDX for protein–nucleic acid interaction study enabled the determination of a protein–DNA complex dissociation constant by a combination of HDX and matrix-assisted laser desorption/ionization mass spectrometry [7]. However, it took five more years until Sperry et al. in 2008 truly successfully probed structural features of protein–DNA interactions of two biological systems (a protein binding to a telomeric oligonucleotide and the protease thrombin interacting with an aptamer). These studies were successfully accomplished thanks to the incorporation of strong anion exchange trap column prior to the liquid chromatography-mass spectrometric (LC-MS) analysis to selectively trap and remove the oligonucleotides [8, 9]. Besides an ion exchanger, the use of protamine sulfate has also been reported as beneficial in the HDX-MS analysis of oligonucleotide-containing samples. Poliakov and coworkers concluded that the protein–DNA co-precipitation upon the lowering of pH is primarily of electrostatic origin. Thus, inclusion of small, highly basic, and hence, positively charged protamine improved protein recovery [10]. Alternatively, in another study, Roberts et al. counteracted the negative effect of DNA on the formation of peptides, solubility, and chromatographic separation in HDX by tuning the concentration of a positively charged denaturing agent (guanidine hydrochloride) in their quench buffer [11]. Use of another denaturant has recently been exploited by Graham et al. to explain structural and mechanistic aspects of DNA unwinding when guanidine

hydrochloride was replaced by urea, and an alternative aspartic protease was used in the quench buffer [12].

In this study, drawing upon the experience gained through our own work, we present an experimental protocol for the characterization of protein interactions with DNA (in the form of an oligonucleotide) as used in our group. We have repeatedly observed that careful optimization and proper control over the quality of the DNA entering an HDX experiment (amount of dsDNA formed, purity, salt concentration) are crucial for boosting its ability to bind to the interaction partner. Ensuring as strong an interaction as possible in this manner then sometimes largely mitigates the problems stemming from the DNA presence, as the amount of oligonucleotides used can be kept to a minimum (optimally down to 1:1 molar ratio to the protein), while still providing virtually complete saturation of the transcription factor (TF). Sometimes, however, even the best quality of sample is not enough, and the aforementioned techniques for coping with DNA come into play.

In the following set of protocols, we describe ways to perform protein–DNA HDX experiments, together with some of the potential pitfalls, with emphasis on the interaction of transcription factors with their DNA-response elements. As we have observed that protein:DNA systems can behave surprisingly differently even when the proteins as well as DNAs used are similar in size, we will be using two example systems. Forkhead box protein O4 (FOXO4)/DAF16 [13, 14] serves as an example where short dsDNA (13 bp) does not interfere significantly with peptide recovery, and only digestion conditions are optimized to reach good HDX spatial resolution. On the other hand, the TEA domain family member 1 (TEAD1)/M-CAT [15] represents the opposite situation when a 15-bp long oligonucleotide has a strong impact on peptide recovery. These striking differences in behavior stress the importance of having a vast array of conditions, proteases, and technical tricks at one's disposal to fine-tune the HDX protocol. The procedures we use, as described herein, should provide prospective users with a set of tools to successfully optimize conditions for HDX and perform their own analyses of a range of TF–DNA complexes.

2 Materials

1. Protein to be analyzed (*see* **Note 2**).
2. Forward and reverse DNA strands (*see* **Note 3**).
3. Buffer suitable for the complex (e.g., 20 mM HEPES, pH 7.4, 150 mM NaCl).
4. Gel casting tray and electrophoresis apparatus (Bio-Rad).

5. Non-denaturing acrylamide gel (12%)—for 10 ml mix the following: 6 ml H_2O, 1 ml 10× concentrated TBE buffer, 80 µl 10% ammonium persulfate (APS), 8 µl TEMED (*N*,*N*,*N*′,*N*′-tetramethylethylenediamine), 3 ml 40% acrylamide:*N*,*N*′-methylenebisacrylamide (1:19 mix).
6. DNA loading dye (Thermo Fisher).
7. Protein staining solution (45% methanol, 10% acetic acid, 0.25% Coomassie Brilliant Blue R250).
8. TBE buffer (90 mM Tris–Cl, 90 mM boric acid, 2 mM EDTA, pH 8.3).
9. Fluorescent DNA stain GelRed (Thermo Fisher).
10. Deuterium oxide (D_2O).
11. Quench buffer (500 mM glycine in LC-MS grade water, pH 2.3 set with concentrated HCl).
12. Urea or Guanidine hydrochloride.
13. Liquid nitrogen.
14. Peltier cooled box or a polystyrene box filled with ice and water mixture (*see* **Note 4**).
15. Two 6-port valves, one of them with injection needle port (Rheodyne/IDEX).
16. PEEK (polyether ether ketone) tubing (OD 1/16″, ID—250 µm and 750 µm) and stainless steel tubing (OD 1/16″, ID—125 µm).
17. UPLC (ultrahigh-performance liquid chromatography) fittings with ferrules (1/16″).
18. Hamilton syringe with blunt end needle (100 µl).
19. Timer.
20. Columns with immobilized acidic protease—pepsin, rhizopuspepsin (protease type XVIII), nepenthesin-1, nepenthesin-2, or aspergillopepsin (protease type XIII) immobilized on POROS™ 20AL (Thermo Fisher Scientific) using previously described [16, 17] protocols and packed into stainless steel guard columns (2 mm × 20 mm or 1 mm × 20 mm, IDEX).
21. Desalting trap column—VanGuard Pre-column (ACQUITY UPLC BEH C18, 130 Å, 1.7 µm, 2.1 mm × 5 mm, Waters).
22. Reversed-phase analytical column (ACQUITY UPLC BEH C18, 130 Å, 1.7 µm, 1 mm × 100 mm, Waters).
23. Loading and desalting pump (1260 Infinity II Quaternary pump (Agilent Technologies)).
24. Gradient pump (1290 series, Agilent Technologies).
25. LC loading solvent: 0.4% formic acid in water.

26. LC solvent A (0.4% formic acid, 2% acetonitrile, 97.6% water).
27. LC solvent B (0.4% formic acid, 95% acetonitrile, 4.6% water).
28. Electrospray ionization (ESI) mass spectrometer (any ESI-equipped MS with reasonable resolution is usable; in our case, we used an electrospray ionization—Fourier transform ion cyclotron resonance 15T solariX XR, Bruker Daltonics).
29. Software: DataAnalysis 4.1 (Bruker Daltonics), ProteinScape 4 (Bruker Daltonics), DeutEx (http://peterslab.org/downloads.php under HDX tools section), mMass (http://mmass.org), PyMol (https://pymol.org/; v 2.1.0), MSTools (http://peterslab.org/MSTools/).

3 Methods

3.1 Verification of Protein–DNA Complex Formation by a Gel Shift Assay

Prior to the HDX-MS experiment, it is advisable to verify whether the DNA duplex is present and that the complex between protein and ligand (here DNA) is indeed formed upon mixing of the components. In this respect, knowledge of the binding constants offers an advantage as the experimental setup can be adjusted to achieve full protein occupancy (complex presence) during the labeling step [18]. Here, we use gel shift assay to check DNA duplex formation and its binding to the studied protein.

1. Dilute both DNA strands in LC-MS purity water and mix them at equimolar ratio.
2. Heat the mixture at 95 °C for 1 min, and let it slowly cool down to room temperature.
3. Dilute/buffer exchange the protein into a buffer suitable for the complex which is later used for H/D exchange (20 mM HEPES, pH 7.4, 150 mM NaCl, *see* **Note 5**) to a concentration at which full occupancy of the protein by DNA is achieved (*see* **Note 6**).
4. Add equimolar amounts of dsDNA to the protein and let the mixture equilibrate for at least 2 min at a temperature where both the dsDNA and protein are stable (*see* **Note 7**).
5. Prepare following samples: 200 ng of both single-stranded oligonucleotides, 200 ng of free dsDNA, 5 μg of free protein, and the equimolar protein–DNA complex equal to 5 μg of the protein (*see* **Note 8**). All components must be prepared in the same buffer.
6. Add loading buffer to all samples.
7. Prepare non-denaturing polyacrylamide gel by mixing all components except APS. Add APS just before pouring the mixture into a gel casting tray (*see* **Note 9**).

8. Load the samples and run the electrophoresis in TBE buffer at a constant voltage of 50 V for one gel until the loading dye reaches half of the gel (*see* **Note 10**).
9. Stain the gel with any kind of fluorescent dye for DNA staining, take a picture, and stain the gel again in protein staining solution to visualize the protein. Only proteins in complex with DNA will migrate into the gel. Free protein will normally not separate as TFs are generally positively charged and interact strongly with negative DNA.
10. Formation of DNA duplex and protein–DNA complex are manifested as shifts in the electrophoretic mobility (Fig. 1). Compare the results from both the staining methods to confirm the complex formation—position of the complex band should be the same for protein and DNA staining.
11. If multiple complex concentrations are tested, select the one where there is the least free dsDNA (of the same gel mobility as dsDNA control) visible.

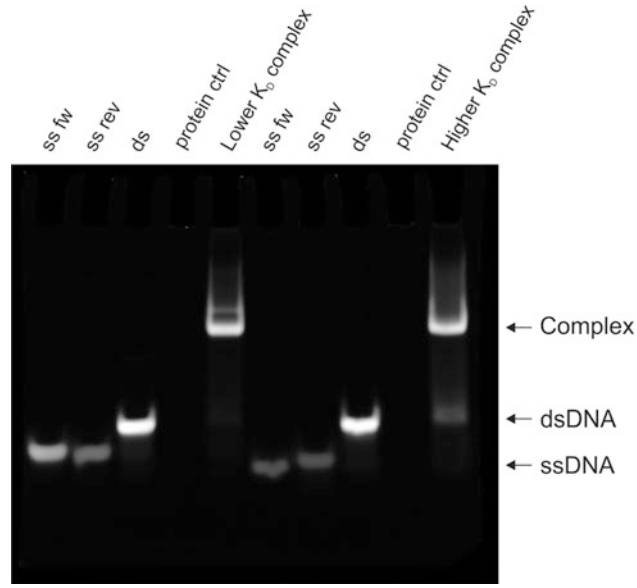


Fig. 1 Gel shift assay. Two oligonucleotides (DNA1 and DNA2) with different binding affinities are shown. Shifts in electrophoretic mobility in the third and fifth lane from the left indicate duplex DNA and protein–DNA complex formation, respectively. Protein control lane stayed empty since no DNA was present. In both the complex lanes, a band of the same mobility as dsDNA can be observed, though in the "higher K_D complex" lane, this band is much thicker, indicating lower bound fraction of the protein

3.2 Optimizing Digestion Conditions

Before starting the H/D exchange, it is crucial to verify how the protein is digested under HDX-MS compatible conditions. Different aspects to be considered are protein sequence coverage, peptide length, and peptide redundancy but also whether the protein remains soluble after quenching and freezing and how the presence of DNA affects the whole procedure. The primary goal is to ideally obtain full sequence coverage with peptides providing good spatial resolution, i.e., not very short or long ones. Recommendations regarding optimal length do vary slightly in the literature, but fragments between 8 and 12 amino acids probably represent the best standard. Besides the spatial resolution obtainable through the generated peptides, one can also target nearly amino acid resolution using suitable fragmentation techniques like electron transfer dissociation [19] or UV photodissociation [20] (*see* **Note 11**). In addition, emphasis should also be put on the generation of large numbers of overlapping peptides as these can further increase the spatial resolution [21, 22] and/or provide higher confidence in the observed changes (*see* **Note 12**). To fulfill all these goals, different proteases, ideally immobilized and packed into a column, should be tested [17, 23–26]. The columns can be used individually or combined in serial or parallel setting [27]. Column size as well as the protease density on the POROS resin, which together dictate the enzyme:protein ratio, can be varied. Other factors affecting the digestion are flow, temperature, pressure, and the use of denaturing agents. Time spent on the proteolytic column is primarily determined by the desalting LC flow. Therefore, higher flow leads to faster digestion and produces longer fragments and higher redundancy (number of overlapping peptides). Slower flow, on the other hand, produces more complete digestion, which however may at times be even detrimental, if the peptides produced are too short. Typical digestion temperature should be close to 0 °C to minimize deuterium-loss (back-exchange), but it can be raised to 15 °C or even 20 °C locally to increase the digestion efficiency. This is usually achieved either in a dedicated separately temperature-controlled chamber for the protease column or by placing the column out of the ice/water bath. However, when tweaking these parameters, one should always keep in mind that optimal balance between efficient digestion and minimal H/D back-exchange conditions must be maintained (hence running the analysis as fast as possible and at the lowest possible temperature). For proteins not offering satisfactory digestion under mild denaturing conditions provided by the acidic environment of the quench buffer alone, addition of denaturing agents is suggested. Urea or guanidine can be used at quite high concentrations with all the immobilized proteases; however, besides rendering the target protein susceptible to digestion, these agents also affect the protease itself as well as peptide binding to the reversed phase resin in the trap column. While guanidine was shown to have mainly a detrimental effect, urea can even enhance

proteolytic activity [17]. In addition, guanidine is difficult to remove completely during the subsequent desalting step, and thus, 4 M urea (final concentration upon quenching) may always be suggested as the first choice.

When analyzing samples containing nucleic acids, specific problems may additionally arise upon the quenching of H/D exchange, as mentioned above. As the pH drops, nucleic acids tend to precipitate, and it is very likely that the protein of interest will coprecipitate as well. Therefore, several strategies described in the Introduction were developed to remove the nucleic acids after quenching or to reduce their detrimental effect on the HDX-MS workflow.

1. Prepare the duplex DNA according to the **steps 2** and **3** of the previous chapter (Subheading 3.1) and mix it with the protein in equimolar ratio and the concentration that will later be used during D_2O labeling. Incubate the mixture for 10–20 min to ensure binding equilibrium. Also, prepare a sample without DNA for the comparison of how DNA affects the digestion pattern. In this step, use normal (H_2O, not D_2O based) buffers.
2. Start the LC-MS/MS system—calibrate mass spectrometer, start analytical gradient pump (Agilent 1290, running at 40 μl/min), loading pump (Agilent 1260, running at 100 μl/min), and pre-cool the LC setup to 0 °C (Fig. 2). If digestion is carried out at higher temperature, make sure that the protease column is well conditioned (*see* **Note 13**).

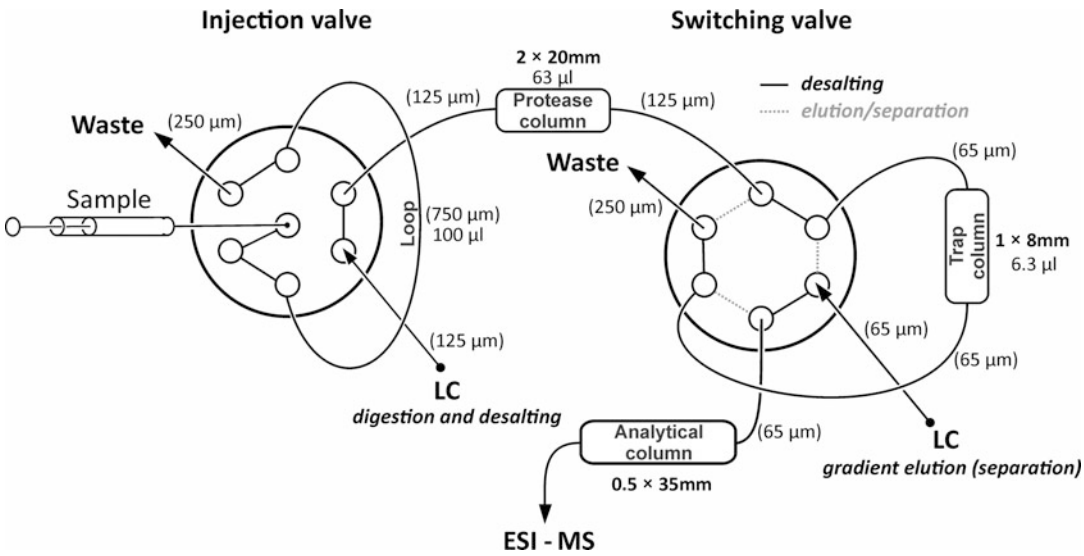


Fig. 2 Schematic representation of HDX LC-MS setup. Dimensions of the columns and tubing are shown. The setup is in position where digestion and desalting are running. Upon switch of the switching valve, the ports are connected by the gray dashed lines and gradient elution of the desalted peptides from trap column onto an analytical column is performed

3. Run a standard protein (e.g., horse heart myoglobin or rabbit muscle phosphorylase B) to verify the LC separation and protease column activity. This standard protein used for system quality control must be different from the protein of interest. Always keep in mind that the protein must be prepared in an acidic buffer as the digestion is done using aspartic proteases. Accidental injection of neutral pH buffer will quickly and irreversibly inhibit pepsin. Other proteases mentioned here are more robust in this regard and can tolerate elevated pH temporarily. Typically prepare 30 pmol of myoglobin in 100 µl of 250 mM glycine-Cl buffer pH 2.3.
4. Run two blank injections (100 µl of pure quench buffer) to clean the system and to ensure proper and stable LC conditions (*see* **Note 14**).
5. Mix the sample with the quench buffer to get the sample volume and concentration as planned for subsequent HDX-MS experiment (*see* **Note 15**). Hence, 50 µl of 500 mM glycine-Cl buffer pH 2.3 with 50 µl of 2 µM protein solution in 20 mM HEPES, pH 7.4, 150 mM NaCl.
6. Inject the sample directly or subject it to a freeze–thaw cycle if you plan to collect aliquots by freezing in liquid nitrogen.
7. Wait for desired time (3 min) until digestion and desalting are finished (*see* **Note 16**) and switch the second LC valve so that the trap column is now in the path of gradient elution and the peptides will continuously elute on the analytical column. Together with the switch, also start the gradient (*see* **Note 17**) and collect the MS data.
8. Process the LC-MS/MS file using the instrument vendor-specific software and generate input for the search engine (*see* **Note 18**).
9. Run a database search using the input file from the previous step. Use a custom-made database containing the sequence of the protein of interest and the protease(s) used and eventually other proteins present in the sample. Do not set any digestion preferences, nor taxonomy filtering. Include possible relevant modifications (fixed or variable) that may occur in the protein. Set the mass tolerance on precursor and fragments according to your instrument performance. Use scoring routines to discard improbable matches. It is also advisable to run the LC-MS/MS analysis 3–4 times and only use peptide identifications that occur in the majority of the analyses.
10. Copy or export the search result (list of identified peptides) to a spreadsheet editor (e.g., Microsoft Excel) and extract the Peptide Start and End columns (peptide limits). Use this in a simple text file to visualize the coverage map using DrawMap

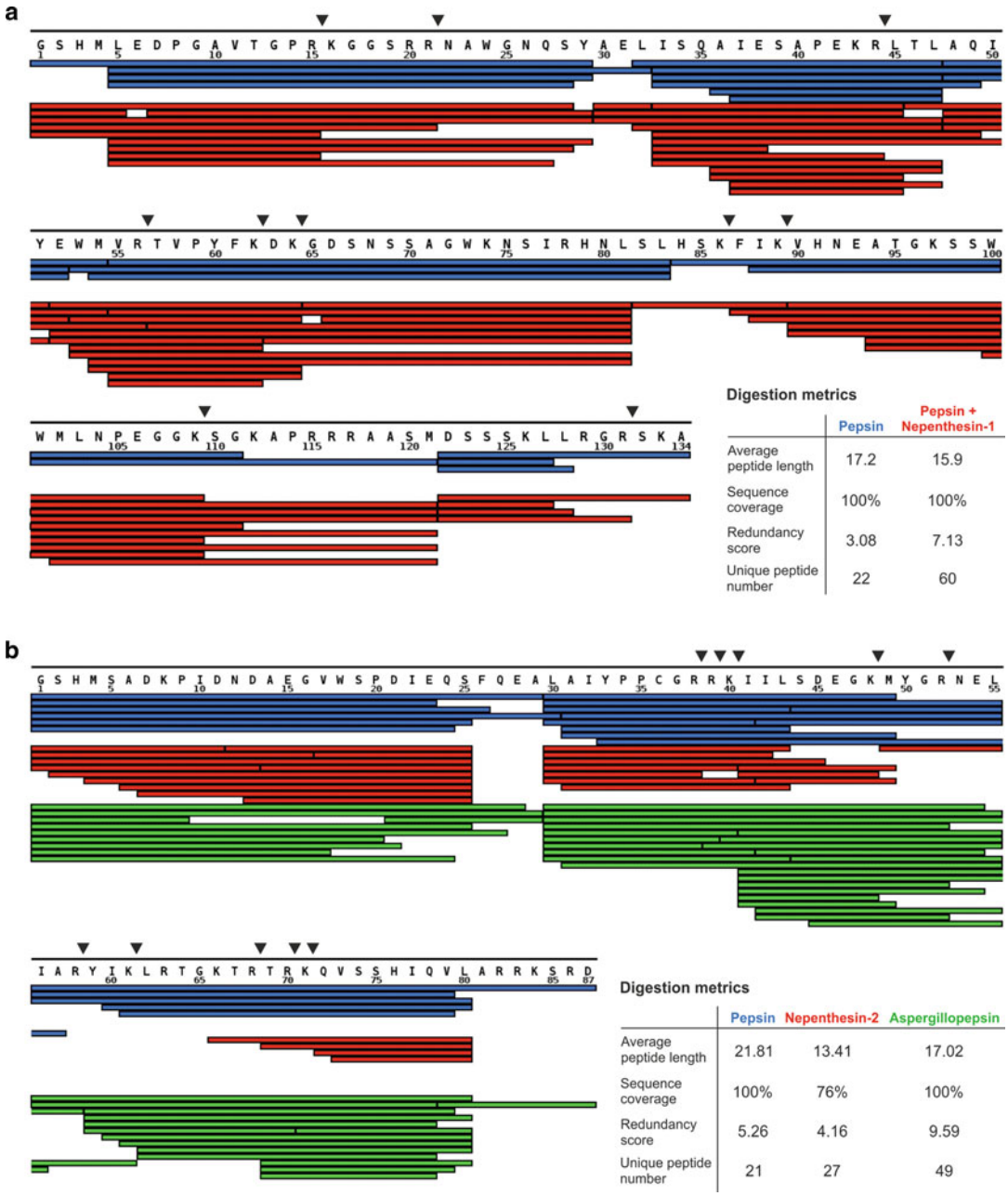


Fig. 3 Sequence coverages of the two studied DNA binding domains. In all instances, immobilized proteases were used. Table at the right lower corner of each panel shows digestion metrics for the individual conditions (coverage, number of peptides, average peptide length, and redundancy—how many peptides are on average covering each residue). (**a**) DNA binding domain of FOXO4 where initial proteolysis by pepsin (blue bars) was after optimization replaced by combined digestion with pepsin followed by nepenthesin-1 (red bars). This led to better spatial resolution and higher redundancy. In case of FOXO4, the presence of oligonucleotides (DAF-16) caused no problems, and thus, only tuning of the digestion itself was necessary. (**b**) Examples from the tuning procedure with TEAD1 DNA binding domain. Pepsin (blue bars) provided full sequence

script at the MSTools website (http://peterslab.org/MSTools/DrawMap/DrawMap.php) and eventually use Excel macro Digestion_metrics.xlsm (http://peterslab.org/downloads.php under HDX tools section) to extract the digestion metrics.

11. Based on the peptide length, redundancy, coverage, and number of gaps in the peptide map, try optimizing the digestion by using alternative proteases or quench buffer additives as described above (example in Fig. 3) (*see* **Note 19**).

3.3 Hydrogen/Deuterium Exchange

The methodology described here represents a manual approach. Even though nowadays there are robotic solutions available that can do the sample preparation and injection automatically, they are quite specific depending on the manufacturer of the robot and the software controlling it. The manual approach represents a straightforward and inexpensive way that can be easily set up in any laboratory possessing an LC system and a mass spectrometer (*see* **Note 20**).

The actual H/D exchange should be started from equilibrated solutions in which the complex is fully formed, and the temperature of all components is stable. The first step in the deuterium labeling reaction is the dilution of the studied protein into a deuterated buffer (*see* **Note 21**) that has otherwise identical composition as the initial buffer used for complex pre-formation. After a predetermined time, the exchange is "stopped" by the addition of acidic quench buffer and by lowering the temperature or eventually by flash freezing in liquid nitrogen.

The experiments should ideally be done (at least) in triplicate to allow for statistical validation. There are several ways how to perform the replication in HDX-MS. The highest validity is achieved by doing a biological replication (in the case of recombinant proteins this means using independent protein production batches). A less significant (though commonly used) level of replication (technical) is reached by separate deuterium labeling, and the lowest tier is represented by measuring the same labeling reaction repeatedly [28]. Replication also allows one to estimate the significance level above which the differences in deuteration between individual states may be considered meaningful [29].

Fig. 3 (continued) coverage but lower spatial resolution—on average it had longest peptides. Nepenthesin-2 (red bars) led to over-digestion, which is indicated by gaps in the sequence, and the identified peptides are mostly short ones. High activity of the protease resulted in the generation of short peptides and incomplete sequence coverage. The green bars show the final conditions where aspergillopepsin (protease type XIII) was utilized in the presence of urea. Addition of denaturing agent was necessary as the protein in the presence of oligonucleotides precipitated. Addition of urea (optimized to final 2 M concentration) led to much better protein recovery. In both examples shown here, the advantage of alternative proteases for the digestion of DNA binding proteins (which are rich in basic residues) is demonstrated through their ability to cleave after the Lys and Arg residues (highlighted by arrowheads above the sequence)

Starting from the moment of quenching and throughout the digestion, desalting, and subsequent analysis, the sample is exposed to protonated solvents. Even though the minimal exchange conditions are maintained through these steps, the sample still undergoes deuterium loss. The level of back-exchange is not uniform and depends on the sequence of individual peptides. However, this loss can be measured by analyzing a fully deuterated sample and then using these data, to correct deuteration levels on partially deuterated peptides [28]. In a typical measurement comparing two protein states, relative deuterium levels are the desired measure, and there is no need for back-exchange correction. Thus, in these cases, this control is often omitted. This is also likely due to the fact that no clear and universal protocol exists and that proteins often tend to precipitate or degrade before their full deuteration is achieved. However, the knowledge of the actual back-exchange is needed when absolute levels of deuteration are sought (e.g., in cases where mutated sequences of the same protein are compared) [30].

1. Prepare the required number of 0.5-ml Eppendorf tubes (for two conditions, three labeling replicates, and six time points, this would mean 36 tubes in total) with quench solution—50 μl of 500 mM glycine-Cl buffer, pH 2.3 (*see* **Note 22**).
2. Prepare (bigger) tubes with the protein alone and with the pre-formed protein–DNA complex. In these, the ten-fold dilution into deuterated buffer and the labeling reaction will be performed. At six preselected time points (*see* **Note 23**), 50 μl aliquots will be removed from the reaction and quenched. As three labeling replicates are to be done, three tubes with protein alone and three with the protein–DNA complex, each containing 40 μl of 2 μM protein solution must be prepared (*see* **Note 24**).
3. Prepare 2 ml of deuterated buffer (20 mM HEPES, 150 mM NaCl, pD 7.4) (*see* **Note 25**).
4. Prepare a time schedule that allows efficient pipetting especially in cases where larger amount of conditions or direct technical (labeling) replicates are performed at once. This can be easily done using MSTools script "Experiment planner" (http://peterslab.org/MSTools/HDExpPlanner/HDExpPlanner.php). For two experimental conditions, each replicated three times, and aliquot collection at 20 s, 2 min, 5 min, 20 min, 1 h, 3 h, this may look as shown in Fig. 4.
5. Prepare automatic pipettes with pre-set volumes for H/D mixing (360 μl) and aliquot collection (50 μl).
6. Use a timer that allows countdown followed by a count-up. Set 20 s count-down and start it. Aspirate 360 μl of the deuterated buffer and when the timer reaches zero (start of the whole

Time	Event
---	Mix TF-r1
20sec	Aliq no. 1 of TF-r1
2min	Aliq no. 2 of TF-r1
3min 30sec	Mix TF-DNA-r2
3min 50sec	Aliq no. 1 of TF-DNA-r2
5min	Aliq no. 3 of TF-r1
5min 30sec	Aliq no. 2 of TF-DNA-r2
7min	Mix TF-r2
7min 20sec	Aliq no. 1 of TF-r2
8min 30sec	Aliq no. 3 of TF-DNA-r2
9min	Aliq no. 2 of TF-r2
10min 30sec	Mix TF-DNA-r2
10min 50sec	Aliq no. 1 of TF-DNA-r2
12min	Aliq no. 3 of TF-r2
12min 30sec	Aliq no. 2 of TF-DNA-r2
14min	Mix TF-r3
14min 20sec	Aliq no. 1 of TF-r3
15min 30sec	Aliq no. 3 of TF-DNA-r2
16min	Aliq no. 2 of TF-r3
17min 30sec	Mix TF-DNA-r3
17min 50sec	Aliq no. 1 of TF-DNA-r3

Time	Event
19min	Aliq no. 3 of TF-r3
19min 30sec	Aliq no. 2 of TF-DNA-r3
20min	Aliq no. 4 of TF-r1
22min 30sec	Aliq no. 3 of TF-DNA-r3
23min 30sec	Aliq no. 4 of TF-DNA-r2
27min	Aliq no. 4 of TF-r2
30min 30sec	Aliq no. 4 of TF-DNA-r2
34min	Aliq no. 4 of TF-r3
37min 30sec	Aliq no. 4 of TF-DNA-r3
1h	Aliq no. 5 of TF-r1
1h 3min 30sec	Aliq no. 5 of TF-DNA-r2
1h 7min	Aliq no. 5 of TF-r2
1h 10min 30sec	Aliq no. 5 of TF-DNA-r2
1h 14min	Aliq no. 5 of TF-r3
1h 17min 30sec	Aliq no. 5 of TF-DNA-r3
3h	Aliq no. 6 of TF-r1
3h 3min 30sec	Aliq no. 6 of TF-DNA-r2
3h 7min	Aliq no. 6 of TF-r2
3h 10min 30sec	Aliq no. 6 of TF-DNA-r2
3h 14min	Aliq no. 6 of TF-r3
3h 17min 30sec	Aliq no. 6 of TF-DNA-r3

Fig. 4 Screenshot of HD Experiment planner script (MSTools: http:/peterslab.org/MSTools/HDExpPlanner/HDExpPlanner.php). Exact planning of aliquot collection for two experimental conditions (TF and TF-DNA) and three labeling replicates. Delays necessary to collect the aliquot or mix the exchange can be varied and are especially useful for larger number of conditions—here 30 s is left for aliquoting and 1 min for mixing

HDX experiment), add the buffer to the protein. Mix thoroughly and prepare the pipette (50 µl) for aliquot collection. Five seconds before the selected time (20 s), aspirate the labeled protein and at the precise time according to the experiment plan, add aspirated solution to the tube with the quench solution. Quickly mix/vortex and flash-freeze the tube in liquid nitrogen. Do this for all the samples according to the time schedule at times indicated by the timer.

7. Transfer the frozen tubes from liquid nitrogen to a deep freezer and store them until analysis (*see* **Note 26**).
8. Also prepare non-labeled reference samples (a non-deuterated control) for each condition. This is the same as preparing the (partially) deuterated samples, except that the buffers used are made with H_2O only.
9. If possible (depends on protein stability), prepare fully deuterated sample (*see* **Note 27**). In order to obtain a fully deuterated control, prepare deuterated buffer containing high concentration (4–6 M final concentration) of urea or guanidine and incubate the protein in it overnight at higher (e.g., 37 °C) temperature. Protein concentration should be as high as feasible, so that it can be lowered before quenching using

non-denaturing deuterated buffer. This will lower the denaturant concentration which can alter the subsequent workflow outcomes. From this perspective, using urea is the better option (*see* Subheading 3.2). Quench the reaction and eventually freeze the sample, if this step is included in the workflow.

3.4 Mass Spectrometric Analysis of Deuterated Samples

1. Use the LC system as in Subheading 3.2 and perform **steps 2–4**. Start all LC pumps, calibrate the mass spectrometer, pre-wash, and pre-condition the protease column. Then run the standard protein followed by two blank injections (*see* **Note 28**).
2. Before the end of previous analysis, take one of the partially deuterated samples from the freezer and start thawing it. Depending on the quench buffer composition, it will melt in 30 s to 1 min. As soon as it is thawed, inject it for analysis.
3. Repeat this for all collected samples (*see* **Note 29**).
4. Analyze fully deuterated control exactly as the partially deuterated samples.
5. Finally run two blanks followed by non-deuterated samples for each experimental condition.
6. Export and/or pre-process all LC-MS acquired data and prepare them for final data processing.

3.5 Interpreting Data from H/D Exchange

Nowadays, there are several different programs that are capable of largely automatic HDX-MS data processing. Waters users rely on their DynamX suite while others can use HDExaminer from Sierra Analytics that supports all native file types as well as open source ones. Another option for Waters and Thermo users is represented by the HDX Workbench [31] from Omics Informatics. Alternatively, there are software tools freely available from different research groups [22, 32–35]. These and other programs and their workflows were recently reviewed in detail by Claesen and Burzykowski and also by Eggertson et al. which in addition offers another view on HDX-MS data processing [36, 37].

The basic principle which is crucial to the understanding of the workflow and its requirements is, however, the same. To make this protocol widely applicable, we will show two possible scenarios. One, relying on a manual interpretation which is nowadays outdated and extremely laborious. However, it demonstrates well the basic principle which the available programs automate. It can also be useful for validation purposes and in specific cases (extraction of EX1/EX2 data). The other workflow presented here employs our own software called DeutEx (*see* **Note 30**).

Data processing in the HDX-MS workflow consists of four major steps. First is the identification of peptides generated during the proteolysis step and their temporal localization within an LC

run. Using the set of coordinates consisting of peptide sequence, its mass-over-charge (m/z) value(s) (*see* **Note 31**), and LC retention time, individual features can be located in the partially deuterated samples. In a subsequent step, information about the number of deuterons carried by individual peptides is extracted. Two major approaches to deuteration readout rely either on centroid readout or on deconvolution and theoretical fitting [36]. Finally, this information is presented in several graphical ways to visualize the effects monitored by the HDX-MS experiment.

3.5.1 Workflow A: (Largely Automated) Data Interpretation Using DeutEx

1. From an MS/MS search engine, e.g., MASCOT (here utilized by Bruker's ProteinScape), export the search result in the form of a csv file (*see* **Note 32**).
2. Prepare a separate simple text file containing the sequence of the studied protein in FASTA format.
3. Open DeutEx and go to *Analysis>Compose digest file from MS*. New window will appear. In this window, upload your protein sequence as a FASTA file, then csv file containing the list of identified peptides from MASCOT and an LC-MS run of a non-deuterated sample exported into simple text files (each MS scan corresponding to one file) containing a list of m/z values and their intensities (*see* **Note 33**).
4. Go to *Digest*, click *Create from report* and then *Extract scan/charge limits*. Each peptide will have its retention limits automatically identified, and extracted ion chromatograms will be drawn for individual charge states (m/z values). User can manually inspect the assignments and either manually correct or completely reject them. Once finished, the result should be exported into a text file that serves as a lookup file for an automated assignment of HDX data for the deuterated samples. Close the "Compose digest from MS window."
5. Go to *Analysis>Open Analysis directory*. It is advisable to pre-compile data for each state into a separate folder (which contains exported LC-MS data in a txt format, with appropriate fasta sequence and digest file—the list of peptides created in **step 4**). However, the data can also be compiled in DeutEx using the "Compose analysis" function.
6. Once all data are opened, check the settings for mass accuracies and filtering under the options menu and run the calculation. Typically, it lasts from seconds to a few minutes, depending on the number of conditions, number of peptides, and size of the exported data on a common laptop/office computer. When calculation is finished, run filtering, which tries to automatically remove incorrect assignments. Then look through the assignments manually, check the data, and correct any misassignments, if needed.

7. The basic output from DeutEx is a set of uptake plots that are displayed either as deuteration percentage or number of exchanged deuterons as a function of time. The data can also be exported in simple text form which allows their further processing in Excel or MSTools (http://peterslab.org/MSTools/) [38].

3.5.2 Workflow B: Manual Data Interpretation

1. In any program that is used for viewing the LC-MS data, open a non-deuterated LC-MS analysis.
2. For each peptide, trace an extracted ion chromatogram based on its *m/z* (usually included in the search engine report). Check also other charge states that are visible and enter this information into an Excel sheet together with the retention time limits for further use. Sum scans across the LC peak and export the spectrum as a simple text file (typically two columns including *m/z* values and intensity).
3. Open a partially deuterated LC-MS analysis and locate the deuterated signal corresponding to the peptide information gathered in the previous step. Repeat the step of summing up the scans and exporting the spectra into text files. Repeat iteratively for all peptides. For such a manual approach, it is advisable to select a minimal set of peptides (preferably short ones) covering the entire protein without many overlaps (*see* **Note 34**).
4. Once the data are exported to *.txt files, run mMass program [39] (*see* **Note 35**). Go to *File>Open* and open a file with non-deuterated data. The spectrum is not by default labeled, but it is not necessary at this point. Make sure that the correct way of peak picking is selected. Go to tab *Processing>Deisotoping* and set *"Label envelope tool"* to *Envelope Centroid* option. Now locate the selected peak (*m/z*) and zoom to the mass range covering the whole isotopic pattern of this peptide. Use *"Label envelope tool"* (select from tab *Tools*). Using Shift + mouse middle wheel scrolling, set the desired charge state—black dots are changing distance as you scroll up and down. Fit it to the desired isotopomer distance and now select the number of dots to cover all visible isotopic peaks from the particular envelope. This is done by Ctrl + Shift + mouse middle button scrolling. This tool helps you to localize all isotopes belonging to the selected peak. Left click and the peak will be labeled. Average *m/z* value is written into the spectrum, and the deconvoluted value (average mass) appears in the right panel (Fig. 5). Write this number down to an Excel file. Eventually, you can now repeat this step also for other charge states of the same peptide. One by one, all *m/z* values of each peptide from all deuterated samples are processed using this approach (*see* **Note 36**).

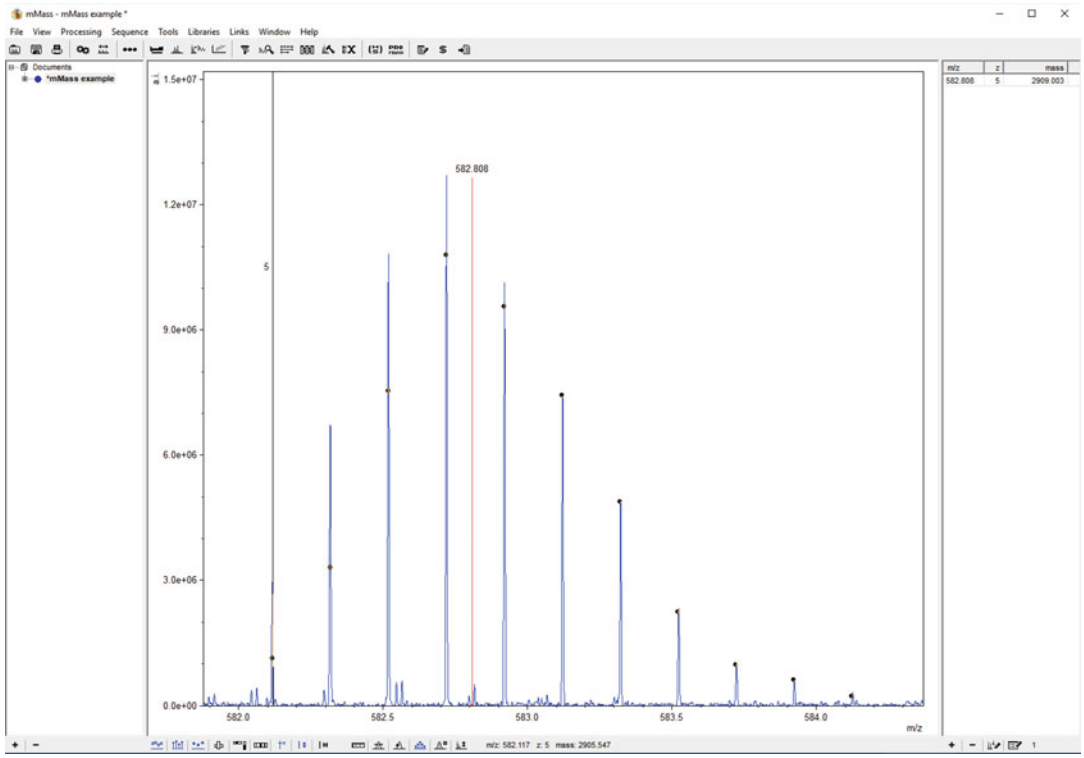


Fig. 5 Screenshot of mMass window with zoom on the isotopic pattern. Black dots are labels showing which isotopes will be used for M_{avg} calculation. The deconvoluted average mass is shown in the right panel

5. When all peptides are processed and the Excel file complete, the percentage of deuteration or the number of deuterons can be easily calculated [2] and used for visualization of the results.

3.6 Data Visualization

There are no set guidelines on how to visualize HDX data, and indeed, a multitude of different plots are used throughout the literature. The first result that is usually obtained is a set of uptake plots in which the percentage of deuteration or the number of deuterons is plotted for each peptide as a function of time. However, from such a display, it is difficult to interpret the data in the structural context and also to efficiently handle the information from overlapping peptides. For a more comprehensive view, heat maps or protection plots [38] are most often utilized. These summarize the information into one picture and present it along the protein sequence. However, their basic form does not allow one to visualize information from the overlapping peptides as the data are plotted using only a subset of selected (usually shortest) peptides continuously covering the entire sequence. To also include the redundancy information, peptide midpoints can be calculated [29] for each peptide, which allows the distinction of closely related ones (e.g., 14–25, 14–27, 16–25, 16–27). Upon this modification,

the data can also efficiently be used in mirror (or so-called butterfly) plots or in bar graphs. The disadvantage of this data display is that the proportions of the peptides versus the protein sequence are lost. Probably one of the most attractive ways of data presentation is thus mapping the observed changes onto an existing high-resolution structure or protein model. In this case, differences between two protein states are calculated (preferably for the number of deuterons not for percentages), and the structure is colored according to the observed changes. Visualization tools are usually included in the HDX specific software (DynamX, HDExaminer, etc.,); however, our web-based MSTools suite makes many visualization tools available for vendor-independent use [38].

Moreover, for structure coloring, PyMol with its scripting capabilities is unparalleled. In this workflow, the b-factors of pdb format's fields in an existing protein structure are hijacked and replaced with the numbers from an HDX experiment. To load the data into PyMol, it is necessary to cut each peptide virtually into individual amino acids and use this as an input for coloring.

1. Prepare a table which contains information about each peptide's start, end, time of deuteration (in seconds), and the difference in deuteration between selected states:

Start	End	Time(s)	(TF-DNA—TF)
1	15	20	0.3
1	15	120	0.5
1	15	300	0.6
1	15	1200	0.8
1	15	3600	1.5
1	15	10,800	2.1
8	17	20	0.2
8	17	120	0.4
8	17	300	0.5

Save this table as tab delimited simple text file into a new folder.

2. Copy Python script HDXPeptideSplitter.py (*see* **Note 37**) into this folder and run it (*see* **Note 38**). Command window will appear, and the script will ask for the name of the input file. Then it requires information about the time to be parsed (selected time or all the times in the file) and finally about the conditions to be extracted (again selected one or all). The script then creates a set of files depending on the number of conditions and time points in the experiment. Names of the files are

created using the name of the input file, time, and number of the column in the input file (e.g., FOXO_20_2.txt means 20-s time point and second column from the input file which was named FOXO.txt).

3. Open PyMol and load the pdb file with the structure. Using "copy" (https://pymolwiki.org/index.php/Copy) command, get the desired number of copies corresponding to the number of states and conditions prepared in the previous step. Also name the objects so that the names correspond to the conditions and times from the dataset.
4. Within the PyMol command line, activate script data2bfactor.py (*see* **Note 39**) by the command "*run path/data2bfactor.py*" (where "path" stands for the absolute path to the folder, where the script is saved on your computer), e.g., *run C:\Python27\Scripts\data2bfactor.py.*
5. Reset the b factors to a number that is outside the dataset (e.g., 99). This will later allow you to color the regions that are not covered by the peptides in the dataset. Enter command "*alter object_name, b=99*" (where "object_name" stands for the name of an individual object/structure chosen by the user).
6. Color the structure(s) to some neutral, uniform color using command "color." E.g., "*color grey70.*"
7. Apply your data onto the individual structures. Use command "*data2b_res object_name, path.*" For example, *data2b_res FOXO_20_WT-MUT, C:\Data\FOXO_20_2.txt* (i.e., second column of the input file for HDXPeptideSplitter.py script contained difference between wild-type and mutant). Do this for all structures/objects in PyMol and all input text files.
8. Now select residues with all negative and positive b values, respectively. Enter commands "*select negative, b<0.0*" and "*select positive, b>0.0.*"
9. Color these selections using commands "*spectrum b, selection=negative, minimum=-X, maximum=0.0, palette=blue_white*" and "*spectrum b, selection=positive, minimum=0.0, maximum=X, palette=white_red*," where X is the highest/lowest number in your dataset. Negative (white-to-blue gradient) differences are regions where the exchange was lowered, and positive (white-to-red gradient) are regions with increased deuteration.
10. Finally, differentiate the regions that were not covered by peptides in the dataset. Run command "*color black, (b=99).*"
11. Optionally show a colorbar with the scale using command "*ramp_new colorbar, none, [-X, 0.0, X], [blue, white, red].*"

4 Notes

1. The influence of pH and temperature on the intrinsic chemical exchange rate must be considered here. If two or more states of a protein are followed at different pH values or temperatures, appropriate correction must be applied. Temperature factor can be calculated from the Arrhenius equation (e.g., comparison of exchange at 10 °C and 37 °C will be corrected by a factor of 13.88). The effect of pH is compensated by 10ΔpH value (e.g., exchange at pH 7 runs 100 times faster than exchange at pH 5) [3].
2. Total amount depends on protein size, the LC system used, and the sensitivity of the mass spectrometer. However, tens (small protein) to hundreds (large protein) of micrograms are enough for the comparison of two states and performing triplicate analysis.
3. Check the oligonucleotides for possible secondary structure formation which may affect their binding properties. When preparing/synthesizing or ordering the oligonucleotides, use HPLC purification as the last step to obtain desalted DNA samples.
4. In case where water/ice mixture is used, the valve body should be properly sealed and isolated from the water (parafilm and heat shrink tubing) as it prolongs the lifetime of the valve.
5. A wide range of buffer components or concentrations is tolerated in H/D exchange workflow as long as they can be removed during desalting on a reversed phase resin that precedes separation and mass spectrometric detection. Nevertheless, the buffer components should be kept at their lowest concentration where both the free protein and the complex are still stable.
6. At least ten times K_D [18]. If K_D is not known, try multiple concentrations. Keep in mind that the complex will be diluted during D_2O labeling. Do not use excess of DNA since high DNA concentration can lead to sample precipitation upon quenching.
7. Depending on the melting point temperature of the used dsDNA, cooling might be needed.
8. Try to keep the volume of samples as low as possible to prevent diffusion from the wells during sample loading—ideally under 5 μl.
9. The concentration of the gel should be adjusted based on oligonucleotide length.

10. If possible, cool the electrophoresis apparatus during the run since the complex or DNA duplex might not be stable at higher temperatures.
11. It should be considered that the fragmentation of a particular peptide does not always lead to sufficient number of fragments. Therefore, the combination of different digestion protocols and dissociation techniques usually represents the best solution.
12. Fragments covering the same region of protein and differing in just few amino acids from the N- or C-terminus should result in fairly similar H/D uptake plots. Hence, the slope, level of deuteration, or eventually the differences between the individual studied states should not be vastly different. If they are, this might be an indication of a possible data processing error. On the other hand, if there are multiple individual peptides spanning one protein region and all show highly similar trend, the confidence in the observed changes is significantly increased.
13. Protease column may suffer from autodigestion upon longer storage. Storing it at conditions that do not affect its activity, while preventing autodigestion (e.g., citrate buffer, pH 4–5, 4 °C), is highly recommended. Also, pre-washing with the loading solvent before connecting the column in front of the trap column should be routinely done during an LC system setup.
14. Specific cleaning solutions containing denaturing agents or higher concentration of acids can be designed to minimize carry-over [40]. Good regeneration of the whole system is likely achieved by combining injection of some other (quality control) protein together with these washing steps.
15. Quench conditions (quench solution composition, concentration, quench ratio) should be tuned prior to the experiment using pure buffers. While protein concentration during the labeling step is given by the requirements on the protein–DNA complex formation, the final concentration and sample volume after quenching depend solely on the requirements of the mass spectrometer and LC setup (injection volume, trap, and protease column size + flow). For the MS analysis, it is worth not considering the true detection limits of the instrument, but staying well above them, as the signal intensity after labeling decreases due to signal spreading to multiple isotopes upon deuteration. Keeping the same protein concentration for LC-MS/MS as for the subsequent LC-MS analyses is crucial as the protein:protease ratio affects the digestion pattern.
16. Sum the volumes of injection loop, protease column, trap column, and the connecting tubing and compare it to the selected flow on loading/desalting pump. Make sure that

the sample will reach the trap column and several volumes of the trap column will pass through to achieve efficient desalting. If guanidine or phosphates are used in sample preparation, make the desalting slightly longer. This step usually lasts from 1 to 4 min, depending on the flow and volumes mentioned above.

17. Duration of the gradient and its slope depend on the column and the LC system used. In general, the gradient should be fast, lasting just a few minutes. However, during this short period, the most efficient elution should be achieved. Therefore, gradient running from 5 or 10% acetonitrile (solvent B) and ending at 25–35% acetonitrile should be used. In specific cases, the upper acetonitrile percentage can be raised to 40 or 50% or even higher.
18. For example, here the data are processed in DataAnalysis using Find>Compounds AutoMS(n) feature and exported into a MASCOT generic file (*.mgf). Two different algorithms are used for peak picking (FTMS and SNAP) and thus two *.mgf files are created for each analysis. The whole procedure can be automated using Visual Basic macros in Data Analysis. These macros can be obtained from the authors.
19. If the coverage map shows many gaps and the covered parts are represented by short peptides, the protein was probably over-digested. Faster flow (even in hundreds of microliters per minute) should be tested. One should, however, keep in mind the trapping efficiency of the trap column at a given flow rate. Alternatively, another protease may be used. These steps are also advisable if higher redundancy is desired. In cases where the peptides are too long, temperature should be increased or addition of a denaturing agent considered. Use of alternative aspartic proteases beside pepsin is advantageous for DNA binding proteins since they tend to be rich in basic residues. While pepsin exerts nearly no cleavage after Lys, His, and Arg residues, the other proteases like nepenthesins, aspergillopepsin, and rhizopuspepsin cleave at these sites with high frequency [17, 24, 26]. These proteases may thus yield good spatial resolution of basic proteins or basic protein regions.
20. It is often stated that the use of robotics leads to higher reproducibility; however, trained user and careful pipetting can lead to very similar outcomes. The main advantage in using automation is the ability to perform the analysis in an unattended manner and thus using the instrument time more efficiently. Most importantly, the manual way utilizes freezing of samples and their subsequent thawing. These steps require additional 0.5–2 min, during which the sample undergoes back-exchange or (in the case of low D_2O dilution by the

quench buffer) artificial, nonnative, in-exchange. In contrast, robotics can set up the mixing of H/D exchange samples consecutively, so that once the exchange is finished by the quench solution addition, the sample is directly injected for analysis. It is also of high importance that some proteins tend to precipitate during freezing/thawing procedures and here the use of robotics provides additional advantage.

21. Dilution is often done ten-fold, but higher (20-fold) or lower (five-fold) ratios are also possible. The key consideration here is the stability of the complex at a given concentration which is influenced by the K_D of the complex.

22. The amount/volume of the quench solution is given by the quench ratio. Optimally, ten-fold dilution is used but other ratios, e.g., 1:1 are also acceptable, if necessary. Quench solution must be acidic and strong enough to override pH of the labeling buffer and should make it 2.3–2.5, instead. Diluted solutions of hydrochloric or phosphoric acid, phosphate buffers, or glycine hydrochloride buffers are typical examples. Quench solution may also contain denaturing agents or other components necessary for the solubility of the protein or its analysis. Only components that are not easily removed by reversed phase desalting and are not compatible with ESI-MS should be avoided (e.g., detergents). However, even for nonionic detergents, specific workflows allowing their nearly complete removal under HDX-MS conditions have been described [41].

    The tubes used for freezing should be PCR tubes as they have to be stable at extreme temperature. Normal Eppendorf microcentrifuge tubes can break or take up some liquid nitrogen, which can lead to their explosion during the subsequent thawing procedure.

23. Preferably run H/D exchange with several time points spanning the region from seconds to a couple of hours. For longer incubation times, a control of protein stability should be performed by incubating the protein for the longest selected time in the normal buffer and then deuterium labeling it for the shortest selected period (e.g., 20 s). If the deuteration is the same as for the sample just deuterated for this time window, the protein is probably stable. If the deuteration profile differs/is indicative of protein unfolding, shorter time windows should be sought or the exchange should be done at lower temperature.

24. Simple setup requires two protein states: protein alone and protein in complex with its DNA binding partner. However, number of conditions to be followed can be much larger, e.g., with different protein:DNA ratios or different ligands.

25. Deuterated buffers must be set to pD (not pH). There is a simple correction factor compensating the difference of measuring deuteron concentration using normal pH electrodes $pD = pH_{read} + 0.41$ [42]. Alternatively, pH reading (clearly marked as such) can be reported instead of the pD value.
26. Samples should be analyzed as soon as possible, e.g., within a few days. Do not store them for more than 1 week, even at very low temperatures.
27. Some proteins are not stable enough to allow the preparation of a fully deuterated control. For the majority of comparative HDX-MS experiments, its inclusion is anyway not crucial, but it can offer valuable insight into the back-exchange levels and provides good quality control of the system used.
28. Tune the mass spectrometer properly to avoid hydrogen scrambling in the ion source [43]. That mainly means lower the desolvation temperature and possible ion activation during ion transfer. These low-scrambling parameters may, however, lead to decrease in signal intensity.
29. It is best not to analyze the samples in any ordered manner, e.g., sequentially by the conditions or replicates. If the analysis is stopped or interrupted, recondition the system by running two or three blanks or a standard protein and two blanks. Cleaning procedures should be included, and potential carryover checked during the analysis. The biggest source of carryover is the protease column, which can, however, be washed during each run [40]. Try to keep the system running continuously so that maximal LC reproducibility is maintained.
30. DeutEx is not yet publicly available but will be offered at our website: http://peterslab.org/downloads.php under HDX tools section.
31. It might be beneficial to search for all detected charge states (m/z values) for each peptide, thus not relying only on those fragmented during the LC-MS/MS run and identified during the subsequent search. This will yield more data points which can be either merged and statistically processed or used separately as redundant information supporting the data validity.
32. DeutEx will accept any text form of the input and has automated detection of the content in the columns. If that fails, the user can correct it and assign the column headers manually.
33. For better clarity, a short video demonstrating the whole workflow is available at http://peterslab.org/downloads/SW/DeutEx.mp4.
34. In Data Analysis, which is delivered together with MS instruments from Bruker Daltonics, this procedure can be partially automated using a Visual Basic macro. This macro can be obtained from the authors upon request.

35. mMass is a freely available software suitable for processing of single spectra. It can be obtained from http://mmass.org/.
36. There are also other ways of data processing. Similar to mMass processing is the use of the MagTran program written by Zhongqi Zhang. Also, HX-Express2 (http://www.hxms.com/HXExpress/) can be used for this purpose [32]. The advantage of HX-Express 2 is its ability to detect and dissect bimodal isotopic envelopes that are typical for mixed EX1/EX2 exchange kinetics.
37. Script can be obtained at http://peterslab.org/downloads.php under HDX tools section.
38. To run Python scripts directly, install python on your PC (https://www.python.org/downloads/) and include it in the PATH variable of the operating system (version used here was 2.7.12).
39. Script was written by Robert L. Campbell and is available here: http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/.

Acknowledgments

Czech Science Foundation projects 16-24309S and 16-20860S are gratefully acknowledged. Additional support was obtained from EU/MEYS projects BioCeV (CZ.1.05/1.1.00/02.0109) and NPU II (LQ1604). R.F. also thanks Charles University Grant Agency (project 1618218) and SVV260427/2019.

References

1. Katta V, Chait BT, Carr S (1991) Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. Rapid Commun Mass Spectrom 5:214–217
2. Zhang Z, Smith DL (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. Protein Sci 2:522–531
3. Bai YW, Milne JS, Mayne L et al (1993) Primary structure effects on peptide group hydrogen-exchange. Proteins 17:75–86
4. Engen JR, Wales TE (2015) Analytical aspects of hydrogen exchange mass spectrometry. Annu Rev Anal Chem 8:127–148
5. Oganesyan I, Lento C, Wilson DJ (2018) Contemporary hydrogen deuterium exchange mass spectrometry. Methods 144:27–42
6. Sorokin VA, Gladchenko GO, Valeev VA (1986) DNA protonation at low ionic strength of solution. Die Makromol Chemie 187:1053–1063
7. Ma L, Fitzgerald MC (2003) A new H/D exchange- and mass spectrometry-based method for thermodynamic analysis of protein-DNA interactions. Chem Biol 10:1205–1213
8. Sperry JB, Wilcox JM, Gross ML (2008) Strong anion exchange for studying protein-DNA interactions by H/D exchange mass spectrometry. J Am Soc Mass Spectrom 19:887–890
9. Sperry JB, Shi X, Rempel DL et al (2008) A mass spectrometric approach to the study of DNA-binding proteins: interaction of human TRF2 with telomeric DNA. Biochemistry 47:1797–1807
10. Poliakov A, Jardine P, Prevelige PE (2008) Hydrogen/deuterium exchange on protein solutions containing nucleic acids: utility of

protamine sulfate. Rapid Commun Mass Spectrom 22:2423–2428

11. Roberts VA, Pique ME, Hsu S et al (2012) Combining HD exchange mass spectroscopy and computational docking reveals extended DNA-binding surface on uracil-DNA glycosylase. Nucleic Acids Res 40:6070–6081
12. Graham BW, Tao Y, Dodge KL et al (2016) DNA interactions probed by hydrogen-deuterium exchange (HDX) Fourier transform ion cyclotron resonance mass spectrometry confirm external binding sites on the minichromosomal maintenance (MCM) helicase. J Biol Chem 291:12467–12480
13. Boura E, Silhan J, Herman P et al (2007) Both the N-terminal loop and wing W2 of the forkhead domain of transcription factor Foxo4 are important for DNA binding. J Biol Chem 282:8265–8275
14. Slavata L, Chmelik J, Kavan D et al (2019) MS-based approaches enable the structural characterization of transcription factor/DNA response element complex. Biomol Ther 9: E535
15. Anbanandam A, Albarado DC, Nguyen CT et al (2006) Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. Proc Natl Acad Sci U S A 103:17225–17230
16. Wang L, Pan H, Smith DL (2002) Hydrogen exchange-mass spectrometry. Mol Cell Proteomics 1:132–138
17. Kadek A, Mrazek H, Halada P et al (2014) Aspartic protease nepenthesin-1 as a tool for digestion in hydrogen/deuterium exchange mass spectrometry. Anal Chem 86:4287–4294
18. Kochert BA, Iacob RE, Wales TE et al (2018) Hydrogen-deuterium exchange mass spectrometry to study protein complexes. In: Methods in molecular biology (Clifton, N.J.). Humana Press, New York, NY, pp 153–171
19. Rand KD, Zehl M, Jensen ON et al (2009) Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. Anal Chem 81:5577–5584
20. Mistarz UH, Bellina B, Jensen PF et al (2018) UV Photodissociation mass spectrometry accurately localize sites of backbone Deuteration in peptides. Anal Chem 90:1077–1080
21. Mayne L, Kan ZY, Sevugan Chetty P et al (2011) Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation-mass spectrometry method. J Am Soc Mass Spectrom 22:1898–1905
22. Kan Z-Y, Walters BT, Mayne L et al (2013) Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. Proc Natl Acad Sci U S A 110:16438–16443
23. Cravello L, Lascoux D, Forest E (2003) Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. Rapid Commun Mass Spectrom 17:2387–2393
24. Rey M, Man P, Brandolin G et al (2009) Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in hydrogen/deuterium exchange mass spectrometry. Rapid Commun Mass Spectrom 23:3431–3438
25. Kadek A, Tretyachenko V, Mrazek H et al (2014) Expression and characterization of plant aspartic protease nepenthesin-1 from *Nepenthes gracilis*. Protein Expr Purif 95:121–128
26. Yang M, Hoeppner M, Rey M et al (2015) Recombinant Nepenthesin II for hydrogen/deuterium exchange mass spectrometry. Anal Chem 87:6681–6687
27. Kadek A, Kavan D, Marcoux J et al (2017) Interdomain electron transfer in cellobiose dehydrogenase is governed by surface electrostatics. Biochim Biophys Acta Gen Subj 1861:157–167
28. Moroco JA, Engen JR (2015) Replication in bioanalytical studies with HDX MS: aim as high as possible. Bioanalysis 7:1065–1067
29. Houde D, Berkowitz SA, Engen JR (2011) The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. J Pharm Sci 100:2071–2086
30. Wales TE, Poe JA, Emert-Sedlak L et al (2016) Hydrogen exchange mass spectrometry of related proteins with divergent sequences: a comparative study of HIV-1 Nef allelic variants. J Am Soc Mass Spectrom 27:1048–1061
31. Pascal BD, Willis S, Lauer JL et al (2012) HDXWorkbench: software for the analysis of H/D exchange MS data. J Am Soc Mass Spectrom 23:1512–1521
32. Guttman M, Weis DD, Engen JR et al (2013) Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra. J Am Soc Mass Spectrom 24:1906–1912
33. Lindner R, Lou X, Reinstein J et al (2014) Hexicon 2: automated processing of hydrogen-deuterium exchange mass spectrometry data with improved deuteration distribution estimation. J Am Soc Mass Spectrom 25:1018–1028
34. Rey M, Sarpe V, Burns KM et al (2014) Mass spec studio for integrative structural biology. Structure 22:1538–1548

35. Kan ZY, Ye X, Skinner JJ et al (2019) ExMS2: an integrated solution for hydrogen-deuterium exchange mass spectrometry data analysis. Anal Chem 91:7474–7481
36. Claesen J, Burzykowski T (2017) Computational methods and challenges in hydrogen/deuterium exchange mass spectrometry. Mass Spectrom Rev 36:649–667
37. Eggertson MJ, Fadgen K, Engen JR et al (2020) Considerations in the analysis of hydrogen exchange mass spectrometry data. Methods Mol Biol 2051:407–435
38. Kavan D, Man P (2011) MSTools - web based application for visualization and presentation of HXMS data. Int J Mass Spectrom 302:53–58
39. Strohalm M, Kavan D, Novak P et al (2010) mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Anal Chem 82:4648–4651
40. Majumdar R, Manikwar P, Hickey JM et al (2012) Minimizing carry-over in an online pepsin digestion system used for the H/D exchange mass spectrometric analysis of an IgG1 monoclonal antibody. J Am Soc Mass Spectrom 23:2140–2148
41. Rey M, Mrazek H, Pompach P et al (2010) Effective removal of nonionic detergents in protein mass spectrometry, hydrogen/deuterium exchange, and proteomics. Anal Chem 82:5107–5116
42. Glasoe PK, Long FA (1960) Use of glass electrodes to measure acidities in deuterium oxide 1,2. J Phys Chem 64:188–190
43. Guttman M, Wales TE, Whittington D et al (2016) Tuning a high transmission ion guide to prevent gas-phase proton exchange during H/D exchange MS analysis. J Am Soc Mass Spectrom 27:662–668

Chapter 12

Integrative Mass Spectrometry–Based Approaches for Modeling Macromolecular Assemblies

Andy M. Lau and Argyris Politis

Abstract

Mass spectrometry (MS)–based strategies have emerged as key elements for structural modeling of proteins and their assemblies. In particular, merging together complementary MS tools, through the so-called hybrid approaches, has enabled structural characterization of proteins in their near-native states. Here, we describe how different MS techniques, such as native MS, chemical cross-linking MS, and ion mobility MS, are brought together using sophisticated computational algorithms and modeling restraints. We demonstrate the applicability of the strategy by building accurate models of multimeric protein assemblies. These strategies can practically be applied to any protein complex of interest and be readily integrated with other structural approaches such as electron density maps from cryo-electron microscopy.

Key words Structural mass spectrometry, Computational modeling, Protein complexes, Hybrid approaches, Modeling restraints

1 Introduction

Integrative structural methods has come of age in building 3D model ensembles of proteins and their complexes [1]. These methods allow bringing together data sets sourced from diverse structural methods that otherwise will be reported independently [2]. Such integration can be achieved through the use of modeling restraints obtained directly from the experimental data [1, 2] (Fig. 1). Recent innovative studies have led to the expansion of the integrative structural toolkit to incorporate a variety of MS-based methods. These include powerful methods such as chemical cross-linking [3–6], ion mobility [7, 8], and native MS [9]. Moreover, these methods can be integrated with other powerful methods such as X-ray crystallography [10], cryo-EM [11], and hydrogen–deuterium exchange (HDX)-MS [11] structural approaches. Impressive examples of the integrative structural toolkit include the elucidation of many structural aspects of large complexes such as the proteasome [5, 6], ATP synthase [12], and nuclear pore complex [13].

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247,
https://doi.org/10.1007/978-1-0716-1126-5_12,

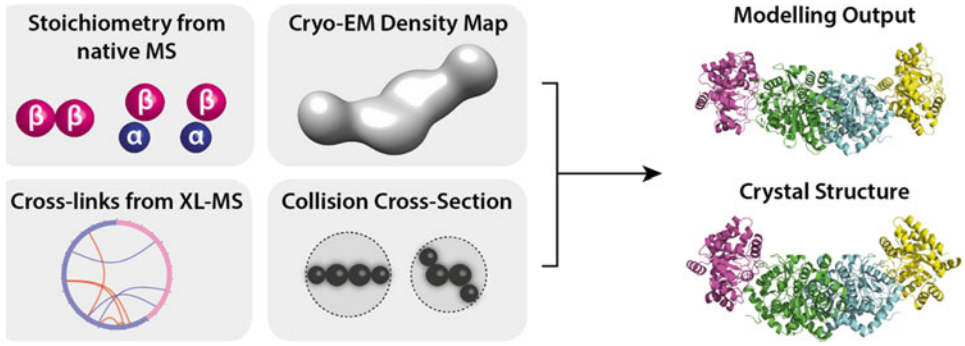


Fig. 1 Integrative modeling with mass spectrometry. Integration of restraints derived from MS-based and cryo-EM data sets lead to computational structural models that can be directly compared with available crystal structures. Such comparisons allow for assessing the ability of the methods used to make accurate predictions

1.1 Native MS

Native MS is a powerful technique for interrogating key structural aspects of proteins and their complexes. It allows examination of the specificity and selectivity of ligands binding to a protein in addition to identifying protein stoichiometry and their oligomeric states [14–16]. The main strengths of native MS are that it can rapidly provide information on the native state of large proteins and their complexes, it does not require high concentrations of pure material, and it can typically deal with heterogeneous samples. As such, native MS has become an integral component of the structural biology toolkit complementing traditional structural methods. Information derived from native MS can be used to model protein complexes [9] and also integrated with data sets from other related methods [1, 2].

1.2 Ion Mobility MS

Native MS is often hyphenated with ion mobility (IM)-MS. IM-MS provides information about protein topologies and dynamics in the gas phase [17, 18]. It is expressed using the collisional cross section (CCS) which reflects the orientationally average cross section of proteins as they tumble through the IM cell in the mass spectrometer [19]. CCS from IM-MS is obtained by measuring the arrival time distributions (ATDs) of different ions [20] and can in turn be used as shape restraints for interrogating structural models generated by computational methods [7, 21, 22]. Structural dynamics and conformational heterogeneity can be inferred from the distribution of mobilities measured by IM-MS [23–25]. Typically, ions with multiple conformations will result in broad, Gaussian ATDs [12].

1.3 Chemical Cross-Linking and Covalent Labelling MS

Chemical cross-linking (CX)-MS allows covalent linkage of reactive amino acid residues enabling information about the protein topology and their fold [6, 26–29]. CX-MS defines protein subunit proximities which can be expressed as distance restraints for downstream modeling analysis [27, 30]. The information can

subsequently be translated into structural restraints, which are then used by a modeling workflow to generate 3D models of protein assemblies. Moreover, several computational methods and tools have been developed to assist the structural interpretation of CX-MS [4, 6, 31]. Another MS-based labeling strategy that can be used for modeling protein complexes includes covalent labeling (CL)-MS. Covalent labeling is a chemical modification technique that allows the mapping of residue-level surface accessibility. It relies on the notion that exposed amino acid residues accessible to the labeling reagents will readily react, while protected (buried) residues respond to a lesser extent [32]. These differences in reactivity can be associated with conformational changes of proteins and therefore be used to infer information regarding protein–protein or protein–ligand interactions [33].

1.4 Hydrogen–Deuterium Exchange MS

HDX-MS has recently gained interest for studying protein conformational dynamics and their associations with ligands [15, 34–36]. HDX-MS can report on change in the structure and dynamics of proteins, through monitoring the exchange of protein backbone amide hydrogens for deuterium in the surrounding labeling buffer. The exchange rate is dependent on both hydrogen bonding and solvent accessibility [37–39]. HDX-MS tolerates a broad range of sample environments and can be used to capture conformational changes of proteins at peptide-level resolution. Data from two protein states (e.g., apo and holo) can be compared in a differential manner, allowing relative changes in protein structure and dynamics to be determined. The resulting significant differences (ΔHDX) between the two states can in turn be mapped onto 3D models of proteins to inform on conformational changes [37] (Fig. 1). Recently HDX-MS has been applied to a range of membrane proteins and their interactions with lipids [36, 40, 41]. Moreover, attempts to model data derived from HDX-MS have utilized protection factors, which can be used as restraints to build model structures directly from ΔHDX [42].

1.5 Integrating with Other Structural Approaches

While MS-based approaches offer a wide range of techniques capable of generating diverse structural information, their integration with other powerful structural methods is often desirable. The most commonly used methods to integrate MS data are X-ray crystallography and cryo-EM. Traditionally, crystallographic information has been the gold standard to either validate the ability of a computational strategy to build accurate models (benchmarking) or be used in conjunction with other methods to refine its conformation. The recent resolution revolution of cryo-EM has yield a game-changing effect in structural biology through its ability to provide atomic-level resolution of proteins previously thought to be insuperable (such as dynamic and membrane proteins), thus significantly expanding the potential number of biological

structures that can be solved to high resolution. It is therefore timely to develop methods that can integrate the data-rich MS-based approaches with cryo-EM.

2 Materials

The following software packages need to be downloaded and installed:

1. PyMOL (https://pymol.org/2/).
2. Chimera (https://www.cgl.ucsf.edu/chimera/).
3. Integrative Modeling Platform (https://integrativemodeling.org/).
4. IMPACT (http://impact.chem.ox.ac.uk/).
   Webservers accessed as part of this protocol:
5. UniProt (https://www.uniprot.org/).
6. BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).
7. T-Coffee (http://tcoffee.crg.cat/).
8. RCSB (http://www.rcsb.org/).
9. PDBe (http://www.ebi.ac.uk/pdbe/node/1).
10. EMDB (https://www.ebi.ac.uk/pdbe/emdb/).
    All tutorial files can be downloaded from:
11. Github (https://github.com/andymlau/MIMB_Integrative_Modelling).

3 Methods

All instructions below assume that tutorial files have been downloaded and that the required software packages have been installed and tested for correct functionality.

3.1 Preparing Inputs for IMP

1. The IMP platform can be used to generate models of protein complexes, informed by experimental techniques which each contribute modeling restraints. To use stoichiometry, cross-linking, and cryo-EM restraints, the following files are needed:
   Files

   initial_model.**pdb**—contains EM-map aligned model of the protein complex.

   sequences.fasta—contains full sequences of each subunit.

   stoichiometry.txt—list of stoichiometric connectivities.

   crosslinks.txt—list of experimental cross-links.

initial_map.gmm.nX.txt—Gaussian mixture model (GMM) representation of the electron density map.

topology.txt—contains topological parameters for initializing each subunit for modeling.

Scripts

modelling.py—script for calling IMP.

calculate_density.sh—bash script for converting density maps to modeling-compatible maps.

create_gmm.py—IMP script for generating the above maps.

2. The following steps will walk through how to prepare these files.

3.2 Sourcing Structures of Constituent Subunits

1. Subunits of the complex may be available as crystal or NMR structures; these may be found at repositories such as RCSB and PDBe. A crystal structure of the tryptophan synthase complex has been deposited under PDB 1WBJ (https://www.rcsb.org/structure/1WBJ). Using PyMOL, enter the "*fetch 1WBJ*" command into the PyMOL command line to automatically download the file from the RCSB database (*see* **Note 1**).
2. PyMOL will display two subunits: alpha (chain A) and beta (chain B). Display the sequence information by pressing the "*S*" button displayed on the bottom right hand side of the Movies Controls or by entering "*set seq_view, on*" into the PyMOL command line.
3. The tryptophan synthase dimer will be coated with ligands and water molecules which will need to be removed prior to modeling. Type "*remove hetatm*" into the PyMOL command line to remove all non-protein atoms.
4. Optionally, type "*util.cbc*" to enable per-chain coloration to aid visualization of the subunits. The alpha subunit (chain A) will be colored green, and beta subunit (chain B) in blue by default. A screenshot of your PyMOL GUI at this point is shown in Fig. 2a.
5. With the sequence view on, check the following:

   Sequence numbering—check the sequence numbering to verify that the numbers are correct. Additionally, locate on the structure the residues involved in cross-linking and verify that their residue numbers are correct, e.g., check that residue LYS46 on the cross-link list correctly points to LYS46 on the structure (*see* **Note 2**).

   Residue type—only certain residue types will be "visible" in the experiment depending on the type of cross-linker used, e.g., BS3 (bis(sulfosuccinimidyl)suberate) is an amine-to-amine cross-linker only targets lysine and N-terminal amine

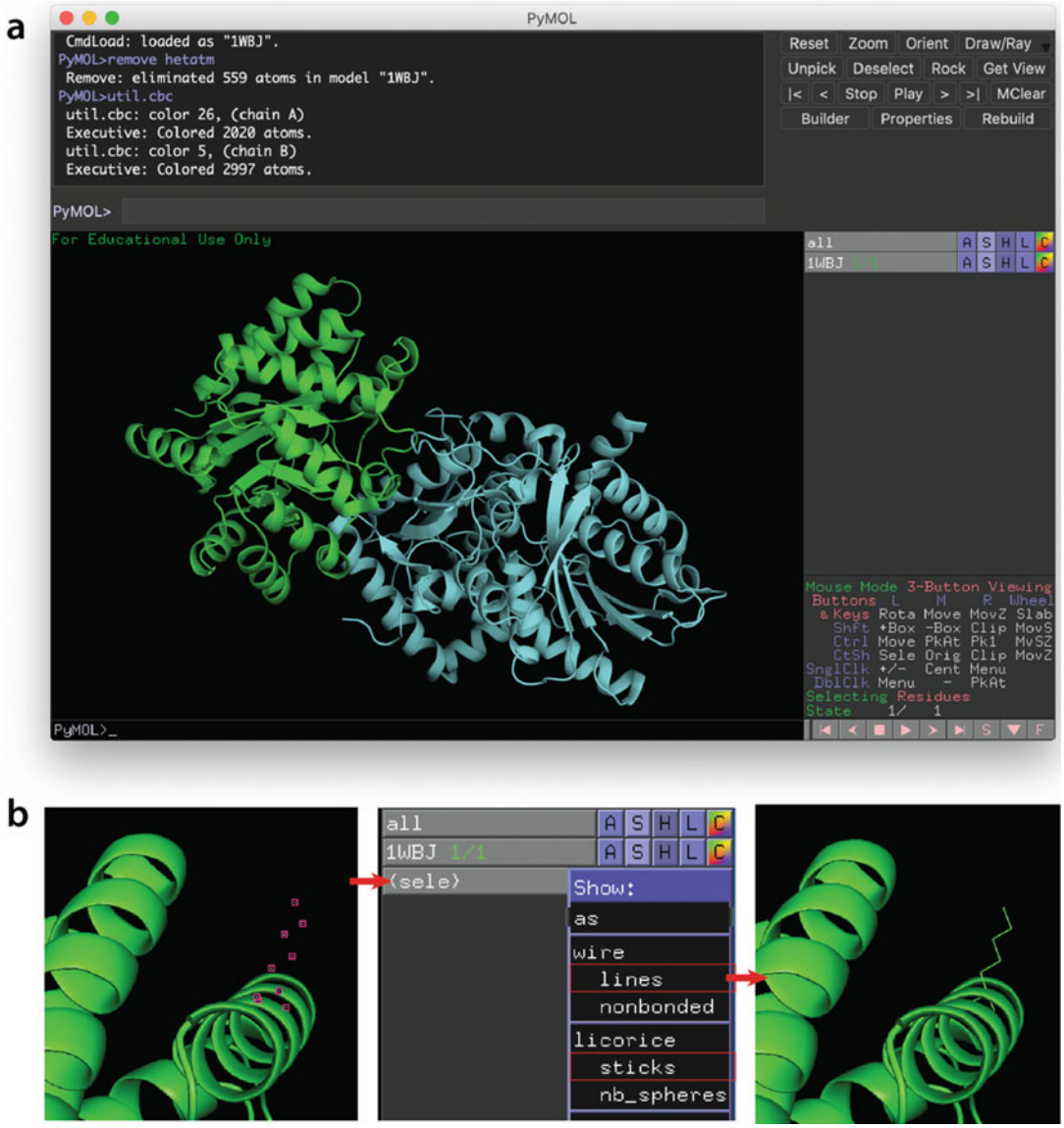


Fig. 2 Visualization of protein structures using PyMOL. (**a**) Ribbon representation of the PDB 1WBJ tryptophan synthase. (**b**) Display of atomic bonds for a selected residue

groups. Check that all cross-linked residues on the model structure are the correct residue type.

Missing atoms—residue sidechains may be missing atoms for various reasons, such as poor electron density from protein crystallography. While checking the residue numbering and type, also check if any of these sidechains are missing atoms (*see* **Note 3**). By default, proteins opened in PyMOL version 2 and above will be represented in cartoon format. To display full atomic information, you will need to select the residue of

interest, and display it in line or stick format. To do this, select the residue by clicking on it in the sequence viewer, click on "S" in the Object Control Panel for "(sele)" and press either "lines" or "sticks" (Fig. 2b).

6. Next, you will need to identify any missing residues in each of the subunits. This can be done by aligning the sequence of each subunit to their canonical sequences (from UniProt) using a multiple sequence alignment program such as T-Coffee (*see* **Notes 4** and **5**).
7. First, save the 1WBJ PDB file in the FASTA format in PyMOL by entering both "save initial_model_A.fasta, chain A" and "save initial_model_B.fasta, chain B" into the PyMOL command line. This will save the sequences of both chain A (alpha subunit) and chain B (beta subunit) separately as FASTA files.
8. Access the UniProt entry of the tryptophan synthase alpha subunit (Accession Number (AN): P00929) and beta subunit (AN: P0A2K1) at: https://www.uniprot.org/uniprot/P00929 and https://www.uniprot.org/uniprot/P0A2K1, respectively.
9. The sequence of each subunit is listed under the "Sequence" tab of the side bar. Click on the "FASTA" download button to open the sequence in the FASTA format.
10. Next, we will align the model and canonical sequences using the T-Coffee webserver. Copy and paste the model and canonical sequences of the alpha chain into T-Coffee and submit. Repeat for the beta chain. Sample results are shown in Fig. 3a.
11. Regions with differences between theoretical and modeled sequences will be identified by sequence alignment, including missing residues (Fig. 3a).
12. If there are many regions of the model with missing coordinates, homology modeling may need to be employed to regenerate these missing residues (*see* **Note 6**).
13. Finally, combine the "initial_model_A.fasta" and "initial_model_B.fasta" sequences into a single "initial_model.fasta." Since there are two copies of alpha and beta subunits in the complex, duplicate both the alpha and beta subunit sequences, and title each subunit as ">TS_alpha_A," ">TS_beta_B," ">TS_alpha_C," and ">TS_beta_D" (Fig. 3b).

3.3 Determining Complex Stoichiometry Using Native MS

1. The native MS spectra of the tryptophan synthase complex can be used to determine any subcomplexes that build up the full complex, and the mass of these subcomplexes (Fig. 4).
2. A native MS spectra of the complex can be found at [33].
3. Use the subcomplexes to determine the known connections between the constituent subunits (*see* **Notes 7** and **8**). The

a

```
T-COFFEE, Version_11.00.d625267 (2016-01-
11 15:25:41 - Revision d625267 - Build 507)
Cedric Notredame
CPU TIME:0 sec.
SCORE=998
*
 BAD AVG GOOD
*
sp|P0A2K1|TRPB_    :  99
TS_A_A             :  99
cons               :  99


sp|P0A2K1|TRPB_    MTTLLNPYFGEFGGMYVPQILMPALNQLEEAFVSAQKDPEFQAQFADLLKNYAGRPTALTKCQ
TS_A_A             MS-LLNPYFGEFGGMYVPQILMPALNQLEEAFVSAQKDPEFQAQFADLLKNYAGRPTALTKCQ
cons               *: ************************************************************
```

b

sequences.fasta - Notepad

File Edit Format View Help

```
>TS_alpha_A
MERYENLFAQLNDRREGAFVPFVTLGDPGIEQSLKIIDTLIDAGADALELGVPFSDPLAD
GPTIQNANLRAFAAGVTPAQCFEMLALIREKHPTIPIGLLMYANLVFNNGIDAFYARCEQ
VGVDSVLVADVPVEESAPFRQAALRHNIAPIFICPPNADDDLLRQVASYGRGYTYLLSRS
GVTGAENRGALPLHHLIEKLKEYHAAPALQGFGISSPEQVSAAVRAGAAGAISGSAIVKI
IEKNLASPKQMLAELRSFVSAMKAASRA

>TS_beta_B
MTTLLNPYFGEFGGMYVPQILMPALNQLEEAFVSAQKDPEFQAQFADLLKNYAGRPTALT
KCQNITAGTRTTLYLKREDLLHGGAHKTNQVLGQALLAKRMGKSEIIAETGAGQHGVASA
LASALLGLKCRIYMGAKDVERQSPNVFRMRLMGAEVIPVHSGSATLKDACNEALRDWSGS
YETAHYMLGTAAGPHPYPTIVREFQRMIGEETKAQILDKEGRLPDAVIACVGGGSNAIGM
FADFINDTSVGLIGVEPGGHGIETGEHGAPLKHGRVGIYFGMKAPMMQTADGQIEESYSI
SAGLDFPSVGPQHAYLNSIGRADYVSITDDEALEAFKTLCRHEGIIPALESSHALAHALK
MMREQPEKEQLLVVNLSGRGDKDIFTVHDILKARGEI

>TS_alpha_C
MERYENLFAQLNDRREGAFVPFVTLGDPGIEQSLKIIDTLIDAGADALELGVPFSDPLAD
GPTIQNANLRAFAAGVTPAQCFEMLALIREKHPTIPIGLLMYANLVFNNGIDAFYARCEQ
VGVDSVLVADVPVEESAPFRQAALRHNIAPIFICPPNADDDLLRQVASYGRGYTYLLSRS
GVTGAENRGALPLHHLIEKLKEYHAAPALQGFGISSPEQVSAAVRAGAAGAISGSAIVKI
IEKNLASPKQMLAELRSFVSAMKAASRA

>TS_beta_D
MTTLLNPYFGEFGGMYVPQILMPALNQLEEAFVSAQKDPEFQAQFADLLKNYAGRPTALT
KCQNITAGTRTTLYLKREDLLHGGAHKTNQVLGQALLAKRMGKSEIIAETGAGQHGVASA
LASALLGLKCRIYMGAKDVERQSPNVFRMRLMGAEVIPVHSGSATLKDACNEALRDWSGS
YETAHYMLGTAAGPHPYPTIVREFQRMIGEETKAQILDKEGRLPDAVIACVGGGSNAIGM
FADFINDTSVGLIGVEPGGHGIETGEHGAPLKHGRVGIYFGMKAPMMQTADGQIEESYSI
SAGLDFPSVGPQHAYLNSIGRADYVSITDDEALEAFKTLCRHEGIIPALESSHALAHALK
MMREQPEKEQLLVVNLSGRGDKDIFTVHDILKARGEI
```

Fig. 3 Sequence analysis. (**a**) T-Coffee sequence alignment of canonical (AN: P0A2K1) and model sequence (TS_A_A) of tryptophan synthase beta chain. The second and third residue of TS_A_A (PDB 1WBJ) has been altered to demonstrate the conservation notation (cons). Asterisk (*) and colon (:) denote full and similar property conservation, respectively. Only the first line of the sequence has been shown. (**b**) Alpha and beta subunit sequence in fasta format

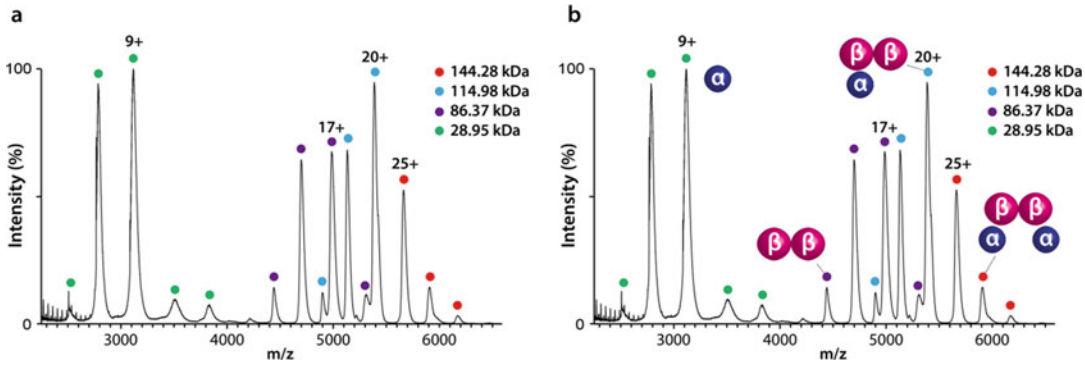


Fig. 4 Native MS of tryptophan synthase. (**a**) Native mass spectra showing four ion distributions of 144, 115, 86, and 29 kDa. (**b**) Annotated spectra of (**a**) with assigned subcomplexes

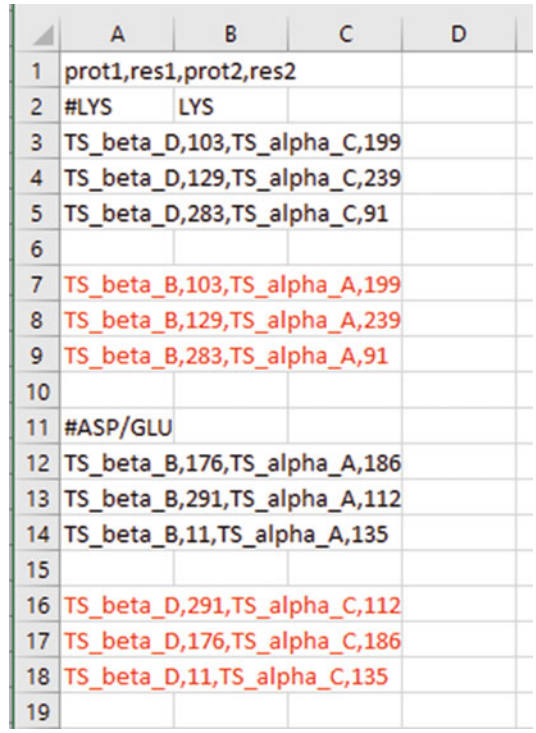
simplest non-monomeric subcomplex has been identified to be a beta subunit dimer. This indicates a beta–beta connection in the complex. A single alpha subunit associates with the beta dimer, forming an alpha–beta–beta subcomplex. This indicates an alpha–beta connection. As there are no alpha dimers seen, we can assume that these do not exist. The connectivity of the complex can therefore be broken down to alpha–beta and beta–beta subunits.

3.4 Preparation of Cross-Link List for IMP

1. A list of cross-links of the tryptophan synthase complex can be found at [33]. To demonstrate a heterogeneous setup of LYS-LYS and ASP/GLU-ASP/GLY cross-links, we have simulated a set of cross-links for this protocol.
2. The cross-links to be used in the modeling have been prepared as "crosslinks.txt" and are in the format shown in Fig. 5, with each cross-link entry listed on a separate line, and each protein and residue separated by commas. The "prot1,res1,prot2,res2" header at the top of the file tells the IMP parser how to interpret the entries.
3. Since there are redundant copies of each alpha and beta subunit in the experiment, the cross-links must be duplicated for each alpha–beta and beta–beta pair as determined by native MS (Fig. 5).

3.5 Preparation of Stoichiometry List for IMP

1. Having determined the subunit connections of the complex via native MS, we will need to convert these into modelable restraints.
2. In the same format as the "crosslinks.txt" file, we can include a pseudo-cross-link between the alpha–beta and beta–beta subunits of the complex. As there are two alpha–beta pairs, we will enforce a pseudo-cross-link between TS_alpha_A-TS_beta_B and TS_alpha_C-TS_beta_D and the beta–beta connection as



	A	B	C	D
1	prot1,res1,prot2,res2			
2	#LYS	LYS		
3	TS_beta_D,103,TS_alpha_C,199			
4	TS_beta_D,129,TS_alpha_C,239			
5	TS_beta_D,283,TS_alpha_C,91			
6				
7	TS_beta_B,103,TS_alpha_A,199			
8	TS_beta_B,129,TS_alpha_A,239			
9	TS_beta_B,283,TS_alpha_A,91			
10				
11	#ASP/GLU			
12	TS_beta_B,176,TS_alpha_A,186			
13	TS_beta_B,291,TS_alpha_A,112			
14	TS_beta_B,11,TS_alpha_A,135			
15				
16	TS_beta_D,291,TS_alpha_C,112			
17	TS_beta_D,176,TS_alpha_C,186			
18	TS_beta_D,11,TS_alpha_C,135			
19				

Fig. 5 Simulated cross-links of the tryptophan synthase complex. Cross-links in black (3–5 and 12–14) have been duplicated (red) to reflect the twofold symmetry nature of the full complex. Not all cross-links are shown

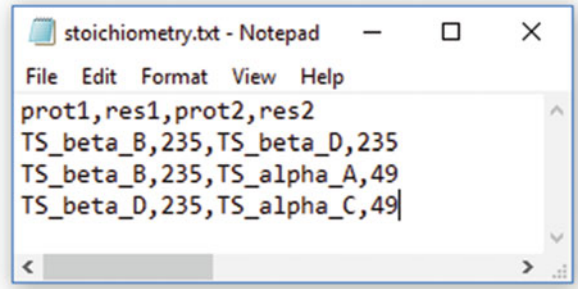


Fig. 6 Stoichiometry of the tryptophan synthase complex. Pseudo-crosslinks between subunits of the complex

TS_beta_B-TS_beta_D. For the residue numbers, provide the center-most residue identified from the 1WBJ PDB file. The center-most residues are residue 49 for alpha subunits and residue 235 for beta subunits. The content of the pseudo-cross-link file is shown in Fig. 6.

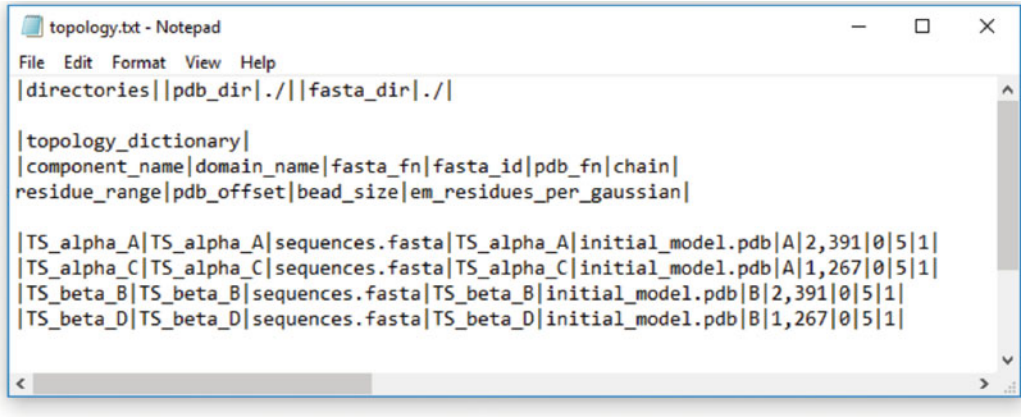
3. The cryo-EM density maps of protein complexes are stored and accessed from the EMDB database.

4. Since no density maps are available for the tryptophan synthase complex, a simulated map at 14 Å "TS_1WBJ_14A.mrc" has been generated and included in the protocol files.
5. All subunits will need to be fitted into the map of the density for the map to restrain the subunit positions during sampling. For this protocol, the structure has been pre-fitted into the mrc map (*see* **Notes 9** and **10**).
6. The mrc density map will need to be converted into a modeling-compatible Gaussian Mixture Model (GMM) format.
7. To do this, have the "initial_map.mrc," "calculate_density.sh" and "create_gmm.py" files in the same directory, and open a terminal window, e.g., Terminal in MacOS/Linux or Bash for Windows.
8. Enter "bash create_gmm.py initial_map.mrc 6" to automatically begin converting the simulated mrc map into a GMM map (*see* **Notes 11** and **12**). The number at the end refers to the number of components that the map will possess—the higher the value, the finer the resolution of the generated map model. The files "initial_map.gmm.n6.mrc" and "initial_map.gmm.n6.txt" will be generated. Optionally open the new and original maps to visualize the conversion.

3.6 Preparing Modeling Script for IMP

1. Open the topology.txt file which has already been filled (Fig. 7a).
2. The topology file contains a "topology_dictionary" for each subunit. Table 1 explains the function of each heading.
3. In the previous directory, there will be a script called "modelling.py" which will call IMP and perform the modeling.
4. "modelling.py" is a python script that will instruct IMP on how to initialize each subunit and perform random sampling while restraining their positions and orientations using the stoichiometry, cross-link, and EM restraints.
5. Within "modelling.py," lines 25–27 under the "Define Input Files" section will link IMP to the necessary "topology.txt" and "initial_map.gmm.n6.txt" files which are located in the "./inputs/" directory (Fig. 7b).
6. The next lines "num_frames" and "num_mc_steps" will set the number of iterations that the sampling will be performed for and the number of Monte Carlo steps that are included in one iteration.
7. "rb_max_trans" and "rb_max_rot" are variables for controlling the step size of each translation and rotation of each rigid body, i.e., each subunit.

a



b

```
#---------------------------
# Define Input Files
#---------------------------
datadirectory = "./inputs/"
topology_file = datadirectory+"topology.txt"
target_gmm_file = datadirectory+"initial_map.gmm.n6.txt"
```

c

```
[["TS_alpha_A"],["TS_beta_B"],["TS_alpha_C"],["TS_beta_D"]] # independent
[["TS_alpha_A","TS_beta_B"],["TS_alpha_C","TS_beta_D"]] # alpha-beta grouped
[["TS_alpha_A","TS_beta_B","TS_beta_D"],["TS_alpha_C"]] # alpha-beta-beta grouped
```

Fig. 7 Preparing the IMP modeling script. (**a**) Content of the topology file topology.txt. (**b**) Definition of input files. (**c**) Examples of rigid body groups

Table 1
Function of headings from the topology file

Heading	Function
component_name	Name of subunit
domain_name	Subunit identifier for modelling
fasta_fn	Fasta filename
fasta_id	References the subunit sequence listed in specified file
pdb_fn	References the PDB file containing all structures
Chain	References the chain belonging to the subunit in the PDB specified
residue_range	Obtains the atomic position of the residues within range from the PDB
pdb_offset	Offsets the PDB by the number specified
bead_size	Number of residues to represent using a single bead for coarse-graining
em_residues_per_gaussian	Number of components for GMM model generation

8. If there are missing residues in the initial subunit, i.e., inherited from crystal or NMR structures, IMP will replace these missing residues using flexible beads. The movement of these beads are controlled by the "bead_max_trans" variable (*see* **Note 13**).
9. A list of subunits will need to be given to IMP via a "rigid_bodies" list. In the current format, each subunit will be permitted translational and rotational movement independently from the other subunits. Several examples of rigid body groups that can be set are shown in Fig. 7c.
10. Next, lines 57–102 are needed to set the coarse-grained representation of the subunits. Each subunit listed in "topology.txt" will be replaced by a bead model which will take into account the "bead_size" parameter as its resolution.
11. Each subunit is automatically given a color for loop on line 74.
12. The ability of each subunit and flexible bead to move is set up by lines 93–102.
13. Finally, the restraint section "Define Scoring Function Components" defines the variables necessary for restraint-guided modeling.

3.7 Modeling Restraints

1. The first restraint is the excluded volume restraint—this prevents unrealistic overlapping in space through calculating the volume of each subunit per iteration and avoiding steps which promote overlap.
2. The excluded volume restraint is computationally expensive to calculate, and thus, the volume is approximated by bead resolution (by default set to 20) on line 115.
3. Lines 117 and 118 add the restraint to the bead model.
4. Next, we will add two cross-link restraints to our tryptophan synthase system, xl1 and xl2.
5. xl1 is the stoichiometry restraint which will enforce alpha–beta–beta–alpha connectivity in our modeling. The xl1 restraint can be adjusted by variables listed in Table 2.
6. The "length" variable passed into xl1 is defined by the approximate distance between the alpha and beta subunits as measured from their centers (approximately equal to the sum radius of alpha and beta subunits). This distance is approximate and serves to enforce the topology of the system by preventing the subunits from moving too far away from their designated binding partners.
7. The xl2 restraint includes cross-links determined for the tryptophan synthase complex. Lines 146–158 are parametrized identically to xl1; however, the distance threshold "length," "slope," and "label" has been altered.

Table 2
Key parameters that can be used for the adjustment of modeling restraints

Heading	Function
Length	Distance threshold to enforce per iteration, set to 38 Å
Slope	Linear gradient for preventing the system from expanding too much
Columnmapping	Designates the format of the stoichiometry.Txt file
Resolution	Sets the bead resolution that the cross-link distance will be measured using
Label	Label for modeling output
Csvfile	Toggle for stoichiometry.Txt file format to be read as a comma separated values file

8. xl2 is set up with a distance threshold of 35 Å. This distance is derived from the BS3 cross-linker molecule length (11 Å), two lysine side chains (2 × 7 Å) and an additional 10 Å which permits small-scale subunit translations and rotations and prevents the system from being locked into a particular conformation.
9. The EM restraint is set by lines 171–178. In the first step, a GMM map is calculated for each subunit. This map will be calculated against the GMM of the original complex map. The EM restraint also has a "slope" variable which will persuade movement of subunits to remain inside the map volume.
10. Finally, the mc1 variable will run the sampling of the subunit bead models which have restraints added. Mc1 is an IMP replica exchange macro which will additionally perform replica exchange steps on flexible regions of the bead model. These parameters will only affect flexible beads (*see* **Note 14**).
11. "crosslink_restraints" in mc1 adds the two xl1 and xl2 restraints to the replica exchange sampler.
12. A Bayesian scoring function is used to calculate the score of each model per iteration (*see* **Note 15**). "number_of_best_scoring_models" controls how many output models are generated. These are saved within the "global_output_directory" directory. The top scoring models are saved in the /output/pdbs folder by default. Model names are updated and ranked such that model.0.pdb has the highest score of the ensemble.
13. The per-iteration trajectory of the sampling simulation is saved as /output/rmfs/0.rmf3. This file can be opened and played using Chimera.
14. The PDB models generated by IMP only contain Ca atoms. Since there are no missing residues (that have been modeled as flexible beads), use PyMOL to align the alpha and beta subunits to model.0.pdb (*see* **Note 16**). You can then save this atomistic model.

15. A PyMOL script has been provided to streamline the conversion between coarse-grained and atomistic models, called "CG2pdb.py" in the same directory as "modelling.py."
16. Open a terminal window and direct to the directory of "CG2pdb.py." Run the script using "<path to pymol binary>-c CG2pdb.py." The <path to pymol binary> will need to be changed to the path of PyMOL on your system. The "-c" flag will suppress the GUI of PyMOL.
17. The directory "output_atomistic" will be made and will contain all 100 best scoring IMP models in atomistic representation. By default, the CG2pdb.py script will convert from model.0.pdb to model.99.pdb (100 models). You may alter the model range within the script by changing "n_first" and "n_last" variables.
18. At this point, you will have a folder called "output_atomistic" containing 100 atomistic models of tryptophan synthase that have been generated using modeling restraints from MS and EM.

3.8 Calculating the CCS of Generated Models

1. The theoretical collision cross section (CCS) of atomic models can be calculated using software such as IMPACT [43] (*see* **Notes 17–19**).
2. This value can be compared to experimentally determined CCS via IM-MS [44].
3. Alternatively, CCS values for complexes may be found in published literature [7].
4. The CCS of the tryptophan synthase complex is approximately 7200 $Å^2$.
5. We will use IMPACT to calculate the projection approximation (PA) of each model.
6. Open a terminal window and direct to the "output_atomistic" folder.
7. With IMPACT callable from your PATH variable, run "impact *.pdb -o output_file_CCS.txt." *.pdb will tell IMPACT to run through file with the .pdb file extension.
8. When IMPACT has finished, output_file_CCS.txt will display a list of each file, with the calculated PA and TJM values (*see* **Note 20**).
9. Linear scaling of the calculated PA by an empirical factor of 1.14 has been shown to provide a good approximation of the true model CCS [45].
10. "output_file_CCS.txt" can be opened using spreadsheet software, and a scaling factor of 1.14 can be applied to each value under "CCS_PA."

11. Calculate the CCS % error between the model (CCSmodel) and experimental CCS (CCSexp) for all tryptophan synthase models using:

$$\%\ \text{error} = \frac{\left|\text{CCS}_{\text{model}} - \text{CCS}_{\text{exp}}\right|}{\text{CCS}_{\text{exp}}} \times 100$$

12. The best-fit models will show the lowest % error. This conformation will have been generated using stoichiometry, cross-linking, and EM restraints and have been filtered using CCS.
13. Finally, evaluate the accuracy of your model against the crystal structure of the full tetrameric tryptophan synthase complex. Open PyMOL and use the "fetch 1WBJ, type=pdb1, multiplex=1" command to automatically download the "biological assembly" of 1WBJ.
14. 1WBJ will download as two objects in PyMOL, 1WBJ_0001 and 1WBJ_0002 which will each be half of the complex. Both alpha chains will be labeled as Chain A, and both beta as Chain B.
15. To align our best-fit model to this complex, we will first need to combine the two halves of 1WBJ into a single object, with the same Chain ABCD arrangement.
16. Use the command "alter 1WBJ_0002 & chain A, chain='C'" and "alter 1WBJ_0002 & chain B, chain='D'" to rename to chains C and D.
17. Next, use "create 1WBJ_assembled, 1WBJ_*" to create a new object with all chains of the tryptophan synthase biological assembly.
18. Align your best fit model to this 1WBJ_assembled object using either "align <best_model_name> 1WBJ_assembled" or by clicking on screen in the PyMOL GUI.
19. In the External GUI, the root mean squared deviation (RMSD) of the fitting will be displayed (Fig. 8). The RMSD represents the similarity score between the two models. The lower the RMSD, the more similar the models.

```
Match: read scoring matrix.
Match: assigning 1314 x 1314 pairwise scores.
MatchAlign: aligning residues (1314 vs 1314)...
MatchAlign: score 6736.000
ExecutiveAlign: 1314 atoms aligned.
Executive: RMSD =    8.221 (1314 to 1314 atoms)
Executive: object "aln_00_atomistic_to_sele" created.
```

Fig. 8 Scoring RMSD between generated model and crystal structure

20. The RMSD value can be improved by either improving the quality of the restraints (i.e., additional experimental information), or by fine tuning the sampling parameters. The number of iterations can also be increased to maximize the search space of the program.

4 Notes

1. A comprehensive PyMOL tutorial can be found at: https://www.mrc-lmb.cam.ac.uk/rlw/text/MacPyMOL_tutorial.html.
2. The sequence of the protein can be edited using PyMOL through the command:

```
alter <selection>, resv += X
```

   where <selection> should be modified to reflect the object or chain to be altered, and the plus sign should be swapped to minus if X is to be deducted from the sequence number. When using this command, <selection> can refer to any selection, e.g. the object name, segment ID, chain ID or the currently highlighted selection. For more information, consult the PyMOL documentation for the Alter command (https://pymolwiki.org/index.php?title=Alter&redirect=no).
3. For cross-linking modeling, residues will need to have the atom that will be involved in distance measurements. For example, LYS-LYS cross-links can be measured from Cα, Cβ, or Nζ atoms. If using the sidechain amide Nζ atom, you will need to have this atom included in all sidechain models. The Mutagenesis tool of PyMOL can be used to quickly regenerate all sidechain atoms, swap between common sidechain rotamers or switch residue types (https://pymolwiki.org/index.php/Mutagenesis).
4. FASTA is a text-based format for displaying sequence information of proteins or nucleotides (*see* also https://zhanglab.ccmb.med.umich.edu/FASTA/). In PyMOL, you can save the sequence information in FASTA format for any object or selection using the "save" command (https://pymolwiki.org/index.php/Save).
5. T-Coffee is a quick and easy-to-use multiple sequence alignment tool [46]. A tutorial is available on the webserver website at: http://tcoffee.crg.cat/apps/tcoffee/tutorial.html.
6. Homology modeling can be performed using the MODELLER software [47]. A detailed tutorial can be found at: https://salilab.org/modeller/tutorial/.

7. If your protein of interest undergoes post translational processing, the canonical sequence reported on UniProt may be different to the true experimental mass. Additional sequence information can be found on UniProt for characterized entries.
8. Identified subcomplexes can be used to determine the connectivity based on its internal constituents. Start from the smallest subcomplex and build up a network of additional subunits.
9. While density maps can be opened in PyMOL, the graphics viewer Chimera is better suited for visualizing density maps and possessed advanced functions for fitting structures into maps, etc.
10. Structures can be fitted into density maps using Chimera's "Fit in Map" function. Open both the map and structures in Chimera. "Fit in Map" can be accessed from the Volume Viewer window when a map file is open, under the "Tools" tab. You may need to manually move the structure in close proximity of the map file for better fitting.
11. The calculate_density.sh script can be opened using any text editor. You will need to replace the "python="/anaconda2/bin/python" line with the path to python for IMP on your own system.
12. GMMs can be calculated for electron density maps as a sum of three-dimensional gaussians. The shape of a structure in PDB format can also be converted into a simulated volume, and likewise converted into a GMM. The fit between each subunit GMM can then be rapidly calculated against the GMM of the complex. Consult the IMP documentation for more information on how GMMs are used.
13. The number of residues that are represented by a single bead is set in the "topology.txt" file under "bead_size."
14. See IMP documentation at: https://integrativemodeling.org/talks/dec_2016_workshop/IMP%20software%20introduction%20and%20tutorial.pdf for additional information on replica exchange simulations.
15. See IMP documentation at: https://integrativemodeling.org/talks/dec_2016_workshop/Modeling%20of%20the%20Nup84%20complex.pdf for more information on scoring functions.
16. Consult the PyMOL documentation for the "Align" function: https://pymolwiki.org/index.php/Align.
17. CCS can be calculated through four different metrics, these are the projection approximation (PA), projection superposition approximation (PSA) [48], exact hard sphere scattering (EHSS), and trajectory method (TM or TJM). These metrics vary in accuracy and computational speed *see* also [19].

18. Since proteins and complexes can exhibit conformational variations, the reported CCS may vary depending on these conformations [49].
19. The CCS of some proteins may not reflect their "native" conformations or those calculated from their crystal structures, due to structural collapse in the gas phase. This phenomenon commonly affects systems which are intrinsically flexible or unstructured [50–52].
20. The TJM value reported by IMPACT is an approximation of the true TJM value [43].

References

1. Alber F et al (2007) Determining the architectures of macromolecular assemblies. Nature 450:683–694. https://doi.org/10.1038/nature06404
2. Ahdash Z, Pyle E, Politis A (2016) Hybrid mass spectrometry: towards characterization of protein conformational states. Trends Biochem Sci 41:650–653. https://doi.org/10.1016/j.tibs.2016.04.008
3. Bullock JMA, Sen N, Thalassinos K, Topf M (2018) Modeling protein complexes using restraints from crosslinking mass spectrometry. Structure 26:1015–1024.e1012. https://doi.org/10.1016/j.str.2018.04.016
4. Degiacomi MT, Schmidt C, Baldwin AJ, Benesch JLP (2017) Accommodating protein dynamics in the modeling of chemical crosslinks. Structure 25:1751–1757.e1755. https://doi.org/10.1016/j.str.2017.08.015
5. Lasker K et al (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. Proc Natl Acad Sci U S A 109:1380–1387. https://doi.org/10.1073/pnas.1120559109
6. Politis A et al (2014) A mass spectrometry-based hybrid method for structural modelling of protein complexes. Nat Methods 11:403–406
7. Hall Z, Politis A, Robinson CV (2012) Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. Structure 20:1596–1609. https://doi.org/10.1016/j.str.2012.07.001
8. Hansen K et al (2018) A mass spectrometry-based modelling workflow for accurate prediction of IgG antibody conformations in the gas phase. Angew Chem Int Ed Engl 57 (52):17194–17199
9. Politis A et al (2015) Topological models of heteromeric protein assemblies from mass spectrometry: application to the yeast eIF3:eIF5 complex. Chem Biol 22:117–128. https://doi.org/10.1016/j.chembiol.2014.11.010
10. Russel D et al (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 10:e1001244. https://doi.org/10.1371/journal.pbio.1001244
11. Faull SV et al (2019) Structural basis of Cullin 2 RING E3 ligase regulation by the COP9 signalosome. Nat Commun 10:3814. https://doi.org/10.1038/s41467-019-11772-y
12. Zhou M et al (2014) ATP binding reduces conformational flexibility in a rotary ATPase—Evidence from ion mobility mass spectrometry. Nat Chem 6(3):208–215
13. Alber F et al (2007) The molecular architecture of the nuclear pore complex. Nature 450:695–701. https://doi.org/10.1038/nature06405
14. Bechara C et al (2015) A subset of annular lipids is linked to the flippase activity of an ABC transporter. Nat Chem 7:255–262. https://doi.org/10.1038/nchem.2172
15. Bechara C, Robinson CV (2015) Different modes of lipid binding to membrane proteins probed by mass spectrometry. J Am Chem Soc 137:5240–5247. https://doi.org/10.1021/jacs.5b00420
16. Ahdash Z et al (2017) Mechanistic insight into the assembly of the HerA-NurA helicase-nuclease DNA end resection complex. Nucleic Acids Res 45:12025–12038. https://doi.org/10.1093/nar/gkx890
17. Pagel K, Natan E, Hall Z, Fersht AR, Robinson CV (2013) Intrinsically disordered p53 and its complexes populate compact conformations in the gas phase. Angew Chem Int Ed Engl

52:361–365. https://doi.org/10.1002/anie.201203047

18. Konijnenberg A et al (2014) Global structural changes of an ion channel during its gating are followed by ion mobility mass spectrometry. Proc Natl Acad Sci U S A 111:17170–17175. https://doi.org/10.1073/pnas.1413118111
19. Gabelica V, Marklund E (2018) Fundamentals of ion mobility spectrometry. Curr Opin Chem Biol 42:51–59. https://doi.org/10.1016/j.cbpa.2017.10.022
20. Jurneczko E, Barran PE (2011) How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. Analyst 136:20–28. https://doi.org/10.1039/c0an00373e
21. Politis A, Park AY, Hall Z, Ruotolo BT, Robinson CV (2013) Integrative modelling coupled with ion mobility mass spectrometry reveals structural features of the clamp loader in complex with single-stranded DNA binding protein. J Mol Biol 425:12. https://doi.org/10.1016/j.jmb.2013.04.006
22. Politis A et al (2010) Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes. PLoS One 5:e12080. https://doi.org/10.1371/journal.pone.0012080
23. Jenner M et al (2011) Detection of a protein conformational equilibrium by electrospray ionisation-ion mobility-mass spectrometry. Angew Chem Int Ed Engl 50:8291–8294. https://doi.org/10.1002/anie.201101077
24. Jurneczko E et al (2013) Probing the conformational diversity of cancer-associated mutations in p53 with ion-mobility mass spectrometry. Angew Chem Int Ed Engl 52:4370–4374. https://doi.org/10.1002/anie.201210015
25. Pacholarz KJ et al (2014) Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling. Angew Chem Int Ed Engl 53:7765–7769. https://doi.org/10.1002/anie.201402863
26. O'Reilly FJ, Rappsilber J (2018) Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. Nat Struct Mol Biol 25(11):1000–1008. https://doi.org/10.1038/s41594-018-0147-0
27. Leitner A et al (2010) Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. Mol Cell Proteomics 9:1634–1649. https://doi.org/10.1074/mcp.R000001-MCP201
28. Sinz A (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass Spectrom Rev 25:663–682
29. Sinz A (2018) Cross-linking/mass spectrometry for studying protein structures and protein-protein interactions: where are we now and where should we go from here? Angew Chem Int Ed Engl 57:6390–6396. https://doi.org/10.1002/anie.201709559
30. Politis A, Schmidt C (2018) Structural characterisation of medically relevant protein assemblies by integrating mass spectrometry with computational modelling. J Proteome 175:34–41. https://doi.org/10.1016/j.jprot.2017.04.019
31. Bullock JMA, Thalassinos K, Topf M (2018) Jwalk and MNXL web server: model validation using restraints from crosslinking mass spectrometry. Bioinformatics 34:3584–3585. https://doi.org/10.1093/bioinformatics/bty366
32. Mendoza VL, Vachet RW (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. Mass Spectrom Rev 28:785–815. https://doi.org/10.1002/mas.20203
33. Schmidt C et al (2017) Surface accessibility and dynamics of macromolecular assemblies probed by covalent labeling mass spectrometry and integrative modeling. Anal Chem 89:1459–1468. https://doi.org/10.1021/acs.analchem.6b02875
34. Wang SC et al (2010) Ion mobility mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing. J Am Chem Soc 132:15468–15470. https://doi.org/10.1021/ja104312e
35. Hopper JT et al (2013) Detergent-free mass spectrometry of membrane protein complexes. Nat Methods 10:1206–1208. https://doi.org/10.1038/nmeth.2691
36. Adhikary S et al (2017) Conformational dynamics of a neurotransmitter:sodium symporter in a lipid bilayer. Proc Natl Acad Sci U S A 114:E1786–e1795. https://doi.org/10.1073/pnas.1613293114
37. Mistarz UH, Brown JM, Haselmann KF, Rand KD (2016) Probing the binding interfaces of protein complexes using gas-phase H/D exchange mass spectrometry. Structure 24:310–318. https://doi.org/10.1016/j.str.2015.11.013
38. Marcsisin SR, Engen JR (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us? Anal Bioanal Chem

397:967–972. https://doi.org/10.1007/s00216-010-3556-4

39. Hebling CM et al (2010) Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry. Anal Chem 82:5415–5419. https://doi.org/10.1021/ac100962c
40. Eisinger ML, Dorrbaum AR, Michel H, Padan E, Langer JD (2017) Ligand-induced conformational dynamics of the *Escherichia coli* Na(+)/H(+) antiporter NhaA revealed by hydrogen/deuterium exchange mass spectrometry. Proc Natl Acad Sci U S A 114:11691–11696. https://doi.org/10.1073/pnas.1703422114
41. Martens C et al (2018) Direct protein-lipid interactions shape the conformational landscape of secondary transporters. Nat Commun 9:415
42. Claesen J, Politis A (2019) POPPeT: a new method to predict the protection factor of backbone amide hydrogens. J Am Soc Mass Spectrom 30(1):67–76. https://doi.org/10.1007/s13361-018-2068-x
43. Marklund EG, Degiacomi MT, Robinson CV, Baldwin AJ, Benesch JL (2015) Collision cross sections for structural proteomics. Structure 23:791–799. https://doi.org/10.1016/j.str.2015.02.010
44. Ruotolo BT, Benesch JL, Sandercock AM, Hyung SJ, Robinson CV (2008) Ion mobility-mass spectrometry analysis of large protein complexes. Nat Protoc 3:1139–1152. https://doi.org/10.1038/nprot.2008.78
45. Benesch JL, Ruotolo BT (2011) Mass spectrometry: come of age for structural and dynamical biology. Curr Opin Struct Biol 21:641–649. https://doi.org/10.1016/j.sbi.2011.08.002
46. Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217. https://doi.org/10.1006/jmbi.2000.4042
47. Webb B, Sali A (2016) Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics 54:5 6 1–5 6 37. https://doi.org/10.1002/cpbi.3
48. Bleiholder C, Wyttenbach T, Bowers MT (2011) A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method. Int J Mass Spectrom 308:1–10. https://doi.org/10.1016/j.ijms.2011.06.014
49. Faull PA et al (2009) Gas-phase metalloprotein complexes interrogated by ion mobility-mass spectrometry. Int J Mass Spectrom 283:140–148. https://doi.org/10.1016/j.ijms.2009.02.024
50. Devine PWA et al (2017) Investigating the structural compaction of biomolecules upon transition to the gas-phase using ESI-T-WIMS-MS. J Am Soc Mass Spectr 28:1855–1862. https://doi.org/10.1007/s13361-017-1689-9
51. Porrini M et al (2017) Compaction of duplex nucleic acids upon native electrospray mass spectrometry. ACS Cent Sci 3:454–461. https://doi.org/10.1021/acscentsci.7b00084
52. Scott D, Layfield R, Oldham NJ (2015) Structural insights into interactions between ubiquitin specific protease 5 and its polyubiquitin substrates by mass spectrometry and ion mobility spectrometry. Protein Sci 24:1257–1263. https://doi.org/10.1002/pro.2692

Chapter 13

Optimization of Sample Preparation for the Observation of Macromolecular Complexes by Electron (cryo-) Microscopy

Alexandre Frechard, Grigory Sharov, Maximilien Werderer, and Patrick Schultz

Abstract

Electron microscopy is a powerful tool for studying the homogeneity and structure of biomolecular complexes. The small wavelength of electron and the availability of electron optics enable the direct visualization of macromolecular assemblies in a large range of sizes between 5 and 100 nm. This informs us about the degree of multimerization or aggregation and provides precise information about their general shape and dimensions. When combined with sophisticated image analysis protocols, three-dimensional (3D) information can be gained from 2D projections of the sample, leading to a structural description. When intermediate steps of a reaction can be imaged, insights into the mode of action of macromolecules can be gained, and structure–function relations can be established. However, the way the sample is prepared for its observation within the vacuum of an electron microscope determines the information that can be retrieved from the experiment. We will review two commonly used specimen preparation protocols for subsequent single-particle electron microscopy observation, namely negative staining and vitrification.

Key words Single particle electron microscopy, Negative stain electron microscopy, Cryo-EM, EM sample preparation

1 Introduction

In the last decades, imaging of biological macromolecules by electron microscopy became a major structural biology method to describe the shape and size of macromolecules, to investigate their oligomeric or aggregated state, and, in combination with single particle image analysis, to describe their structure [1]. Ultimately this approach aims at understanding the mode of action of molecular assemblies by correlating their structure with their function. Handling macromolecular assemblies a few nanometers in size for their observation in the vacuum of a transmission electron microscope (TEM) while preserving the structural integrity of the

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_13,

sample is challenging. Over the years, many obstacles were overcome to reveal biomolecules in atomic details.

For most biological applications, the sample is applied onto 3 mm EM grids that can be made of different material. Continuously carbon-coated grids are used for negative staining where carbon acts as a support for the sample, whereas perforated carbon-coated grids can in addition be used in cryo-EM when adsorption has to be avoided. The carbon surface is generally hydrophobic and has to be rendered hydrophilic to allow proper spreading of the sample and adsorption of the biomolecules [2]. The properties of the carbon surface are altered by glow discharge, a process where the grids are exposed to an ionizing plasma formed by applying an electric current through a gas volume at low pressure. The ionized gas molecules will modify the properties of the carbon surface of the grids and render it hydrophilic.

We will describe two popular specimen preparation protocols for subsequent single-particle electron microscopy observation, namely negative staining and vitrification.

1.1 Negative Staining

The most straightforward negative staining approach was first described by Brenner and Horne in 1959 [3]. This method brings important preliminary information on sample quality and may be used on a daily basis to follow and optimize a purification protocol. Associated with single particle image analysis, it may provide important information on homogeneity or oligomerization states. This information may not be portable for cryo-EM sample preparation but guaranties a firm starting point. Negative staining consists of embedding the protein into a matrix of heavy atom salt to increase the contrast of unstained biomolecules and prevent, to some extent, their collapse during the drying process when exposed to the vacuum of the microscope. Stains of heavy metal salts, such as uranyl acetate, uranyl formate, phosphotungstic acid, and others, are commonly used, since they strongly scatter electrons and produce high amplitude contrast. It is important to notice that negative staining reveals only the surface and the overall shape of protein molecules and does not provide information about their inner structure. The resolution that can be achieved after image analysis of multiple images is limited to 15–20 Å, by the graininess of the stain and particle distortions during the drying process. This method is, however, very useful to check particle homogeneity and to determine initial 3D models. Furthermore, the sample preparation is easy to perform and to observe at room temperature in a standard transmission electron microscope. The stained sample, or more specifically, its heavy metal cast, is more resistant to radiation damage produced by the electron irradiation, and the high image contrast favors image interpretation.

1.2 Vitrification

The major drawback of negative staining is that the sample is dried during preparation and introduction into the microscope vacuum. To overcome this limitation, several options have been explored to keep the specimen hydrated including the construction of hydrated chambers around the specimen holder. Attempts were made to replace water with sugar polymers to functionally replace the shielding properties of water, and remarkable results were obtained for 2D crystals of membrane proteins. Cryogenic methods were tried early on and proved efficient to reduce radiation damage resulting from the interaction of electrons with matter by a factor of 2 at liquid nitrogen temperature. However, freezing water invariably resulted in ice crystal formation which affects sample distribution and image quality. Being able to obtain vitreous ice at atmospheric pressure was a major breakthrough for studying biological macromolecules in a fully hydrated state in the electron microscope. The method consists of flash freezing a thin layer of an aqueous suspension of the purified biomolecule to form a vitrified sample that can be transferred and observed in a cryo electron microscope at liquid nitrogen temperature. A layer of suspension thinner than 100 nm is generally obtained by removing the excess of a drop of sample applied onto the electron microscopy grid by blotting with a filter paper. Alternatively, sample micro droplets can be sprayed onto the grid or applied by a nanoliter dispensers [4]. Vitrification is generally performed by plunging the EM grid into liquid ethane cooled by liquid nitrogen for better heat dissipation. High pressure freezing devices have been developed for the vitrification of up to 300 nm thick sample [5]. This approach, which was rewarded the Nobel Prize in chemistry in 2017, preserves the hydrated structure of the molecules at atomic resolution. The vitrification process can be performed once the sample is adsorbed on a carbon film, but adsorption can be avoided by using holey carbon films. In this case, observations are made on the thin layer of vitrified suspension stretched over the carbon-free holes. However, the sample still experiences the interaction with the air–water interface which can be deleterious when the specimen is prone to denaturation or can result in preferential orientations of the molecules [6].

2 Materials

We list here the equipment needed to prepare purified biological molecules for their observation in electron microscopy. The negative staining and vitrification methods require common material listed in Subheading 2.1 and have their own specificities detailed in Subheadings 2.2 and 2.3, respectively.

2.1 Common Materials and Equipment

1. The EM grid, a universal, and versatile specimen support. The universal support to manipulate EM samples is a thin metallic grid 3 mm in diameter adapted for most electron microscope holders. EM grids are made of copper, copper/rhodium, or gold to be conductive and eliminate electron charge accumulation. Square holes of different sizes are separated by grid bars that absorb electrons. Copper mesh grids with 300 grid squares per inch are commonly used because they are conductive, stable under the electron beam, and inexpensive. The grid is generally covered by a 5–20 nm thin carbon film onto which protein complexes are adsorbed for subsequent negative staining. To avoid sample adsorption in cryo-EM, the grids can also be covered with a self-made carbon film perforated with holes of various diameter ranging between 0.3 and 10 μm in diameter [7]. More recently, perforated grids with calibrated circular holes arranged into regular arrays were developed to serve automated data acquisition [8]. Such perforated grids can be covered by a continuous thin carbon film when adsorption cannot be avoided. Holey carbon films can also be covered by a single monolayer thick graphene or graphene oxide crystalline layer instead of amorphous carbon, thus leading to a less grainy background [9] (Oxford Instruments, quantifoil micro tools GMBH).
2. High precision tweezers. To properly manipulate the EM grids without distortion and contamination, high precision and sharp tweezers are recommended (Dumont style 5 tweezers, Fine Science Tools GMBH ref.: 11252-30). Make sure to grab the grids by the rim to not damage or distort the support. Standard straight tweezers are sufficient to manipulate grids, but other types of tweezers can be useful in electron microscopy: anti-capillarity tweezers have a wider angle at the tip when closed, limiting the amount of liquid retained by capillarity. Inverted tweezers are also found useful since they remain closed until squeezed, limiting the risks of dropping grids when handling the tweezers.
3. Glow discharge and plasma cleaner devices. These table top instruments are used to produce a gas plasma to charge the surface of the grids by an electrostatic potential in order to produce a hydrophilic surface suitable for protein absorption. The devices differ by the quality of the vacuum (between 10^{-2} and 10^{-6} Torr, or 1.3 and 1.3×10^{-4} Pa), the way the plasma is produced, and the mixture of gases that can be used to form the plasma (ELMO, Cordouan Technologies; Nanoclean model 1070, Fischione).
4. Filter paper. Different steps of the protocol require blotting the grids to remove excess of staining solution or water (Macherey-Nagel, ref.: 202009).

5. Cross-linkers. Chemical cross-linkers such as glutaraldehyde (Sigma, ref.: G5882) or bis-sulfosuccinimidyl suberate (BS3, Thermo Fisher Scientific, ref.: 21580) are often used to stabilize fragile sample by forming chemical bonds between proximal lysine residues. Chemical cross-linkers have a short life time and may form oligomeric states. They can be aliquoted and frozen to extend their life time.

2.2 Negative Staining

1. Heavy atom salt solution. Uranyl acetate (Agar Scientific, ref.: AGR1260A; Polysciences ref.: 21447-25) uranyl formate, phosphotungstic acid (PTA), and other heavy metals are commonly used to embed the biological sample into a shell of electron-dense material. The heavy atom salt is dissolved in water at 2% (w/v). Most of the solutions are at low pH, and only few of them, such as PTA, can be buffered to pH 7.0 without forming aggregates. In this respect, phosphate buffer present in the sample solution may provoke uranyl acetate precipitation observed as dark clusters under the microscope.
2. Handling heavy metal salts requires strict safety protections: use gloves, mask and safety glasses and work under hood when preparing solutions. Uranyl salts are slightly radioactive, and specific safety precautions are necessary: store uranyl salts in a lead protection box, identify the workbench used to work with uranyl acetate, weigh powders in a secured work station, work under a hood protected by a Plexiglas screen when handling concentrated solutions, and finally, make sure to dispose any contaminated waste according to your lab policy. Radiations emitted from 2% solutions are shielded by water molecules and the glass bottle. Depending on national policies, uranyl-based reagent might be difficult or even illegal to buy. Please make sure your experiments comply with your local regulations.
3. Note that there are now non-radioactive alternatives to uranyl acetate: based on methylamine vanadate (NanoVan, Nanoprobes), methylamine tungstate (Nano-W, Nanoprobes), UranyLess (TAAB), and UA-Zero (Agar Scientific).
4. Parafilm. The protocol includes several washing steps aiming at solution replacement. Parafilm may be used to deposit small solution drops on top of which the EM grid can be placed (VWR international SAS, ref.: 97949).

2.3 Cryo-EM

Preparing the sample for cryo-EM requires more material and equipment than negative staining. Although automated spray systems and microfluidic-based devices are emerging, vitrification is usually performed using a cryo-plunger. Make sure that all the materials are at your disposal before you start since the protocol cannot be interrupted.

1. Liquid nitrogen. Liquid nitrogen is used at several steps in the protocol to refrigerate and liquefy ethane, to store the vitrified grid and cryo-box, and to transfer the cryo-box to a dedicated short-term or long-term storage device. Once vitrified, the specimen has to be kept at liquid nitrogen temperature (boiling temperature of −196 °C under normal atmospheric pressure) to avoid water devitrification.
2. Plastic beaker. A 1 L polypropylene beaker with a handle to pour nitrogen (Vitlab, ref.: 442941).
3. Cryo-boxes. These round boxes are used to store the vitrified samples in liquid nitrogen. They generally hold four EM grids and have a transparent cover hold in place by a screw. The cover has a single slit and can be rotated to select one of the four EM grid slots (Oxford Instruments, ref.: G3539).
4. Screw driver. A screw driver is necessary to fasten and unfasten the cryo-box cover immersed in liquid nitrogen.
5. Long isolated tweezers. Twenty-centimeter long tweezers are used to manipulate the cryo-boxes submerged in nitrogen. To protect the operator from frostbites the tweezers should be cold-isolated by a protective layer.
6. Liquid nitrogen tank. A 5 L liquid nitrogen container is required to store over longer periods the cryo-boxes which contain the grids. The cryo-boxes can be put into Falcon tubes perforated for liquid nitrogen influx and fixed to a 2–3 mm thick plastic-coated metallic cord which hangs out of the tank once the tube is immersed. More elaborate systems have been developed to organize multi-user grid storage, efficient cryo-box retrieval, and automated nitrogen refill (EMS Cryo Pucks G2).
7. Cryo-plunger. Specimen vitrification is commonly achieved by forming a thin layer of the sample on the EM grid through blotting away the excess of sample and subsequent plunging of the grid into a liquid ethane slush refrigerated by liquid nitrogen [5]. Simple, versatile, and portable devices were designed in which a tweezer holding the grid was fixed on a weighted friction-reduced rod that moved by gravity to plunge the grid into the cryogen container. Ethane is liquefied in a 2 mL aluminum container inserted in a styrofoam container of liquid nitrogen. Since the boiling temperature of liquid nitrogen is lower than the melting temperature of ethane (−183.3 °C), the ethane will tend to solidify. An ethane heating device or a properly designed ethane tank assures that ethane does not reach liquid nitrogen temperature and stays as a liquid slush. Highly automated devices are currently available, providing a temperature and humidity controlled chamber to avoid

evaporation, a force- and time-controlled blotting and automated plunging [10] (Vitrobot Mark IV, Thermo Fisher; EM GP2, Leica).

8. Ethane gas cylinder. A 5 L cylinder containing compressed ethane gas is required to fill the ethane container. The cylinder has to be equipped with a pressure reducer to fine tune the gas flow and with a thin flexible plastic tube (diameter of 2 mm) to flow the ethane stream into the liquid nitrogen-cooled cryogen container. The extremity of a pipette tip can be mounted at the end of the plastic tube for increased precision in driving the ethane flow.
9. Safety issues. Handling liquid nitrogen and ethane which are very cold and vaporize into large amounts of gas requires strict safety protections. Liquid ethane close to its solidification temperature is dangerous and will freezes any hydrated object upon contact. Care must be taken to avoid spills on skin or eyes. The use of a face shield is highly recommended. Thick gloves must be used to manipulate cold objects.

3 Methods

3.1 Negative Staining

Negative staining is a fast method to assess the quality of your sample by checking its stability, homogeneity, and propensity to aggregate. Key steps of sample preparation are illustrated in Fig. 1 and detailed below.

3.1.1 Preparation of a Negatively Stained Sample

1. Glow discharge a carbon-coated EM grid to render the carbon support hydrophilic. Place the EM grid in the glow discharging device, carbon side facing up. The nature of the gas present in the chamber, the partial gas pressure, the voltage applied, the glow discharge time, and the geometry of the device will all affect the ionization of the carbon support. Using air as the residual gas at a pressure of 1.8×10^{-1} bar, a current of 2 mA should be applied for 45 s to assure a suitable hydrophilicity (Fig. 1a).
2. Deposit 3–5 μL of your sample onto the glow discharged side of the grid held with tweezers and wait for 1 min for protein adsorption. The protein concentration should be around 30 μg/mL (Fig. 1b).
3. Place a square of parafilm onto the bench and deposit four separate 40 μL drops, three drops of water or buffer, and one drop of phosphotungstic acid at 2%.
4. After adsorption (**step 2**), place the grid sequentially on top of the first three droplets of water to remove the material not adsorbed to the carbon (Fig. 1c). Then remove the excess of

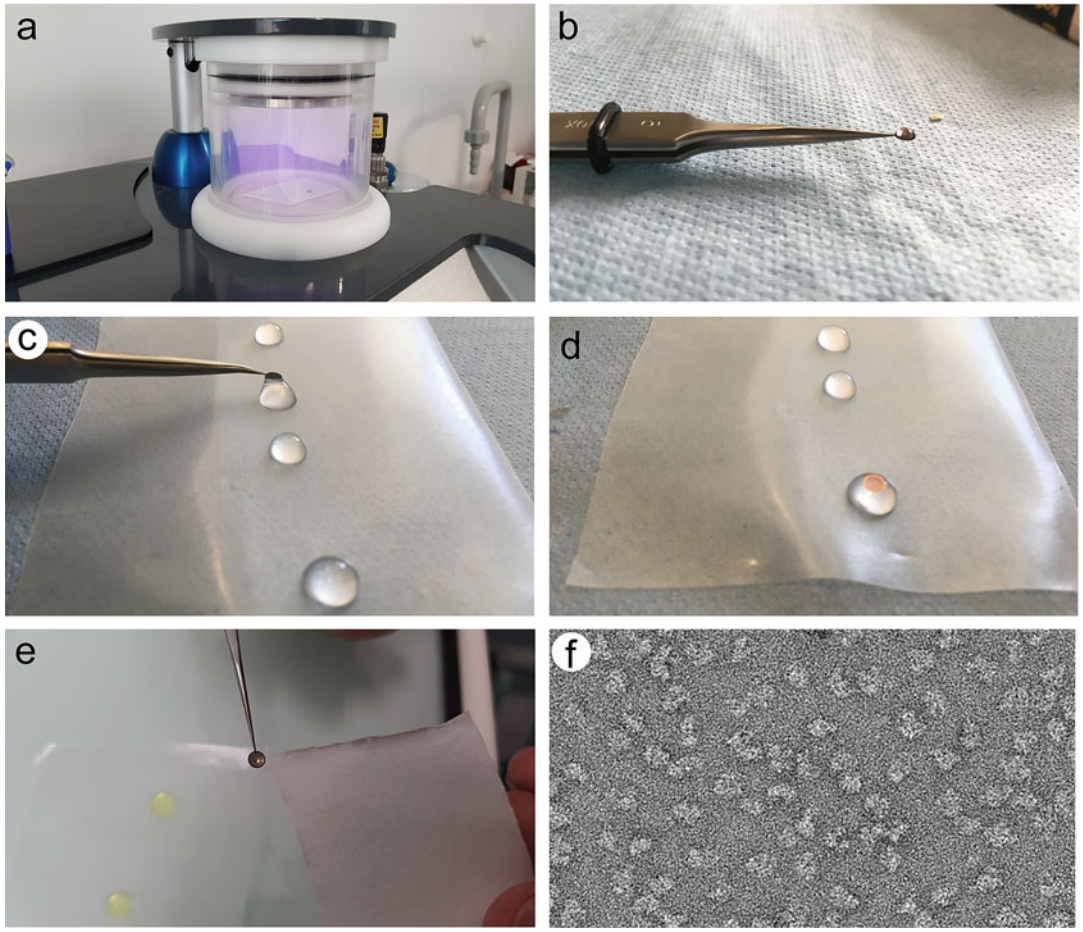


Fig. 1 Negative staining. (**a**) The EM grid is glow-discharged in a plasma of ionized air to render the support hydrophilic. (**b**) 3 μL of the purified biomolecules solubilized in buffer is allowed to adsorb on the EM grid support for 1 min. (**c**) The excess of sample and detrimental buffer components are diluted away by several passes on top of buffer drops. (**d**) Sample staining: the grid is deposited on the top of a drop of stain—here a 2% of uranyl acetate—for 30 s. (**e**) The staining solution is blotted away using filter paper. (**f**) Electron micrograph of a well-stained sample, particles are well dispersed, show little aggregation and optimal embedding in the stain layer

water using filter paper and transfer the grid to the last drop containing the phosphotungstic acid and wait for 30 s (Fig. 1d).

5. Grab the grid with tweezers, approach the filter paper from the side of the grid, gently remove the excess of stain by capillarity, let the grid dry completely for 1 min, and store it in a grid storage box before observation in the transmission electron microscope (Fig. 1e).

3.1.2 Observation and Trouble Shooting

Ideally the sample should be evenly distributed with little interactions between individual particles (compare Figs. 1f and 3a). The macromolecules should be included in a layer of stain of the same thickness as the particle to sustain optimally the sample, to reveal all parts and to obtain the best contrast. It is often necessary to screen several places on the grids to find optimal staining. The following indications may help to improve staining.

1. Poor staining. In a standard bright field transmission electron microscope, your protein sample should appear as a lighter area embedded in a darker background. If your sample is poorly stained, you will only see fingerprints of your protein (Fig. 3b) or worse, and your sample may be positively stained and appear as dark spots on a lighter background. You may adjust the glow discharge conditions to render the support more hydrophilic and make sure that the staining solution is not removed too fast by wetting the filter paper beforehand. Adding 0.5% glycerol may help to reduce the drying speed and improve staining.
2. Excess staining. Your sample may be well embedded, but a thick layer of stain may be superimposed to the particles, thus reducing the contrast and obscuring fine details of your sample (Fig. 3c). Sometimes wiping away the stain faster can help. Check that your sample buffer does not contain large excess of sucrose of glycerol which increases its viscosity. In this case, an additional washing step can be added.
3. Presence of stain crystals. Uranyl salts precipitate at neutral pH or in contact with phosphate buffer and form small crystals visible in EM. Make sure that these reagents are absent or sufficiently dilute in the last staining step and that the pH is suitable.
4. Damaged/destroyed proteins. Some fragile protein complexes may be altered during grid preparation and even fall apart during the dilution, adsorption, or staining steps. Chemical cross-linkers such as glutaraldehyde or BS3 may stabilize the sample. As a standard procedure, glutaraldehyde is added at a final concentration of 0.1–0.5% for 2 min, before adsorption on the EM grid. Cross-linkers may lead to protein aggregation, and therefore, cross-linking time and concentration need to be adjusted. Adding 1% glycerol in the staining solution may help sustaining fragile samples and prevent their collapse during the drying step.
5. Small number of particles. When a smaller number of particles than expected is observed, the sample may form aggregates. Aggregates are not always detected because of their small interaction interface with the carbon support which can lead their dissociation during the washing or staining steps. Alternatively,

the sample suspension may contain reagents (elution peptides, detergents, lipids) that compete with the protein for adsorption. In this case, these reagents have to be removed whenever possible.

3.2 Vitrification

3.2.1 Preparation of the Vitrified Suspension

The protocol presented here is adapted for a Vitrobot, but most of the steps can be adjusted to a handmade plunger or a device from another manufacturer. Key steps are illustrated in Fig. 2 and detailed below.

1. Prepare four EM grids coated with holey carbon (non-adsorbed sample) or plain carbon (adsorbed sample) and glow discharge them as described in Subheading 3.1.1.
2. Prepare the plunging device by adjusting the temperature and humidity of the chamber.

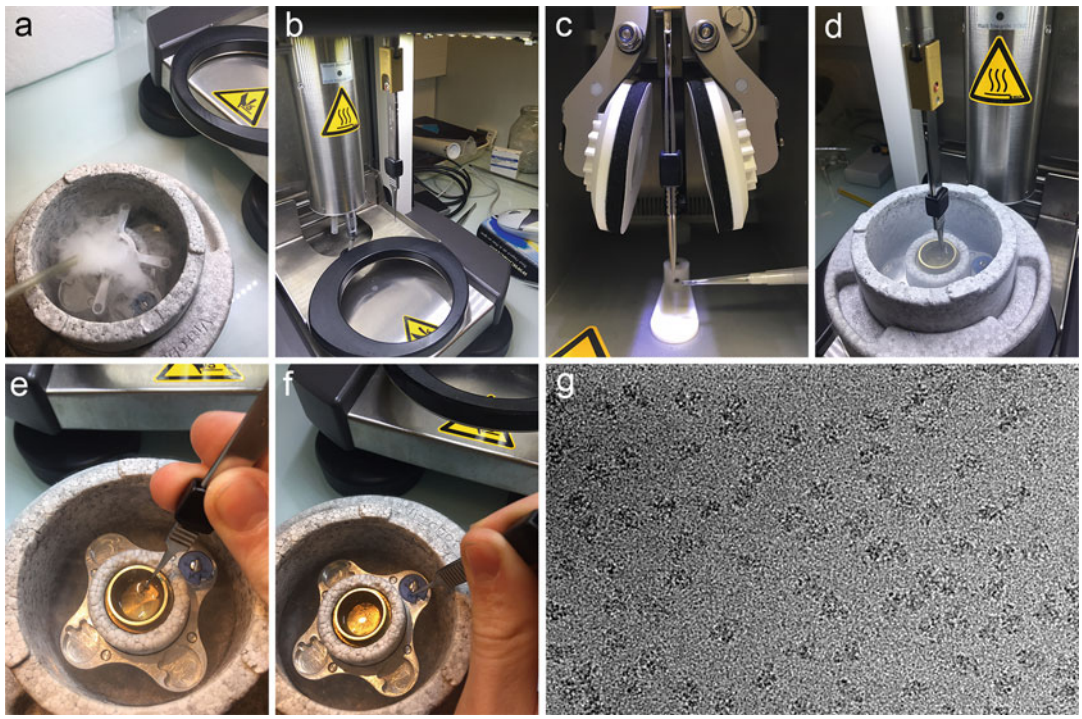


Fig. 2 Sample vitrification. (**a**) Styrofoam holder filled with liquid nitrogen, showing the central ethane container. The ethane gas was liquefied, holding the tube appearing on the left into the cold container. (**b**) The grid is loaded on the tweezers attached to the plunger. (**c**) 3 μL of sample is deposited on the grid which will be lifted up to the filter paper pads that will be automatically pressed against the grid to blot away the excess of liquid. (**d**) The plunger quickly submerges the grid in the liquid ethane. (**e**, **f**) The vitrified sample is transferred into a cryo-box for storage. (**g**) Electron micrograph of a frozen hydrated sample showing well-dispersed particles, little aggregation, thin ice, and no ice contamination

3. Cool down the liquid nitrogen and ethane containers with liquid nitrogen. Once it has reached liquid nitrogen temperature remove liquid nitrogen from the ethane container and place it in its holder.
4. Liquefy ethane in the ethane container by introducing a flow of ethane gas with a polyethylene tubing connected to the ethane cylinder. The gas flow can be manually adjusted by clamping the tubing. Place the tube at the bottom of the container with a weak gas flow until some liquid ethane forms. At that point, the gas flow may be increased until the liquid level reaches a few millimeters from the top of the container. Wait for the ethane to solidify at the rim of the container, which signals that the ethane is cold enough.
5. Adjust the liquid nitrogen level and place the container under the plunger.
6. Fix the grid onto the plunger tweezer and transfer it into the chamber.
7. Apply 3 μL of the sample onto the grid. In the absence of adsorption, the sample concentration needs to be higher, in the order of 0.3–1 mg/mL.
8. Blot the grid using the filter paper supplied by the manufacturer after having adjusted the blotting time and blotting force.
9. Plunge the grid down into liquid ethane.
10. Transfer quickly the grid into a cryo-box and store it in liquid nitrogen.

3.2.2 Observation and Trouble Shooting

1. Opaque grid. If your grid is not transparent to electrons, it generally means that the ice is too thick. Blot time and force can be increased. Make sure that reagents that increase viscosity such as glycerol or sucrose are absent. The thickness of the vitrified layer is of key importance to reach high resolution in single particle cryo-EM. The thinnest is the better, and small amounts of nonionic detergents may be added to form even thinner layers.
2. Poor contrast. Thick ice will reduce image contrast. Some reagents such as glucose or glycerol may affect the contrast. Make sure that their concentration is lower than 1%.
3. Ethane contaminants. Long filaments forming aggregates are sometimes present and originate from impurities present in the ethane which do not evaporate in the temperature and vacuum conditions of the microscope (Fig. 3d). Avoid solid ethane accumulation on the EM grid which forms when transferring the grid from the ethane to the liquid nitrogen container after vitrification. Slow retraction of the grid from the container is the best way to prevent ethane accumulation. Solid ethane may

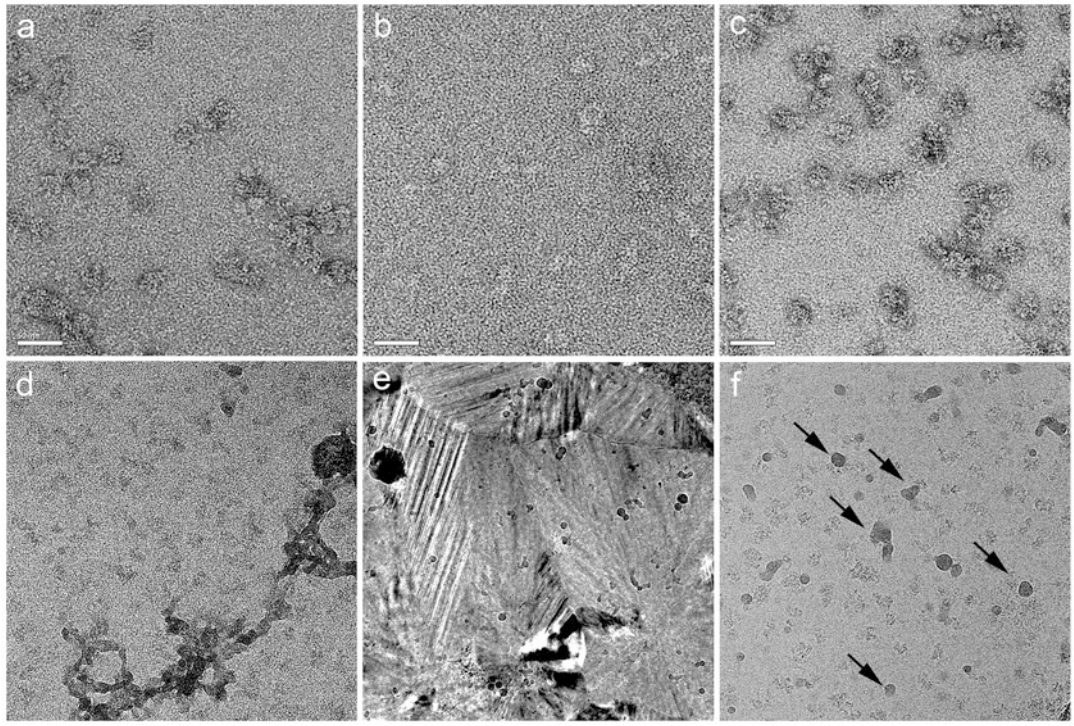


Fig. 3 Troubleshooting. Negative staining: (**a**) Electron micrograph showing aggregated particles. (**b**) Electron micrograph showing a heavily stained sample in which the details of the particles are lost, and the overall contrast is weak due to stain accumulation on top of the particle. (**c**) In this field, the particles are not well-stained and embedded, and only a small rim of stain surrounds each particle whose structure is poorly sustained. *Sample vitrification*: (**d**) Contaminant filaments arising from residues remaining after ethane evaporation. (**e**) Transformed ice image due to warming up of the sample during transfer in the microscope. (**f**) Partially molten hexagonal ice crystals (arrows) are spread over the grid surface and may have been collected during the grid transfer or mounting steps and exposition to humid environment

be gently removed under liquid nitrogen with tweezers. Storage for a few days in liquid nitrogen may detach the solid ethane from the grid.

4. Crystalline ice. During freezing, parts of the water layer may not have cooled down fast enough, thus forming hexagonal ice. This should be restricted to the thicker parts of the suspension layer. If this problem persists, the plunging speed or depth may need to be adjusted. Alternatively if the grid heats up during its manipulation, the vitreous ice will transform into cubic ice and deteriorate image quality [5] (Fig. 3e). To avoid this problem, it is imperative to keep the grid cold during the transfer and mounting steps.
5. Ice contamination. During the grid transfer and mounting steps, ambient humidity can condense on the grid and form ice crystals on its surface (Fig. 3f). Such crystals may also be

collected from the liquid nitrogen. Furthermore, ice crystals loosely attached to the grid may evaporate and re-condense onto the cold grid surface, thus forming a continuous or speckled ice layer sometimes difficult to detect. This continuous contamination accumulates over time due to water molecules present in the vacuum of the microscope. Working in a dry environment, changing liquid nitrogen frequently to avoid ice crystal accumulation, handling the grids in liquid nitrogen or very close to the surface, and heating/emptying the containers between two freezing sessions will reduce ice contaminations. If contamination builds up rapidly in the microscope, check the anticontamination devices of your microscope.

6. Damaged sample. During cryo-EM sample preparation, protein complexes may be damaged when placed in a thin layer of suspension due to its interactions with the air–water interface. This may result to partial denaturation, dissociation of subunits, and aggregation. Sample cross-linking with glutaraldehyde or BS3 as described in Subheading 3.1.2 generally helps stabilizing the molecular assemblies. Alternatively, the air–water interface may be screened with amphiphilic molecules such as detergents. In this case, the particle concentration may need to be increased since the air–water interface does not concentrate the biomolecules any more. The possibility to adsorb the proteins on a continuous carbon film is worth exploring to avoid the air–water interface.
7. Lack of particles. It is difficult to predict the protein concentration needed to produce projection image with densely packed particles. Recent experiments showed that in most cases as much as 90% of the particles accumulate at the air–water interface [6]. The number of particles adsorbed on the grid depends on their propensity to accumulate at the interface. In the absence of surface accumulation, an extremely rare situation, particle concentrations of 100 mg/mL (10% of the volume) may be needed to produce densely populated images. Absence of particles may therefore also arise from competition for the surface and high concentrations of detergents or of hydrophobic peptides that may shield the surface and prevent surface accumulation of the sample. In extreme cases, most of the particles are denatured and form a grainy background. A cross-linking experiment generally helps sorting out this problem.
8. The quality of the frozen hydrated grid depends critically on sample homogeneity and stability. It is essential to give the highest priority to protein purification to increase the chances to produce an exploitable grid.

Acknowledgments

We acknowledge the support from the Institut National de la Santé et de la Recherche Médicale (INSERM), the Centre National pour la Recherche Scientifique (CNRS), the Ligue contre le Cancer, and the grant ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02. We acknowledge the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC.

References

1. Cheng Y (2018) Single-particle cryo-EM-how did it get here and where will it go. Science 361 (6405):876–880
2. Dubochet J et al (1971) A new preparation method for dark-field electron microscopy of biomacromolecules. J Ultrastruct Res 35 (1):147–167
3. Brenner S, Horne RW (1959) A negative staining method for high resolution electron microscopy of viruses. Biochim Biophys Acta 34:103–110
4. Jain T et al (2012) Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. J Struct Biol 179(1):68–75
5. Dubochet J et al (1988) Cryo-electron microscopy of vitrified specimens. Q Rev Biophys 21 (2):129–228
6. Noble AJ et al (2018) Routine single particle CryoEM sample and grid characterization by tomography. Elife 7:e34257
7. Fukami A, Adachi K (1965) A new method of preparation of a self-perforated micro plastic grid and its application. J Electron Microsc 14 (2):112–118
8. Russo CJ, Passmore LA (2016) Progress towards an optimal specimen support for electron cryomicroscopy. Curr Opin Struct Biol 37:81–89
9. Palovcak E et al (2018) A simple and robust procedure for preparing graphene-oxide cryo-EM grids. J Struct Biol 204(1):80–84
10. Vos MR et al (2008) The development of a glove-box/Vitrobot combination: air-water interface events visualized by cryo-TEM. Ultramicroscopy 108(11):1478–1483

Chapter 14

Solubilization and Stabilization of Native Membrane Proteins for Drug Discovery

Vincent Corvest and Anass Jawhari

Abstract

Membrane proteins (MPs) are stable in their native lipid environment. To enable structural and functional investigations, MPs need to be extracted from the membrane. This is a critical step that represents the main obstacle for MP biochemistry and structural biology. Here we describe detergent solubilization screening of MPs using dot-blot and Western-blot analyses. Good solubilization conditions are ranked for their best capacity to stabilize MPs using thermal shift assay. The protein functionality is evaluated by radioligand binding (for G-protein-coupled receptor) and ATPase activity (ABC Transporter) and finally the aggregation status as well as protein homogeneity are assessed by Native-polyacrylamide gel, chemical cross-linking, and size exclusion chromatography.

Key words Membrane proteins, Solubilization, Stabilization, Detergent, Native-PAGE, Size exclusion chromatography, Chemical-crosslinking, Function

1 Introduction

Membrane proteins (MP) represent ~30% of human proteins [1]. They are crucial for cellular physiology as they are directly involved in cell adhesion, cell–cell communication, signal transduction, and transport [1]. This is why large majority (~70%) of therapeutic targets are membrane proteins (MPs) [2]. To better understand MP function and to enable drug development using structure-based drug design or fragment-based drug discovery, MPs need to be extracted from their membrane environment [3]. MPs extraction and solubilization represent the most critical step of MPs isolation [4], since depending on the used surfactant/detergent, they can go through significant structural constraints which may significantly compromise their structural and functional integrities. To date, no detergent has been reported which tackles membrane protein solubilization and stabilization universally. Therefore, it is essential to screen for the most suitable detergent [5]. Here we show that solubilization efficiency of different

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_14,

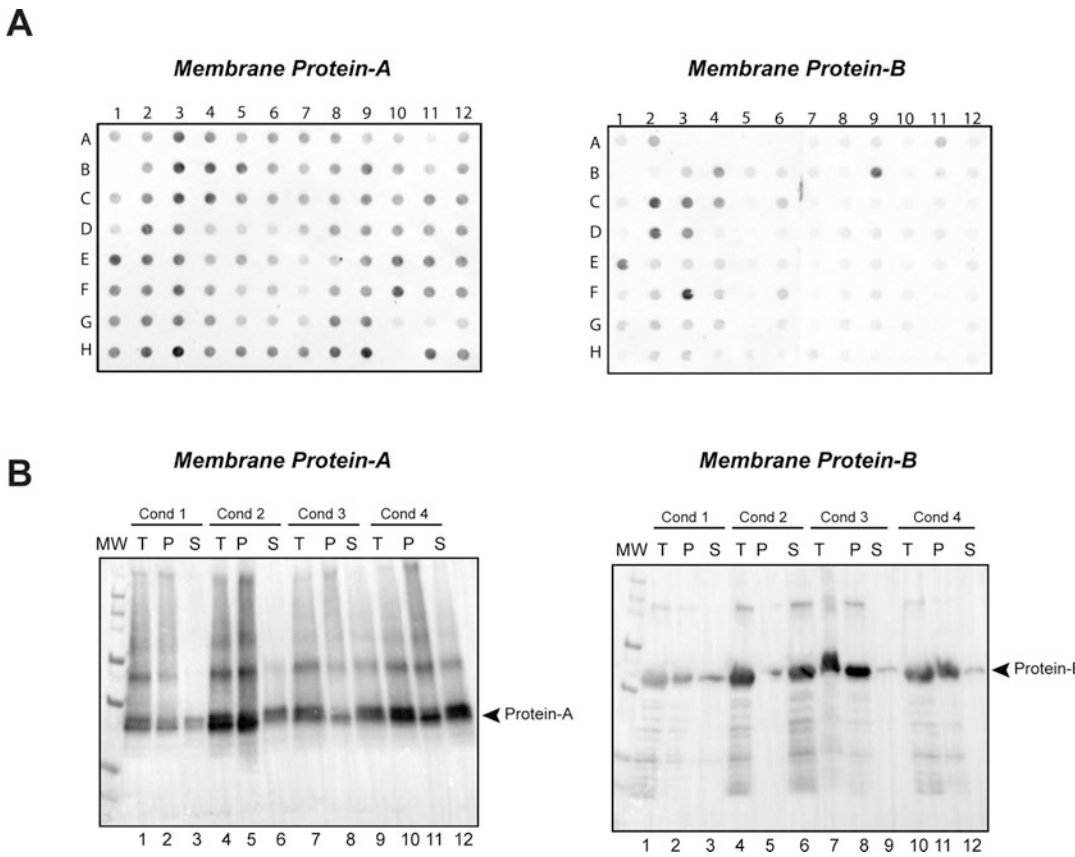


Fig. 1 Screening of membrane proteins solubilization. (**a**) Large solubilization screening performed in 96-well plate and monitored by dot-blot. (**b**) Confirmation of solubilization efficiency by Western blotting. Membrane protein A and B correspond to typical easy and difficult to solubilize targets, respectively (Adapted from [7])

detergents can be monitored in a 96-well plate dot blot format for easy and difficult to solubilize MPs (Fig. 1a). To avoid false positives, Western blot experiments are performed to confirm the presence of the solubilized proteins of interest at the expected molecular weight (Fig. 1b). This approach was previously reported [6, 7]. In the following example, the best stabilizing detergent is evaluated on its ability to solubilize and stabilize MPs in a Western blot–based thermal-shift assay (Fig. 2a). This assay relies on the assumption that unstable heated proteins will aggregate, and after ultracentrifugation and Western blot, the band intensity corresponding to the protein will decay proportionally to its instability [8]. Few examples using this approach were described [8–10]. To make sure that solubilization and purification steps were not harmful to the protein, functional assays are performed. Figure 2b shows two examples, a radioligand binding assay for a G-protein-coupled receptor ($A_{2A}R$) and an ATPase activity assay for an ABC transporter (BmrA). The structural integrity and more

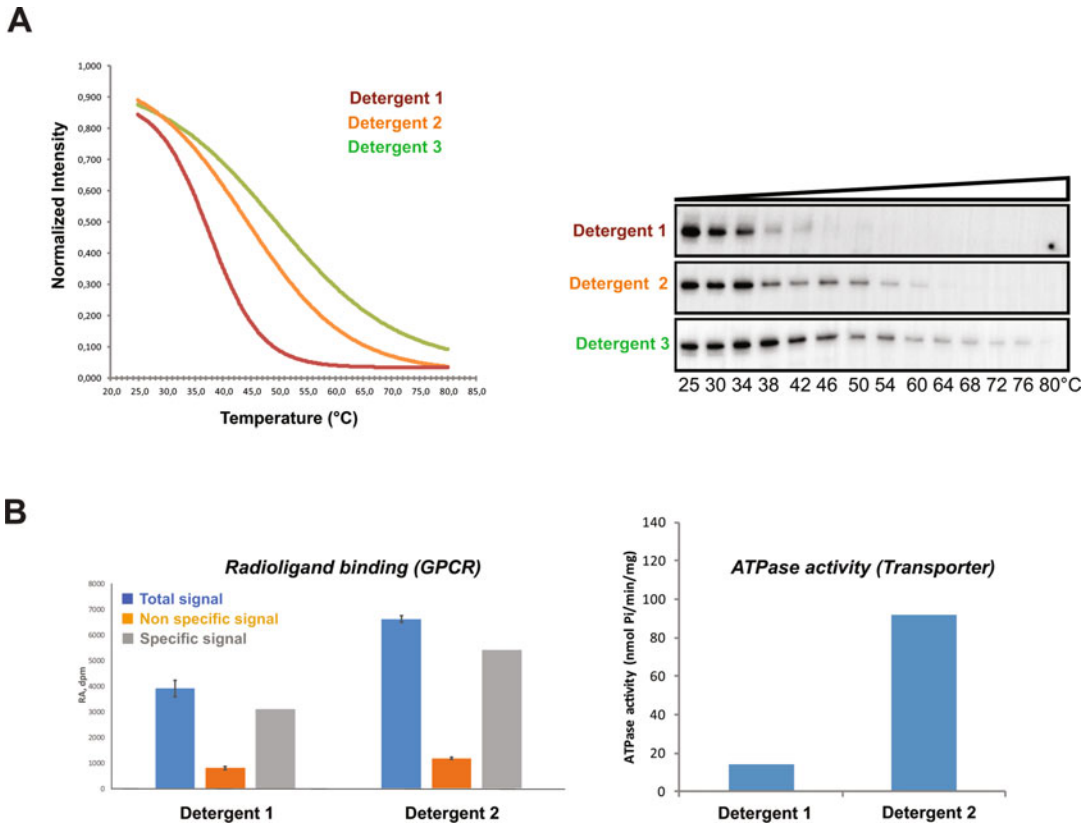


Fig. 2 Membrane protein stability assessment. (**a**) Western blot based thermostability assay (adapted from [12]). (**b**) Functional investigation using radioligand binding and ATPase activity assays for a GPCR and a Transporter target, respectively. Two examples were previously described [9, 11]

particularly the aggregation status and homogeneity of the MPs are evaluated by different methods. Native-PAGE helps to ensure that the protein can migrate inside of the gel, suggesting a non-aggregation state (Fig. 3a). This is further confirmed by size exclusion chromatography (SEC, Fig. 3c) which do not show any protein signal at the void volume. The protein homogeneity and oligomeric state can also be investigated by SEC, Native-PAGE, and chemical cross-linking (Fig. 3b).

2 Materials

All solutions are prepared using ultrapure water (18.2 MΩ-cm conductivity at 25 °C) and analytical grade reagents. All chemicals are obtained from Sigma-Aldrich (unless indicated otherwise).

2.1 Solubilization Screening

1. Bio-Dot® microfiltration apparatus (Bio-Rad).
2. Phosphate-buffered saline (1× PBS): 25 mM Na_2HPO_4, 150 mM NaCl, pH 8.0.

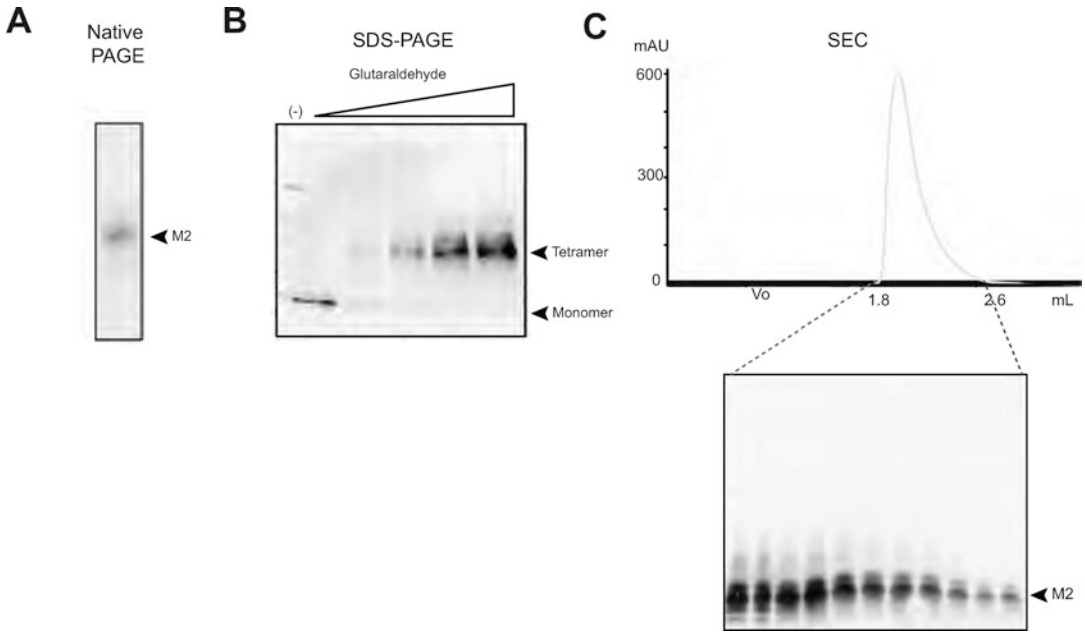


Fig. 3 Membrane protein behavior in solution. (**a**) Native-PAGE on purified Matrix 2 ion channel revealed by western blot. (**b**) Chemical cross-linking assay using different amounts of glutaraldehyde, analyzed by SDS-PAGE and revealed by Western blot. (**c**) Gel filtration (5-150GL) profile and corresponding Western blot analysis. (Adapted from [13]). A similar approach was also previously reported for the characterization of the human transcription factor TFIIE [14]

3. Protease inhibitor cocktail (1× PIC): 1 tablet/100 mL buffer (SigmaFAST EDTA free).
4. Blocking buffer: 1× Roti®-block buffer (Carl Roth).

2.2 Radiobinding Assay

1. MicroBeta filtermat 96-cell harvester (Perkin-Elmer).
2. MicroBeta^2 microplate counter (Perkin-Elmer).
3. PEI-coated GF/B glass fiber filter-bottom 96-well microplate with BackSeal and TopSeal (Perkin-Elmer).
4. Ultima Gold MV liquid scintillation cocktail (Perkin-Elmer).
5. Wash buffer: 50 mM Tris–HCl, pH 7.4.
6. Radioligand solution: 3.6 μM ^{3}H-CGS21680 (Perkin-Elmer) in wash buffer.
7. Nonspecific ligand solution: 3.6 μM ^{3}H-CGS21680, 1 mM CGS21680 in wash buffer.
8. γ-Globulin solution: 0.1% (w/v) γ-globulin in wash buffer.
9. PEG6000 solution: 25% (w/v) PEG6000 in wash buffer.

2.3 ATPase Activity Assay

1. Inorganic phosphate (Pi) standard stock solution: 10 mM Na_2HPO_4, pH 8.0.
2. Adenosine 5′-triphosphate (ATP).
3. Sodium orthovanadate (Na_3VO_4) heated at 95 °C.
4. Tris-buffered saline (1× TBS): 50 mM Tris–HCl, 50 mM NaCl, pH 8.0.
5. Reaction buffer: 50 mM Tris–HCl, 50 mM NaCl, 10 mM $MgCl_2$, 6 mM NaN_3, pH 8.0.
6. Revelation buffer (to be prepared just before use): 45 mM ascorbic acid, 17.5 mM ammonium molybdate, 15 mM zinc acetate, pH 5.0.

2.4 Clear-Native-PAGE

1. 5× native loading buffer: 250 mM Tris–HCl, 50% glycerol, 0.25% bromophenol blue, pH 8.0.
2. CN-PAGE anode buffer: 25 mM imidazole, pH 8.0.
3. CN-PAGE cathode buffer: 50 mM tricine, 0.05% Na-deoxycholate, 0.01% *n*-dodecyl β-D-maltoside (DDM, Carl Roth), 7.5 mM imidazole, pH 8.0.

2.5 Glutaraldehyde Cross-Linking Assay

1. Glutaraldehyde (Carl Roth).
2. 5× SDS-DTT loading buffer: 250 mM Tris–HCl, 10% sodium dodecyl sulfate (SDS), 0.5 M dithiothreitol (DTT), 50% glycerol, 0.25% bromophenol blue, pH 8.0.

3 Methods

3.1 Screening of Membrane Proteins Solubilization

Here we describe a 96-well microplate solubilization screening for membrane proteins using the micro-ultracentrifugation to separate soluble from insoluble fractions. For limited micro-ultracentrifugation capacity, an alternative using the biotinylated membranes solubilization and separation method is described in [7].

3.1.1 Cell Lysis and Membrane Fractionation

1. Resuspend dry cell pellet at 2 mL/g with 1× PBS supplemented with 1× PIC.
2. Lyse cells on ice using a Bead Beater homogenizer with 0.1 mm diameter glass beads (0.5 mm for yeast) with five cycles of 30 s lysis followed by 2 min cool down (*see* **Note 1**).
3. Centrifuge at increasing speeds at 4 °C to collect different membrane fractions: 1000 × *g* for 5 min (whole cell and cell debris), 15,000 × *g* for 30 min (internal membranes, ER, mitochondria, etc.) and 100,000 × *g* for 45 min (plasma membrane).

4. Resuspend membrane fractions at 0.5 mL/g of initial dry cell pellet with 1× PBS, 1× PIC, 10% (v/v) glycerol.
5. Quantify total membrane proteins of each fraction using the micro BCA protein assay kit (Thermo Scientific) and determine the protein of interest localization by SDS-PAGE and Western blot.
6. Aliquot, flash-freeze in liquid nitrogen and store membrane fractions at −80 °C until use.

3.1.2 96-Well Microplate Solubilization

1. Prepare 1 mL of 2× stock solutions in 1× PBS, 1× PIC of each detergent or combination of detergents to be tested in a 96-deep well plate (sufficient for 20 tests of 96-well microplate solubilization) (*see* **Note 2**).
2. Dilute the membrane fraction containing the protein of interest at 2 mg/mL in 1× PBS, 1× PIC.
3. Aliquot 50 µL diluted membranes into a 96-well microplate (*see* **Note 3**).
4. Using a multi-channel pipet transfer 50 µL of detergent solution from the 96-deep well 2× stock solutions plate to the 96-well solubilization microplate containing the membranes.
5. Incubate the 96-well solubilization microplate at 4 °C for 2 h under agitation.
6. Transfer the 100 µL solubilisates into 0.5 mL micro-ultracentrifuge tubes and ultracentrifuge at 150,000 × *g* for 15 min at 4 °C.
7. Transfer the supernatants (soluble fractions) in a 96-well microplate at 4 °C.

3.1.3 Dot-Blot

1. Soak the nitrocellulose membrane in 1× PBS for 10 min at room temperature.
2. Assemble the Bio-Dot® microfiltration apparatus (Bio-Rad).
3. Lay the membrane on the sealing gasket so that it covers all of the holes and remove any air bubbles trapped between the membrane and the sealing gasket (*see* **Note 4**).
4. Place the sample template on top of the membrane and finger-tighten the four screws (*see* **Note 5**).
5. Attach a vacuum pump to the flow valve with a waste trap set up and positioned between the vacuum outlet and the flow valve. Turn on the vacuum pump (−20 kPa pressure) and set the flow valve to apply vacuum to the apparatus (flow valve position 2).
6. Repeat the four screws tightening process (*see* **Note 6**).
7. Set the flow valve to stop vacuum (flow valve position 1) and turn off the vacuum pump (*see* **Note 7**).

8. Apply 100 μL 1× PBS to all 96 wells.
9. Turn on the vacuum pump and set the flow valve to position 2. Gently remove the buffer from the wells. As soon as the buffer drains from all the wells, set the flow valve to position 1 and turn off the pump.
10. Fill the appropriate wells with 50 μL of supernatant from solubilization and ensure that there are no air bubbles in the wells.
11. Allow the samples to filter through the membrane by gravity for 5 min.
12. Turn on the vacuum pump, set the flow valve to position 2, and allow complete sample filtration with vacuum.
13. With vacuum applied, loosen the four screws and remove the sample template.
14. Set the flow valve to position 1, turn off the vacuum pump, and remove the nitrocellulose membrane.
15. Block the nitrocellulose membrane with 30 mL blocking buffer and immunodetect the protein of interest using a specific antibody.
16. Confirm the optimal conditions of solubilization by SDS-PAGE and Western blot analysis and Western blot–based thermostability assay.

3.2 Membrane Protein Stability Assessment

The Western blot–based thermostability assay is adapted from [8].

3.2.1 Western Blot–Based Thermostability Assay

1. Solubilize the protein of interest in each detergent condition to be evaluated and in a final volume of 700 μL.
2. Aliquot 50 μL of solubilisate into 14 PCR microtubes.
3. Using a thermocycler, submit the PCR microtubes to a gradient of temperature from 25 °C to 80 °C for 30 min.
4. Immediately transfer to fresh microtubes and centrifuge at 20,000 × *g* for 40 min at 4 °C.
5. Analyze the supernatants by SDS-PAGE and Western blot using an antibody raised against the protein of interest.
6. Quantify the intensity of the immunodetection signal, corresponding to the protein of interest, plot as function of temperature, and fit to the Boltzmann equation to determine the T_m.

3.2.2 Radioligand Binding Assay

This protocol is developed for A_{2A} receptor and is adapted from [9]. A single protein concentration and a single radioligand concentration are used to compare the different protein preparations. This assay is performed in triplicate for each protein sample.

1. Prepare the purified membrane proteins at 24 μg/mL in their own buffer.
2. Using a 96-well microplate with a U-bottom, aliquot 50 μL of protein sample at 24 μg/mL into two wells (well #1 for the total signal and well #2 for the nonspecific signal).
3. Add 10 μL of radioligand solution to the well #1 (total signal) and 10 μL of nonspecific ligand solution to the well #2 (nonspecific signal).
4. Mix by pipetting up and down, and incubate at 4 °C for 2 h.
5. Add 60 μL of γ-globulin solution and 120 μL PEG6000 solution to each well (*see* **Note 8**).
6. Mix by pipetting up and down, and incubate at 4 °C for 15 min.
7. Using the FilterMate Harvester:
   (a) Rinse twice the PEI-coated GF/B filter 96-well microplate with 200 μL of wash buffer.
   (b) Transfer the protein samples to the PEI-coated GF/B filter 96-well microplate.
   (c) Wash four times the PEI-coated GF/B filter 96-well microplate with 200 μL of wash buffer.
   (d) Dry the PEI-coated GF/B filter 96-well microplate for 2 min under vacuum.
8. Seal the bottom of PEI-coated GF/B filter 96-well microplate with the adhesive BackSeal.
9. Add 25 μL of scintillation reagent to each well.
10. Seal the top of PEI-coated GF/B filter 96-well microplate with the adhesive TopSeal.
11. Incubate the plate for 1 h at room temperature.
12. Top count the CPM for 5 min per well using a microplate counter. Determine the specific signal by subtracting the nonspecific signal (well #2) from the total signal (well #1).

3.2.3 ATPase Activity Assay

This protocol is developed for BmrA transporter and is adapted from [11].

1. Prepare a range of 20 μL Pi standards at 0 mM, 2 mM, 4 mM, and 8 mM in 1× TBS from the Pi standard stock solution.
2. Prepare the membrane protein samples at 0.5 mg/mL in their own buffer.
3. Using a 96-well microplate with a transparent flat-bottom, prepare 2 series of 10 μL for each Pi standards and protein samples (series #1 for total signal, series #2 for nonspecific signal).

4. Add 40 μL of reaction buffer to each well.
5. Add 2 μL of 150 mM ATP to each protein sample well, and 2 μL 1× TBS to each Pi standard well.
6. Add 2 μL of 1× TBS to series #1 well, and 2 μL of hot 6 mM orthovanadate to series #2 well (*see* **Note 9**).
7. Incubate at 37 °C for 60 min under gentle agitation.
8. Add 30 μL 10% SDS to stop the reaction.
9. Incubate at room temperature for 10 min under gentle agitation.
10. Add 180 μL of revelation buffer (freshly prepared).
11. Incubate at 37 °C for 20 min under gentle agitation.
12. Measure the absorbance at 620 nm.
13. Calculate for each protein sample the total ATPase activity (series #1) and the nonspecific ATPase activity (series #2) using the corresponding standard curve.
14. Determine the specific ATPase activity for each protein sample by subtracting the nonspecific signal (series #2) from the total signal (series #1).

3.3 Behavior of Membrane Protein in Solution

3.3.1 Clear-Native-PAGE

1. Add appropriate volume of 5× native loading buffer to the membrane protein samples leading to 1× final concentration (*see* **Note 10**).
2. Load the sample in a 4–15% Mini-PROTEAN® TGX™ precast protein gel (Bio-Rad) and run the gel for 70 min at 4 °C using CN-PAGE anode and cathode buffers.
3. Detect the protein by Coomassie staining or Western blot for an immunodetection (*see* **Note 11**).

3.3.2 Glutaraldehyde Cross-Linking Assay

1. Prepare a range of glutaraldehyde stock solutions in water: 0%, 0.0625%, 0.125%, 0.25%, 0.5%.
2. Aliquot 10 μL of purified membrane protein into five microtubes.
3. Add 2.5 μL of different glutaraldehyde stock solutions in order to reach 0%, 0.0125%, 0.025%, 0.05%, and 0.1% final concentrations.
4. Incubate the samples on ice for 2 h.
5. Stop the crosslinking reaction by adding 3.5 μL of 5× SDS-DTT loading buffer.
6. Analyze the samples by SDS-PAGE and Western blot.

3.3.3 Size Exclusion Chromatography (SEC) Analysis

1. Equilibrate a Superdex™ 200 Increase 5/150 GL column (GE Healthcare) with a flow rate of 0.1 mL/min at 4 °C with 3 column volumes (CV) of eluent buffer corresponding to the buffer of the membrane protein to be analyzed (*see* **Notes 12** and **13**).
2. Inject the membrane protein sample onto the column (volume $\leq$ 50 μL).
3. Elute the protein at 4 °C with 1.5 CV of eluent buffer at 0.1 mL/min.
4. The elution fractions can be analyzed by SDS-PAGE and Western blot to confirm the presence of the membrane protein.

4 Notes

1. As the cell lysate will warm up, due to the high energy mixing with glass beads, make sure to use an ice-water cooling jacket with the Bead Beater homogenizer to control the temperature during the cell disruption. The 2 min cool down step between two cell lysis cycles is critical to avoid protein denaturation.
2. 2× stock solutions of each detergent or combination of detergents are prepared in 1× PBS, 1× PIC at ≥20-fold the Critical Micelle Concentration (CMC) in order to reach a final concentration of ≥10 CMC in the solubilization tests.
3. Keep one well with no membrane that will serve as negative control and one well with 2% SDS that will serve as positive control. An example of 96-well plate solubilization conditions is described in [7].
4. The membrane should not extend beyond the edge of the gasket after the Bio-Dot apparatus is assembled.
5. Use a diagonal crossing pattern when tightening the screws to ensure uniform application of pressure on the membrane surface.
6. Tightening the four screws while vacuum is applied ensures a tight sealing, preventing leaking between the wells and contamination between slots.
7. To stop the vacuum, set the flow valve to a position that does not expose the apparatus to air. That will maintain a slight depression in the Bio-Dot apparatus preventing leaking between the wells, while the vacuum pump is off.
8. This step will drastically improve the membrane protein retention on the GF/B filter.
9. The orthovanadate inhibits ATPase activity, enabling to determine the nonspecific inorganic phosphate background.

10. The ClearNative gel electrophoresis works without using an anionic dye. Therefore, this method can be used for the separation of membrane proteins prepared with an anionic detergent. For nonionic or zwitterionic detergents, the suitable method is the BlueNative gel electrophoresis, which uses a negatively charged Coomassie Brilliant Blue G250 dye that forms a complex with the protein to enable the gel migration.
11. The choice for the method of protein detection depends on the amount of protein loaded to the gel. The immunodetection by Western blot is more sensitive than the Coomassie staining but requires more steps.
12. For size exclusion chromatography, it is recommended to degas and filter all solutions through a 0.22 μm filter. However, some detergents may interact with the filter membrane and be retained. If in doubt, do not filter the eluent buffer.
13. For size exclusion chromatography, it is recommended to choose the optimal fractionation range of the column to ensure better separation between aggregates and non-aggregates for high molecular proteins.

References

1. Almen MS et al (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biol 7:50
2. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? Nat Rev Drug Discov 5(12):993–996
3. Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. Biochim Biophys Acta 1666(1–2):105–117
4. Duquesne K, Sturgis JN (2010) Membrane protein solubilization. Methods Mol Biol 601:205–217
5. Eshaghi S (2009) High-throughput expression and detergent screening of integral membrane proteins. Methods Mol Biol 498:265–271
6. Agez M et al (2017) Molecular architecture of potassium chloride co-transporter KCC2. Sci Rep 7(1):16452
7. Desuzinges Mandon E et al (2017) Novel systematic detergent screening method for membrane proteins solubilization. Anal Biochem 517:40–49
8. Ashok Y, Nanekar R, Jaakola VP (2015) Defining thermostability of membrane proteins by western blotting. Protein Eng Des Sel 28 (12):539–542
9. Igonet S et al (2018) Enabling STD-NMR fragment screening using stabilized native GPCR: a case study of adenosine receptor. Sci Rep 8(1):8142
10. Nguyen KA et al (2018) Glycosyl-substituted Dicarboxylates as detergents for the extraction, overstabilization, and crystallization of membrane proteins. Angew Chem Int Ed Engl 57 (11):2948–2952
11. Matar-Merheb R et al (2011) Structuring detergents for extracting and stabilizing functional membrane proteins. PLoS One 6(3): e18036
12. Hardy D et al (2018) The yin and yang of solubilization and stabilization for wild-type and full-length membrane protein. Methods 147:118–125
13. Desuzinges Mandon E et al (2017) Expression and purification of native and functional influenza a virus matrix 2 proton selective ion channel. Protein Expr Purif 131:42–50
14. Jawhari A et al (2006) Structure and oligomeric state of human transcription factor TFIIE. EMBO Rep 7(5):500–505

Part III

Analysis in a Cellular Context

Chapter 15

Practical Aspects of Super-Resolution Imaging and Segmentation of Macromolecular Complexes by dSTORM

Leonid Andronov, Jean-Luc Vonesch, and Bruno P. Klaholz

Abstract

Super-resolution fluorescence microscopy allows imaging macromolecular complexes down to the nanoscopic scale and thus is a great tool to combine and integrate cellular imaging in the native cellular environment with structural analysis by X-ray crystallography or high-resolution cryo electron microscopy or tomography. Here we describe practical aspects of SMLM imaging by dSTORM, from the initial sample preparation using mounting media, antibodies and fluorescent markers, the experimental setup for data acquisition including multi-color colocalization and 3D data acquisition, and finally tips and clues on advanced data processing that includes image reconstruction and data segmentation using 2D or 3D clustering methods. This approach opens the path toward multi-resolution integration in cellular structural biology.

Key words Super-resolution microscopy, dSTORM, SMLM, Immunofluorescence, Fluorescence microscopy

1 Introduction

Fluorescence microscopy is a key technique for the specific observation of proteins in their native cellular environment. Unfortunately, the resolution of fluorescence microscopy is limited by the wavelength of light to around 200–300 nm, which limits the studies to the scale of cellular organelles. The recently emerged super-resolution fluorescence microscopy allows to overcome this limit and improves the resolution by about one order of magnitude, i.e., down to the nanoscopic scale of large macromolecular complexes. Super-resolution microscopy is therefore a powerful technique that bridges the gap between high-resolution crystallographic and cryo electron microscopy (cryo-EM) structures of proteins and their functions in the cellular context [1].

Among the super-resolution techniques, single-molecule localization microscopy (SMLM), which includes techniques such as

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_15,

(direct) stochastic optical reconstruction microscopy ((d)STORM) [2, 3], photoactivated localization microscopy (PALM) [4], and point accumulation for imaging in nanoscale topography (PAINT) [5], has proven to produce the highest resolution even using basic optical setups and common sample preparation protocols. SMLM relies on stochastic switching of fluorophores between a fluorescently active "on" and a fluorescently inactive "dark" state that can be controlled by the intensity of the excitation light and composition of the sample mounting medium. Using appropriate buffer and imaging conditions, many different fluorophores can be made suitable for SMLM, including organic fluorophores [6] and fluorescent proteins [7]. SMLM not only provides super-resolution images but also gives access to properties of individual molecules, allowing cluster analysis [8], segmentation [9], colocalization estimation [10], molecular counting [11], and probing of stoichiometry of macromolecular complexes [12].

In this chapter, we describe how to perform a typical dSTORM experiment from sample preparation to image acquisition and data processing, which includes event list generation, image reconstruction and post-processing of SMLM data that can eventually be analyzed and segmented by 2D and 3D clustering methods.

2 Materials

2.1 Cell Culture and Immunolabeling

In addition to your favorite adherent cell line, you need:

1. Glass-bottom petri dishes with a diameter of 35 mm for cell culture (CELLView, Greiner Bio-One).
2. Phosphate-buffered saline (PBS), 1 and 10 times concentrated (PBS 1× and 10×).
3. Paraformaldehyde (PFA) 4% solution in PBS.
4. 0.1% Triton X-100 in PBS (PBS/Tx).
5. Primary antibody.
6. Secondary antibody coupled to compatible fluorophores (if primary antibody was not directly labeled).
7. Bovine serum albumin (BSA).
8. Normal goat serum (NGS).
9. Fetal bovine serum (FBS).

   Compatible fluorophores:

   Alexa Fluor family (Thermo Fisher Scientific): Alexa 647 (the most commonly used fluorophore for SMLM), Alexa 555, Alexa 532, Alexa 488; Atto 488 (ATTO-TEC GmbH); Cy3B [13] and others [6].

2.2 Mounting Media: Composition and Choice

Gloxy [6]: Cysteamine hydrochloride 50 mM, glucose 10% w/v, glucose oxidase 0.5 mg/ml, catalase 40 μg/ml in TN buffer (50 mM Tris (pH 8.0) and 10 mM NaCl).

We use the following stock solutions:

1. Cysteamine hydrochloride (Sigma-Aldrich M6500, 1 M in H_2O, stored at −20 °C).
2. Glucose 400 g/l, stored at −20 °C.
3. Glucose Oxidase (Sigma-Aldrich G2133, 5 mg/ml, stored at 4 °C).
4. Catalase (Sigma-Aldrich C3515, 4 mg/ml, stored at 4 °C).

Vectashield/TDE [14]: Vectashield 20% v/v, TDE 70% v/v in PBS.

1. Vectashield® Antifade Mounting Medium (Vectorlabs H-1000, stored at +4 °C).
2. 2,2′-Thiodiethanol (TDE, Sigma-Aldrich 166782).

   Mix 20% v/v Vectashield with 70% v/v TDE and 10% v/v PBS 10×. For mounting, incubate in solutions with increasing concentrations of TDE in PBS: 10%, 25%, 50%, 10 min each. Then replace with the final solution, Vectashield/TDE.

   OxEA [15]: Cysteamine hydrochloride 50 mM, OxyFluor™ 3% v/v, sodium DL-lactate 20% v/v in PBS adjusted to pH 8–8.5 with 1 M aqueous NaOH.

We use the following stock solutions:

1. Cysteamine hydrochloride (Sigma-Aldrich M6500, 1 M in PBS, stored at −20 °C).
2. OxyFluor™ (Sigma-Aldrich SAE0059, stored at −20 °C) [16].
3. Sodium DL-lactate (Sigma-Aldrich L1375, 60% w/w syrup, stored at +4 °C).

2.3 Instrumentation

The super-resolution experiments were performed on a Leica SR GSD system built on the base of the DMI6000 B inverted wide-field microscope, consisting of the HCX PL APO 100×/1.47 Oil CORR TIRF PIFOC objective and 1.6× magnification lens that provide an equivalent pixel size of 100 nm on an EMCCD camera (Andor iXon3 DU-897U-CS0-#BV); continuous wave fiber lasers (MPBC Inc., 488 nm 300 mW, 532 nm 1000 mW, 642 nm 500 mW), a diode laser (405 nm 30 mW) and the suppressed motion (SuMo) sample stage.

3 Methods

3.1 Specimen Preparation

Procedure (adapted from a Leica protocol):

1. Seed cells in glass-bottom Petri dishes $d = 35$ mm (CELLView, Greiner Bio-One) at a density of ~100 cells/mm^2.
2. Aspirate the cell culture medium.
3. Wash briefly 1× with PBS.
4. Fix with 4% PFA in PBS for 20 min.
5. Permeabilize with PBS/Tx twice for 10 min.
6. Incubate with primary antibody overnight at 4 °C (*see* **Notes 1** and **2**).
7. Wash 3× with PBS/Tx over 2 h each time.
8. Incubate with secondary antibody for 2 h at RT.
9. Wash 3× with PBS/Tx over 2 h each time.
10. Postfix with 4% PFA in PBS for 10 min (*see* **Note 3**).
11. Wash briefly 3× with PBS.
12. Store at 4 °C in PBS until mounting.
13. Before imaging, replace PBS with an appropriate mounting medium.

The mounting medium for dSTORM (also known as the imaging buffer) has to be chosen based on the following criteria:

1. Fluorophores used. Alexa-647 works best in Gloxy or Vectashield-containing media, while OxEA allows to use a wider range of dyes, e.g., for multi-color imaging. Vectashield-containing media are usually unsuitable for the detection of light with wavelengths $\lesssim$550 nm and for excitation with wavelengths $\lesssim$500 nm due to autofluorescence.
2. Depth of the region of interest (ROI) inside the sample. A ROI close to the coverslip (<200 nm) can be imaged with total internal reflection fluorescence microscopy (TIRF) which allows to reduce background fluorescence. This needs a mounting medium with a refractive index significantly lower than that of glass. Therefore, usually water-based media are used ($n \approx 1.33$), e.g., Gloxy or OxEA. These media, however, are less suitable for deeper (more than several μm from the coverslip) imaging because of spherical aberrations due to the refraction index mismatch. For such situations, buffers with a refractive index close to that of glass ($n \approx 1.52$) are better suited, e.g., Vectashield/TDE. Using highly inclined and laminated optical sheet (HiLo) illumination, the sample is excited with a light beam under an angle smaller than the critical angle for the given coverglass/specimen interface that

allows to image deeper in the sample as compared to TIRF microscopy but allows to keep low background intensity [17]. HiLo illumination can be used with any mounting medium.

3. Labeling method. Cysteamine- or β-mercaptoethanol-containing buffers may cleave antibodies which can decrease the specific labeling and increase background, while TDE-containing media destabilize phalloidin labeling [18].
4. Possibility to refresh the buffer. Gloxy buffer needs to be replaced each 2–4 h, OxEA—each 6–12 h, while Vectashield-based buffers do not need to be replaced.

3.2 Basic SMLM Experiment

Workflow of acquisition processing:

A typical acquisition session is composed of the following steps.

1. Preview. It consists of the search of the ROI and focusing using conventional epifluorescence or TIRF/HiLo illumination with low laser power (1–5% of maximal power). In the case of the TIRF/HiLo mode the best direction (azimuth) of laser illumination to a sample has to be chosen as well. The step is completed by capturing a wide-field image (Fig. 1a). It is important that the dSTORM experiment will be carried out under the same conditions (x, y, z position, epi/TIRF, direction of laser illumination if TIRF/HiLo) to have a direct comparison with the classical (epifluorescence) microscopy image and avoid artifacts in the super-resolution image.
2. Pumping. Illumination of the sample with high intensity light in order to bring most fluorescent molecules into the dark state. At the very beginning, when all the molecules are in the bright state, due to very strong excitation, the fluorescence is very intense because all the dyes emit at the same time. After a little time (in the order of seconds, depending on kinetics of the fluorophore and excitation intensity), with occupation of the dark states and depletion of the on state, the intensity of fluorescence decreases because progressively less molecules stay in the on state until the number of shining molecules becomes less than one per diffraction-limited region. At this time one begins to see the light from each "on" molecule separately, and it is the time to start the acquisition, the next step of the experiment.
3. Acquisition. With constant laser intensity, which may be as high as during pumping or lower, one gets a number of shining single-molecule events ("blinks") per frame with certain exposure time (Fig. 1b). This number is progressively dropping with photobleaching of the molecules, but it can be increased using "backpumping" (e.g., with a 405-nm laser). With gradual increase of backpumping laser intensity, it is possible to keep

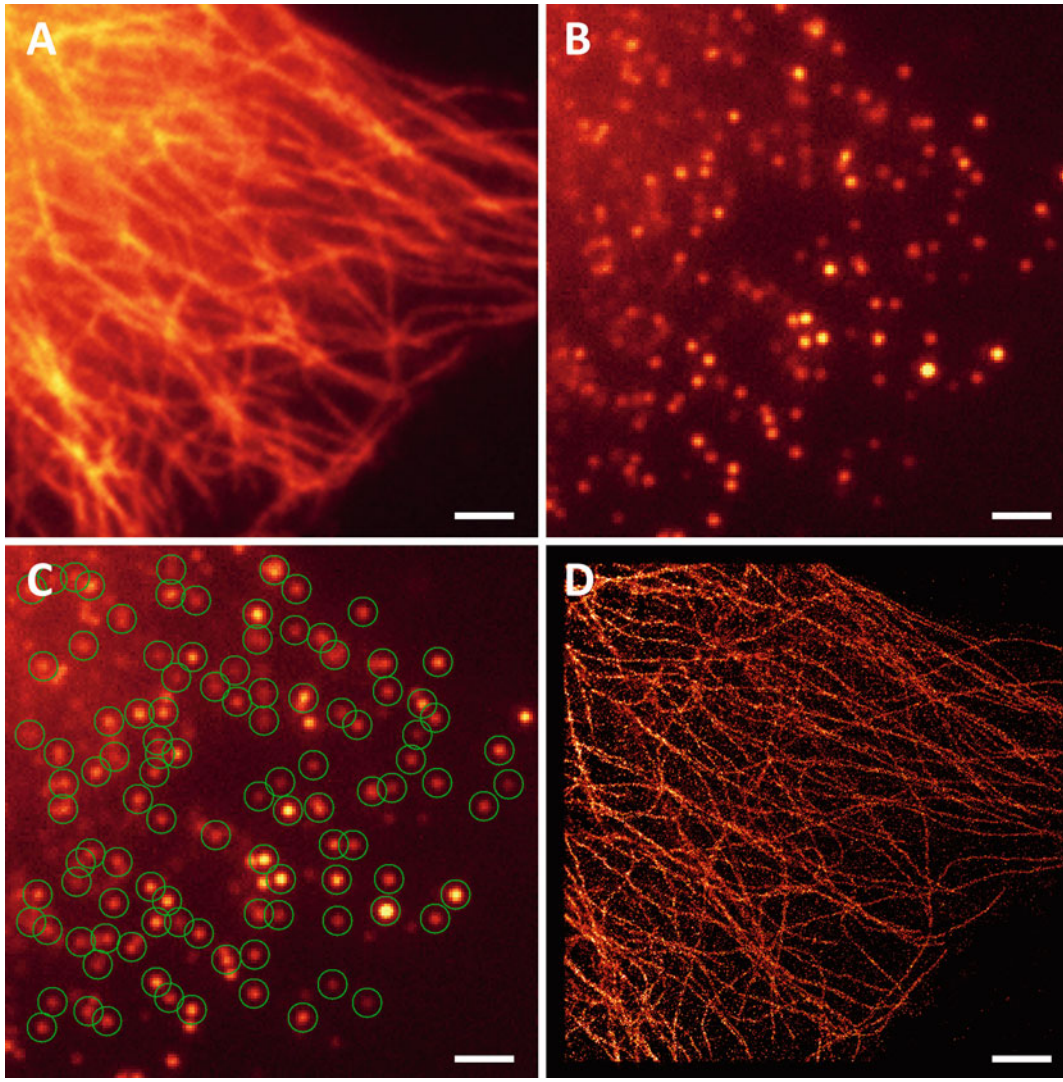


Fig. 1 Example of a typical dSTORM data acquisition cycle (see methods). HeLa cells, β-tubulin marked by Alexa Fluor-647-conjugated secondary antibodies and mounted in the "Vectashield/TDE" medium. (**a**) Preview in HiLo mode. (**b**) One of the frames of the acquisition (exposure time is 50 ms/frame). (**c**) Circles indicate the spots that are selected by the Leica LAS AF software as single-molecule localizations. (**d**) Final reconstructed super-resolution image using 17,622 frames with 1,160,875 localization events (collected here over 15 min); the image is in the histogram mode with 20 nm/pixel, corrected for drift and rendered in SharpViSu [29]. Scale bars, 2 μm

the number of events per frame at a nearly constant value, until most of the molecules become bleached and the acquisition is finished.

4. Post-processing. Although real-time super-resolution image reconstruction is possible depending on the installed software, a post-processing step is usually required to improve the result

by optimizing the processing parameters. For example, to reduce the noise level, it is necessary to increase the threshold of single-molecule detection, if it was quite low during the experiment. Vice versa, if the threshold was pretty high, it is desirable to reduce it in order to get more events and obtain a better reconstructed image. The frames with undesirable artifacts (non-blinking areas, especially in the very first frames) should be truncated, thus also improving the results.

3.3 Experimental Parameters of dSTORM Data Collection and Processing

Numerous experimental parameters that can be changed by the user are summarized here.

1. Laser power, can be chosen differently for pumping, acquisition, and backpumping. Strong excitation intensity during acquisition increases the speed of "blinking," reduces the number of "on" molecules, and ensures faster completion of the acquisition. However, too strong excitation can also lead to increase of the background brightness and reduction of the number of the localized molecules. This parameter has to be adjusted for each combination of fluorophore-imaging buffer. Backpumping intensity is normally set to zero for pumping and for the beginning of acquisition and is turned on with gradually increasing intensity when the number of events per frame drops.
2. Frame exposure time. Ideally, it has to be equal to the average on-time of the fluorophores. If the exposure time is too short, molecules will appear on many consecutive frames, their signal to noise ratio (SNR) will be reduced. On the other side, if the exposure time is too long, the density of localizations per frame will be higher and the SNR can also be reduced because the background is acquired during the whole frame exposure time while the signal from the fluorophores is detected only during their on-time.
3. Illumination mode: epi-/HiLo/TIRF, azimuth of laser illumination (when available), affects the penetration depth and the axial resolution. It is advised to make pumping in epi-illumination mode to send higher energy to the specimen and move all the volume of the sample to the dark state; this provides lower background during the acquisition.
4. EM camera gain for EMCCD detectors. Normally it is set up to maximum, but to be diminished in case of saturation due to strong fluorescence and/or high exposure times (i.e., adjust and calibrate the dynamic range of the camera).
5. Threshold of single-molecular detection in photons/pixel. Affects the real-time reconstructed super-resolution image (if available); it can be changed in the post-processing step. Usually lower threshold increases the noise level, a high

threshold reduces the number of events, so a compromise value should be found. It is convenient to follow which spots on the images are detected and adjust the threshold in a way that only the single-molecule spots are detected (Fig. 1c).

6. Pixel size for real-time reconstruction (Fig. 1d), if available. Commonly set to around 20 nm, and can be reduced for post-processing (e.g., 5 or 10 nm sampling).

3.4 3D SMLM Experiment

One of the easiest and most common ways for 3D SMLM imaging is realized by the modification of the point spread function (PSF) of the microscope with an astigmatic aberration.

Calibration. The astigmatic PSF deformation as a function of the axial position of fluorophores should be calibrated by imaging fluorescent beads using several known axial positions of the objective.

1. Install a sample with fluorescent beads with subdiffraction size in the microscope. For example, multi-color TetraSpeck™ Microspheres, 0.1 or 0.2 μm, can be used.
2. Put in place the cylindrical lens or activate astigmatism in the adaptive optics module.
3. Focus the microscope in a way that the images of the beads are closest to circular symmetry.
4. Adjust the imaging parameters (excitation intensity, exposure time, and electron-multiplying gain of the camera) to get good SNR. It is preferable to increase first the excitation intensity in order to keep the exposure time minimal.
5. Image the beads as a Z-stack with a step of 50 nm around ±0.6–1 μm of the focus point determined in **step 3**.
6. Repeat the procedure (**steps 1–5**) for each spectral channel.
7. Focus the microscope on the beads in one of the spectral channels. Image the beads through all the channels in this Z-position. It is necessary to take into account axial chromatic aberration.
8. Fit all the acquired images with a software for single-molecule detection, which will be used for the following experiments. Refer to the manual of the software.

   Once the calibration is done, 3D SMLM experiments can be performed with the conventional SMLM procedure, with the difference that the maximal density of switched-on molecules on a frame should be reduced to avoid overlap of stretched PSF spots.

3.5 SMLM Data Processing

The first step in SMLM data processing is the determination of the localization of each individual fluorophore to provide a full event list from which a super-resolution image can be reconstructed. The

single-molecule localization precision and therefore the best attainable resolution can be determined by the following formula [19]:

$$\left\langle \left(\Delta x^2\right)\right\rangle = \frac{\sigma^2}{N} + \frac{a^2}{12N} + \frac{8\pi\sigma^4 b^2}{a^2 N^2}, \quad (1)$$

where Δx is the standard deviation (SD) of the localization error, σ is the SD of the PSF, N is the number of detected photons, a is the pixel size, and b is the SD of the background noise. For best localization precision, it is therefore important to keep a high number of detected photons from individual fluorophores, low background noise, small PSF, and a sufficiently small camera pixel size (usually around 100 nm).

Acquired images have to be processed for the detection of single molecules (could be done in real time during experiments depending on microscopy system). This will produce a list of localizations, from which a super-resolution image can be built or other information can be extracted using advanced data processing methods. Most dedicated super-resolution microscopes are supplied with a software for single-molecule detection, image reconstruction, and basic post-processing of localization data. In practice, however, specialized software packages are needed for advanced data processing.

Single-molecule detection requires first of all a good choice of the image fitting method. The most popular methods are:

1. Center of mass assessment—the fastest method, but produces good results only in conditions of low density of localizations and low homogeneous background. This is a good method for the initial processing of the data (e.g. during acquisition).
2. Least-square fitting—usually intermediate in speed group of methods and provides best results when the models for the PSF or for the noise are hard to determine. Some of these methods can be used for the localization of overlapping fluorophores [20].
3. Maximum-likelihood estimation—usually the slowest family of algorithms, requires a model of the PSF and of the noise. Gives best results, especially for weak signals, provided that the PSF shape and the noise model are correctly set [21]. Some of these methods can be used for localization of overlapping fluorophores [22].

   Even though algorithms for single-molecule detection are often supplied by the manufacturer of the super-resolution system, the available tools are often not sufficient, especially for difficult imaging conditions, such as overlapping localizations, non-homogeneous background, weak signal, etc. For such conditions, external software packages such as ThunderSTORM [23], B-recs [24], SimpleSTORM [25], and others (reviewed in [26]) can be recommended.

3.6 Super-Resolution Image Reconstruction

Unlike in conventional microscopy, the image in SMLM cannot be acquired in a direct optical way, but should be calculated from the list of localizations, obtained after fitting. Therefore, there are several methods to calculate an image from SMLM data:

1. Histogram mode. The gray value of a super-resolution pixel equals to the number of events detected within the pixel's area. The fluorescence intensity (number of photons) is not taken into account. The pixel size chosen by the user strongly affects the gray levels of the image: in case of too small pixel sizes, the image would contain only "zeros" and "ones," i.e., become binarized. This mode is fast to calculate, but it can produce noisy and pixilated images, especially for weak density of localizations.
2. "Gaussian" mode. Every localization is represented as a Gaussian function with the width equivalent to the localization precision of the fluorophore, taking into account the number of detected photons. This image is slower to calculate than the histogram image, but it provides a smooth representation and allows to decrease pixel size without risk of pixilation. However, this mode can reduce the resolution of the image [27].
3. Local density mode. The pixel's gray values are calculated to be proportional to the local density of fluorophores in the neighborhood of the pixel. The local density can be estimated, e.g., using Voronoi diagrams [9] or Delaunay triangulations [27]. Even though this method is slower than the others (but still reasonable as in the order of minutes), it provides smooth images with preserved resolution even for weak signals.

3.7 Post-processing of Localization Data

Post-processing of localization data includes methods such as correction of drift and chromatic aberrations that are essential for most of experiments and advanced processing methods, such as clustering and colocalization analysis that are necessary for some experiments, depending on biological question and sample type or acquisition setup (e.g., multi-color imaging or 3D data acquisition).

For drift correction, there are two most feasible possibilities: (1) correction using fiducial markers and (2) correction with cross-correlation. For the first approach, photostable fiducial markers, such as fluorescent beads, gold nanoparticles, or quantum dots [28], need to be incorporated and fixed within the sample, usually on the surface of the coverslip. The second approach does not need fiducials and uses properties that are intrinsic to localization data and rely only on computing [29]. The cross-correlation method works well for contrasted structures with well-defined shape but can be less efficient for diffuse structures. It can also be recommended for imaging far from the coverslip because the fiducials usually can be robustly immobilized only on a glass surface.

For multi-color imaging, chromatic aberrations of the objective can cause a shift between images, captured with different wavelengths. To correct for this, first, the aberrations should be calibrated by imaging a multi-color fiducial marker (e.g., beads) with the spectral channels to be used for the experiments. The shift detected from the fiducials should be subtracted from the single-molecule coordinates [30].

If the on-time of fluorophores is higher than the frame exposure time, their images will appear on several consecutive frames. These localizations can be combined by searching localizations within a circle of a given radius (usually around 50 nm) around every localization in following frames. Using this option, it is possible to increase the number of photons per localization and therefore the localization precision. An inconvenience is that localizations of different fluorophores can be combined by error if they appear on consecutive frame (this is less likely when the densities of switched-on fluorophores is low). In some cases, it can be also useful to remove localizations that appear on too many consecutive frames, eliminating non-blinking regions that can produce artifacts.

Localization event lists can be conveniently analyzed with the help of histograms as a function of the number of photons per localization (Fig. 2) and the number of localizations per frame (Fig. 3). Localizations with too low photon counts can be removed, thus improving average localization precision. Frames with too high number of localizations that correspond to very dense

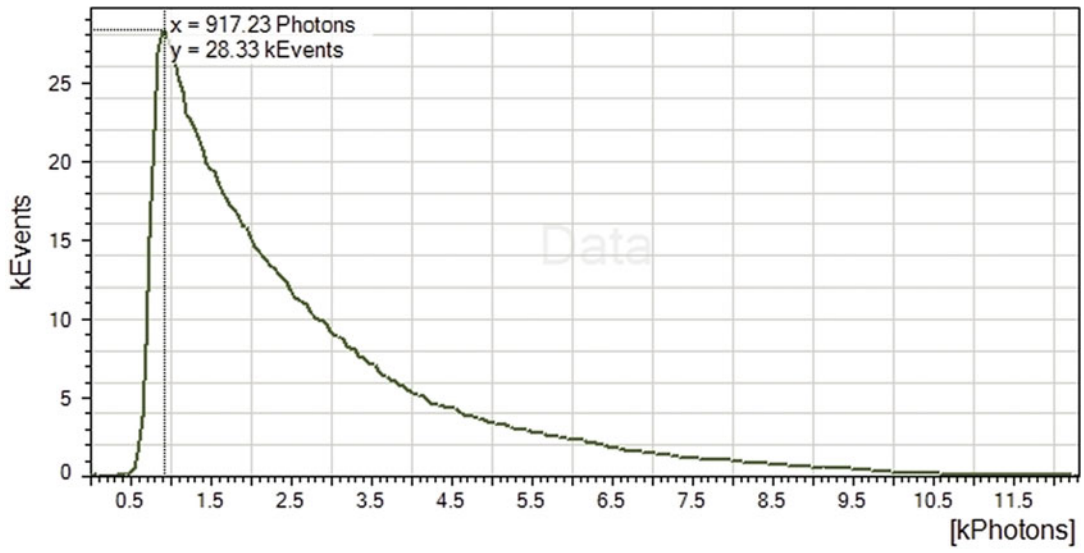


Fig. 2 Typical histogram of the photon count of localizations of the Alexa Fluor-647 dye. The exposure time was 50 ms/frame, the detection threshold is 60 photons/pixel, 17,622 frames were analyzed producing 1,160,875 events, the average photon count is 2675 photons/event. The analysis was done in the Leica LAS AF software

Fig. 3 Typical histogram of the number of localizations per frame for the Alexa Fluor-647 dye. The curve is built using blocks of 100 frames each for smoothing reasons. Note increases in the number of localizations (middle part of the graph) due to an increase in the intensity of the "backpumping" with a 405-nm laser. Exposure time is 50 ms/frame, detection threshold is 80 photons/pixel; 18,620 frames with 1,161,303 events were analyzed. The analysis was done in the Leica LAS AF software

localizations that may compromise single-molecule detection can be removed as well. An additional problem for data quantification can be relocalization events (fluorophores that appear more than once during the acquisition); these can be addressed by taking into account the blinking behavior of fluorophores [31].

3.8 SMLM Data Segmentation and Cluster Analysis, from 2D to 3D

Advanced SMLM data processing techniques include co-localization and cluster analysis. While for classical microscopy data, these are commonly done on images, analysis of SMLM images is less optimal, because the image is not the primary data type, and moreover, they are not built in a unique way (*see* Subheading 3.6). To get most information from the data, it is preferred to process directly the localization coordinates.

Cluster analysis or segmentation aims for the detection of accumulations (clusters) of fluorescently labeled objects, such as proteins or nucleic acids. These detected clusters should then be analyzed for their size, density distribution, etc. Two types of cluster analysis can be distinguished: determination of global properties of data and detailed analysis of individual clusters. For the first part of the analysis, the method of choice is Ripley's K- and L-functions [32, 33]. It allows to, first, test whether or not the experimental distribution follows a given pattern, and, second, to find the characteristic size of clusters [8]. A similar method was also developed for the estimation of colocalization between two experimental distributions [10], it provides a colocalization value for a region that reflects the global colocalization extent within this region.

For detailed cluster analysis, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method [34] can be used. However, to get optimal results, it needs optimization of its parameters that can introduce a bias in the results. Recently emerged Voronoi tessellation-based methods [9, 35–37] allow not only for an unambiguous local density estimation and visualization of SMLM data but also for testing of datasets for clustering via comparison of the Voronoi polygons built on the experimental data with that of the reference distribution of points and for automated unbiased segmentation without parametrization [9, 37].

Some of these processing aspects can be found in packages for microscope control, but in many cases, external tools have to be used. There are several packages dedicated for processing of localization data, namely ThunderStorm [23] that includes post-processing aspects such as image reconstructions in several modes, filtering and combination of localizations, drift correction, and estimation of colocalization. ViSP [38] is a convenient tool for 3D visualization and basic segmentation. Lama [39] provides tools for quality control, resolution estimation, cluster analysis with Ripley's functions, DBSCAN and Ordering Points To Identify the Clustering Structure (OPTICS) [40] algorithms, coordinate-based colocalization, and estimation of stoichiometry of molecular complexes [12]. MIiSR [41] is focused on colocalization estimation and cluster analysis using the Ripley's, DBSCAN, and OPTICS methods. PALMsiever [42] allows for the combination of localizations, drift correction, DBSCAN cluster analysis, and different rendering possibilities. SharpViSu [29] offers comprehensive tools for the correction of chromatic aberrations and drift, selection and filtering of localizations, different visualization modes, resolution estimation and, through ClusterViSu [9], cluster analysis with the Ripley's function method and Voronoi-based automated segmentation. Our latest development, 3DClusterViSu, allows to perform Voronoi-diagram-based segmentation of 3D data [37]. Clearly, if it can be done for a given experiment, 3D analysis becomes the method of choice nowadays as compared to 2D analysis because it avoids artefacts from superposed structures (*see* details in Supplementary Material in [37]), and it provides direct 3D information about the molecular organization in the cell, thus also facilitating the integration with molecular and atomic structures and opening the path toward cellular structural biology [1].

4 Notes

1. Absence of any blocking reagents and blocking step is possible with highly specific antibodies. If necessary, BSA 1% or a mixture of decomplemented NGS (5%) and FCS (1%) can be used before **step 6**.

2. To avoid aggregates, all stock solutions of antibodies should be centrifuged at 16,000 × *g* for 5 min. All working solutions can be filtered on 0.22 μm filters.
3. The purpose of the post-fixation of the immune complexes is twofold: first, avoiding cleavage of disulfide bounds of the antibodies (destroying the immunoreactivity) in the mounting buffers that often contain reducing agents like cysteamine or β-mercaptoethanol. Moreover, cysteamine solutions are also weakly basic (pH 8–8.5), weakening the bindings. Second, reducing the mobility of the antibodies on the antigens (off and on reactions) that may affect the quality of super-resolution images.

Acknowledgments

We thank Yves Lutz from the Imaging Centre for discussions on the adaptation of a Leica protocol and his contribution during earlier published work. This work was supported by CNRS, Association pour la Recherche sur le Cancer (ARC), Institut National du Cancer (INCa), Ligue nationale contre le cancer (Ligue), Agence National pour la Recherche (ANR), and the USIAS research fellowship program of the University of Strasbourg. The super-resolution microscope setup was supported by the Alsace Region and by the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INSB-05-01) and Instruct-ERIC.

References

1. Orlov I, Myasnikov AG, Andronov L et al (2017) The integrative role of cryo electron microscopy in molecular and cellular structural biology. Biol Cell 109:81–93
2. Rust MJ, Bates M, Zhuang X (2006) Stochastic optical reconstruction microscopy (STORM) provides sub-diffraction-limit image resolution. Nat Methods 3:793–795
3. van de Linde S, Löschberger A, Klein T et al (2011) Direct stochastic optical reconstruction microscopy with standard fluorescent probes. Nat Protoc 6:991–1009
4. Betzig E, Patterson GH, Sougrat R et al (2006) Imaging intracellular fluorescent proteins at nanometer resolution. Science 313:1642–1645
5. Sharonov A, Hochstrasser RM (2006) Wide-field subdiffraction imaging by accumulated binding of diffusing probes. Proc Natl Acad Sci U S A 103:18911–18916
6. Dempsey GT, Vaughan JC, Chen KH et al (2011) Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. Nat Methods 8:1027–1036
7. Lippincott-Schwartz J, Patterson GH (2009) Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. Trends Cell Biol 19:555–565
8. Owen DM, Rentero C, Rossy J et al (2010) PALM imaging and cluster analysis of protein heterogeneity at the cell surface. J Biophotonics 3:446–454
9. Andronov L, Orlov I, Lutz Y et al (2016) ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. Sci Rep 6:24084
10. Malkusch S, Endesfelder U, Mondry J et al (2012) Coordinate-based colocalization analysis of single-molecule localization microscopy data. Histochem Cell Biol 137:1–10
11. Lee S-H, Shin JY, Lee A et al (2012) Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy

(PALM). Proc Natl Acad Sci U S A 109:17436–17441
12. Fricke F, Beaudouin J, Eils R et al (2015) One, two or three? Probing the stoichiometry of membrane proteins by single-molecule localization microscopy. Sci Rep 5:14072
13. Cooper M, Ebner A, Briggs M et al (2004) Cy3B™: improving the performance of cyanine dyes. J Fluoresc 14:145–150
14. Olivier N, Keller D, Rajan VS et al (2013) Simple buffers for 3D STORM microscopy. Biomed Opt Express 4:885–899
15. Nahidiazar L, Agronskaia AV, Broertjes J et al (2016) Optimizing imaging conditions for demanding multi-color super resolution localization microscopy. PLoS One 11:e0158884
16. Adler H, Spady G (1997) The use of microbial membranes to achieve anaerobiosis. J Rapid Methods Autom Microbiol 5:1–12
17. Tokunaga M, Imamoto N, Sakata-Sogawa K (2008) Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat Methods 5:159–161
18. Staudt T, Lang MC, Medda R et al (2007) 2,2′-thiodiethanol: a new water soluble mounting medium for high resolution optical microscopy. Microsc Res Tech 70:1–9
19. Thompson RE, Larson DR, Webb WW (2002) Precise nanometer localization analysis for individual fluorescent probes. Biophys J 82:2775–2783
20. Holden SJ, Uphoff S, Kapanidis AN (2011) DAOSTORM: an algorithm for high- density super-resolution microscopy. Nat Methods 8:279–280
21. Abraham AV, Ram S, Chao J et al (2009) Quantitative study of single molecule location estimation techniques. Opt Express 17:23352–23373
22. Quan T, Zhu H, Liu X et al (2011) High-density localization of active molecules using structured sparse model and Bayesian information criterion. Opt Express 19:16963–16974
23. Ovesný M, Křížek P, Borkovec J et al (2014) ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. Bioinformatics 30:2389–2390
24. Rouault H (2013) Brecs: reconstruction of dense localization microscopy datasets https://github.com/hrouault/Brecs
25. Köthe U, Herrmannsdörfer F, Kats I et al (2014) SimpleSTORM: a fast, self-calibrating reconstruction algorithm for localization microscopy. Histochem Cell Biol 141:613–627
26. Sage D, Kirshner H, Pengo T et al (2015) Quantitative evaluation of software packages for single-molecule localization microscopy. Nat Methods 12:717–724
27. Baddeley D, Cannell MB, Soeller C (2010) Visualization of localization microscopy data. Microsc Microanal 16:64–72
28. Glushonkov O, Réal E, Boutant E et al (2018) Optimized protocol for combined PALM-dSTORM imaging. Sci Rep 8:8749
29. Andronov L, Lutz Y, Vonesch J-L et al (2016) SharpViSu: integrated analysis and segmentation of super-resolution microscopy data. Bioinformatics 32:2239–2241
30. Erdelyi M, Rees E, Metcalf D et al (2013) Correcting chromatic offset in multicolor super-resolution localization microscopy. Opt Express 21:10978–10988
31. Annibale P, Vanni S, Scarselli M et al (2011) Quantitative photo activated localization microscopy: unraveling the effects of Photoblinking. PLoS One 6:e22678
32. Ripley BD (1976) The second-order analysis of stationary point processes. J Appl Probab 13:255–266
33. Nicovich PR, Owen DM, Gaus K (2017) Turning single-molecule localization microscopy into a quantitative bioanalytical tool. Nat Protoc 12:453–460
34. Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. KDD-96 Proceedings 226–231 (http://www.aaai.org)
35. Voronoi G (1908) Nouvelles applications des paramètres continus à la théorie des formes quadratiques. J Reine Angew Math 1908:97–102
36. Levet F, Hosy E, Kechkar A et al (2015) SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. Nat Methods 12:1065–1071
37. Andronov L, Michalon J, Ouararhni K et al (2018) 3DClusterViSu: 3D clustering analysis of super-resolution microscopy data by 3D Voronoi tessellations. Bioinformatics 34 (17):3004–3012
38. Beheiry ME, Dahan M (2013) ViSP: representing single-particle localizations in three dimensions. Nat Methods 10:689–690
39. Malkusch S, Heilemann M (2016) Extracting quantitative information from single-molecule super-resolution imaging data with LAMA—LocAlization microscopy analyzer. Sci Rep 6:34486

40. Ankerst M, Breunig MM, Kriegel H-P et al (1999) OPTICS: ordering points to identify the clustering structure. In: Proceedings of the 1999 ACM SIGMOD international conference on Management of Data. ACM, New York, NY, USA, pp 49–60
41. Caetano FA, Dirk BS, Tam JHK et al (2015) MIiSR: molecular interactions in super-resolution imaging enables the analysis of protein interactions, dynamics and formation of multi-protein structures. PLoS Comput Biol 11:e1004634
42. Pengo T, Holden SJ, Manley S (2015) PALMsiever: a tool to turn raw data into results for single-molecule localization microscopy. Bioinforma Oxf Engl 31:797–798

Chapter 16

Multicolor FRET-FLIM Microscopy to Analyze Multiprotein Interactions in Live Cells

Abdullah Ahmed, Jennifer Schoberer, Emily Cooke, and Stanley W. Botchway

Abstract

The need to describe and understand signaling pathways in live cell is seen as a primary route to identifying and developing targeted medicines. Signaling cascade is also seen as a complex communication and involves interactions between multiple interconnecting proteins. Where subcellularly and how different proteins interact need to be preserved during investigation. Furthermore, these complex events occurring simultaneously may lead to a single or multiple end point or cell function such as protein synthesis, cell cytoskeleton formation, DNA damage repair, or autophagy. There is therefore a need of real-time noninvasive methods for protein assays to enable direct visualization of the interactions in their natural environment and hence overcome the limitations of methods that rely on invasive cell disruption techniques. Förster resonance energy transfer (FRET) coupled with fluorescence lifetime imaging microscopy (FLIM) is an advanced imaging method to observe protein–protein interactions at nanometer scale inside single living cells in real-time. Here we describe the development and use of two-channel pulsed interleave excitation (PIE) for multiple protein interactions in the mTORC1 pathway. The proteins were first tagged with multiple color fluorescent protein derivatives. The FRET-FLIM combination means that the information gained from using standard steady-state FRET between interacting proteins is considerably improved by monitoring changes in the excited-state lifetime of the donor fluorophore where its quenching in the presence of the acceptor is evidence for a direct physical interaction.

Key words mTOR, Enzymes, Protein complex, Fluorescence imaging, Confocal microscopy, Excited-state lifetime, TCSPC, FRET, FLIM GFP-tag, DNA, Three-color

1 Introduction

All cellular processes are driven by the action of proteins, from ion channels across phospholipid membranes to DNA synthesis and cell growth. For example, the mechanistic or mammalian target of rapamycin (mTOR) pathway functions in the coordination of energy, nutrients, and growth factor availability to regulate key biological processes including cellular growth, metabolism, and protein synthesis through the phosphorylation of downstream substrates, ribosomal protein, S6 Kinase 1 (S6K1), and 4E-binding

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_16,

protein 1 (4E-BP1/eIF4E) [1–4]. The phosphorylation of these effectors is to regulate cell growth, aging and adiposity [5], memory [6], immunity [7], and muscle hypertrophy [8]. The human mTOR protein exists as a multi-protein complex with Rheb, raptor, mLST8, PRAS40, and DEPTOR proteins termed mTOR Complex 1 (mTORC1). This complex is the main rapamycin target via the FKBP12 protein. Furthermore, rapamycin differentially inhibits S6Ks and 4E-BP1 toward mRNA translation [9]. It has also been observed that the concentration of rapamycin required for an effect varies significantly between cell lines with a range of nM-mM reported [10]. Majority of our understanding of the mTOR signaling pathway and protein interaction has come from classical cell disruption pull-down assays. Hence, information where within the cell these assemblies are localized that lead to subsequent downstream targets phosphorylation in real time is lost. Cell immunofluorescence staining following cell fixation also has several drawbacks and limitations such as mislocalization of proteins after fixation, poor antibody sensitivity and specificity in some cases, as well as antibodies not reaching their target of interest.

Förster or fluorescence resonance energy transfer (FRET) is an excellent and powerful technique to determine relevant distances in biological processes under physiological conditions [11]. It relies on the non-radiative energy transfer from an excited fluorescent donor molecule to an acceptor molecule with non-excited fluorescence in its close vicinity through a dipole–dipole interaction and may be used to investigate physical interactions between two or more small or macro-molecules. Thus molecular scale distances (1–10 nm) can be measured using energy transfer processes. The very short distances required for this process to occur (<10 nm) mean that the molecules, in this case two or more macromolecules such as proteins, need to be physically close to one another for FRET to occur. The key general requirements are that the donor emission spectrum must overlap sufficiently with the acceptor absorption spectrum while the donor and acceptor dipoles display a mutual molecular orientation. The efficiency E or the rate of the energy transfer k_T can be described and calculated simply using Eq. (1):

$$k_T = \left(\frac{1}{\tau_D}\right)\left(\frac{R_0}{R}\right)^6 \quad \text{or} \quad E = \frac{{R_0}^6}{R + {R_0}^6} \tag{1}$$

where τ_D is the donor excited-state lifetime in the absence of the acceptor, R is the distance between D (donor) and acceptor, and R_0 is the Förster radius. At the Förster radius ($R = R_0$), 50% of the donor molecules will emit fluorescence while the rest will undergo energy transfer. Since the energy transfer process is strongly distance dependent with $1/R^6$, FRET can be used to measure distances and examine molecular interactions on a nanometer spatial

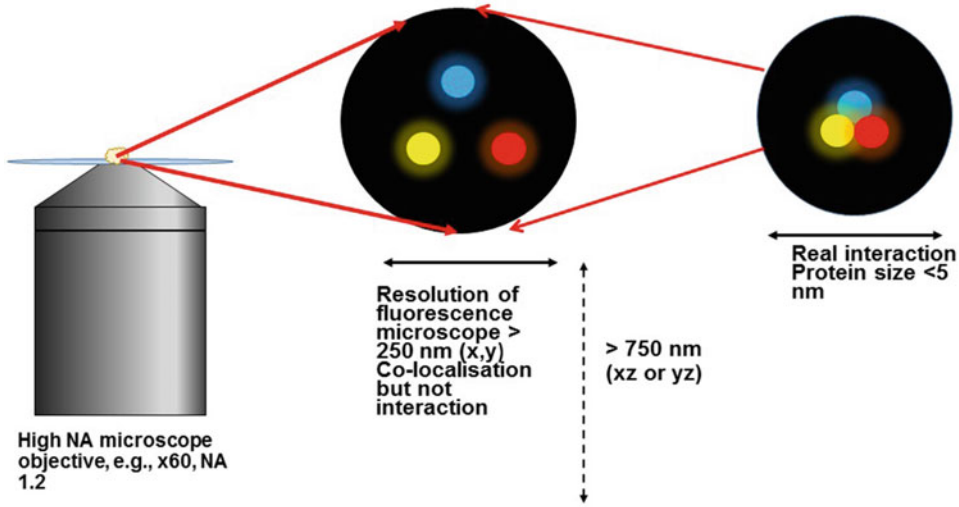


Fig. 1 Schematic representation of colocalization versus interaction of proteins in a confocal volume of a high numerical aperture microscope objective

scale. In experimental conditions, values for FRET efficiencies are typically between ~50% and 5%, thus $R_0 \leq R \leq 1\frac{1}{2}R_0$ since FRET pairs are around 5 nm for R_0.

However, imaging steady-state FRET under microscopy conditions alone may lead to erroneous and false-positive results since the protein size (~5 nm and volume) is less than a factor of 50 compared to that of the microscope objective with a diffraction-limited spot size (~250 nm, Fig. 1). Furthermore, the strict microscopy condition required to be met to avoid false results during normal FRET, including instrument error such as thresholding and emission bleed-through, is rarely accounted for routinely [12].

For multiple acceptors and donors, and going beyond a one to one protein interaction, the efficiency equation above apparently becomes less accurate. The probability for FRET coupling for a donor increases with the number of acceptors. Similarly, the distance sensitivity of FRET also increases with the number of multiple potential acceptors for each donor without a real change in molecular distance [13–15]. It has been suggested that in cases where multiple acceptors modify the FRET coupling, it may be advantageous to work with normalized transfer rates, k'_{T}, as these are additive while efficiency of transfer E (Eq. 1) is not linear with multiple acceptor–donor situations [16] and can be described in Eq. (2) for i number of acceptors (A_i).

$$E = \frac{\sum_i k_{D\to Ai}}{\sum_i k_{D\to Ai} + \frac{1}{\tau_{\mathrm{D}}}} \quad \text{so that} \quad k_{D\to A} = \frac{E}{\tau_{\mathrm{D}}(1-E)} \tag{2}$$

The efficiency of transfer from two acceptors is therefore:

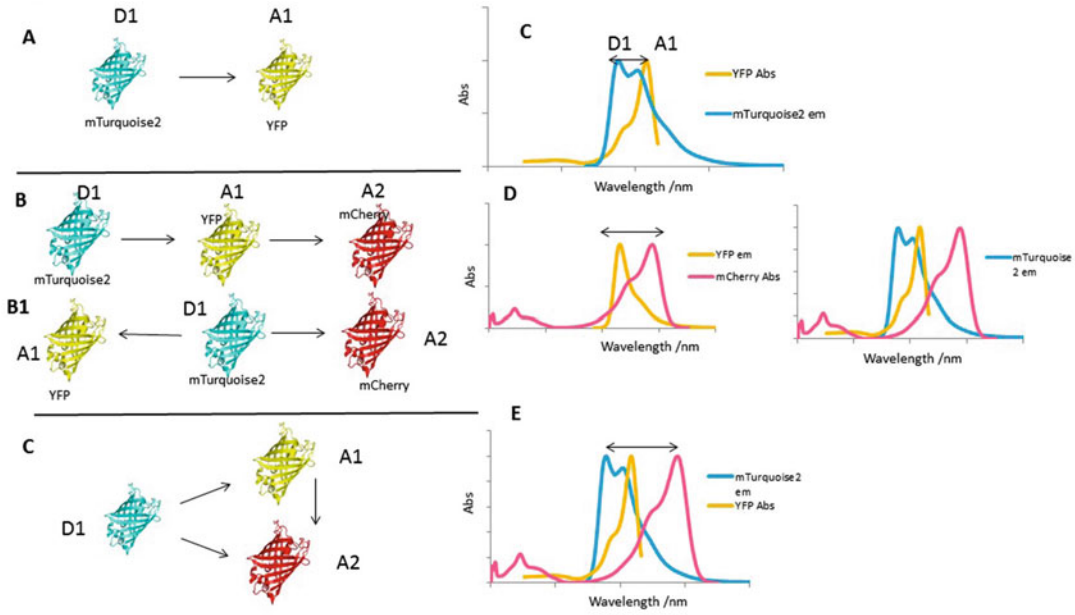


Fig. 2 Simplified types of Förster resonance energy transfer (FRET) multiplexing. (**a**) Standard dual channel (1:1) FRET process. For example, small two molecular probes or intramolecular FRET sensors co-expressed in a cell, each containing a donor (D) and acceptor (A) fluorophores that are spectrally separate but emission of the D overlaps with excitation of A, shown in (**c**). Both fluorophore needs to be within <10 nm. (**b**) Sequential two-step FRET between three fluorophores that form two consecutive FRET donor and acceptor pairs. The acceptor of the first pair acts as the donor for the second acceptor. In (**b**1), the donor (D1) is able to transfer energy to both A1 and A2, so that D interacts with both A1 and A2. But A1 does not interact with A2. (**c**) In this triple-fluorophore FRET used to detect multiple interactions in a complex, the overlap between all three does need to have some spectral overlap. FRET detection between the donors and acceptors can be determined when they form a donor–acceptor triple complex. (**c**, **d**), and (**e**) are representative excitation and emission spectra (donors and acceptors pairs) for the fluorescent probes described in the work. The double-headed arrows indicate the spectral overlap between donor emission and acceptor excitation spectra. (Modified from Ref 17)

$$E = \frac{k_{D\to A1} + k_{D\to A2}}{k_{D\to A} + k_{D\to A} + \frac{1}{\tau_D}} \tag{3}$$

where E is the efficiency, k is the transfer rate, D is the donor, and τ is the lifetime.

The cumulative FRET efficiencies in a complex multiplexed acceptor(s)–donor(s) (Fig. 2) further developed by FRET from single to multiplexed signaling events [17] can be calculated using the normalized transfer rates related as:

$$k'_{\mathrm{T}} = \frac{k_{\mathrm{T}}}{\Gamma + k_{\mathrm{nr}}} = \left(\frac{R_0}{R}\right)^6, \quad E = \frac{k'_{\mathrm{T}}}{k'_{\mathrm{T}} + 1} \tag{4}$$

where k_T is the transferred rate, k'_{T} is the normalized transferred rate, k_{nr} is the non-radiative decay, R is the distance between donor and acceptor, R_0 is the Förster radius, and E is the efficiency.

During FRET, the rate of decay is reduced by a quenching process that depletes the excited state of the donor fluorophore, i.e., the donor fluorescence lifetime is shortened. Thus, by measuring changes in the excited-state lifetime of the donor at each pixel making up an image, steady-state FRET can be more accurately determined. This is described as fluorescence lifetime imaging microscopy or FRET-FLIM [18–22]. Generally, the two proteins under investigation are tagged with green fluorescent protein (GFP) and its variants. In our work, we have used a combination of blue FP (mTurquoise), yellow FP, and red FP (mCherry) as donors and acceptors to investigate some of the multiple complex mTORC1 proteins. We have also applied the FLIM imaging method to investigate DNA structure as well as protein–protein interactions in several systems including mammalians, viruses, and plants [23–27], thus demonstrating wide applicability of FRET-FLIM. We describe here a multi-color FRET-FLIM microscopy of two-channel pulsed interleave excitation (PIE) to analyze multi-protein interactions in live cells.

Measuring the change in donor(s) lifetime in a FLIM experiment provides an elegant way to answer biological questions as well as protein signaling that may involve multiple proteins at the same time. It is important to excite only the donor(s) individually, in the complex, and this is provided by the pulsed interleave excitation (PIE) method (Fig. 3).

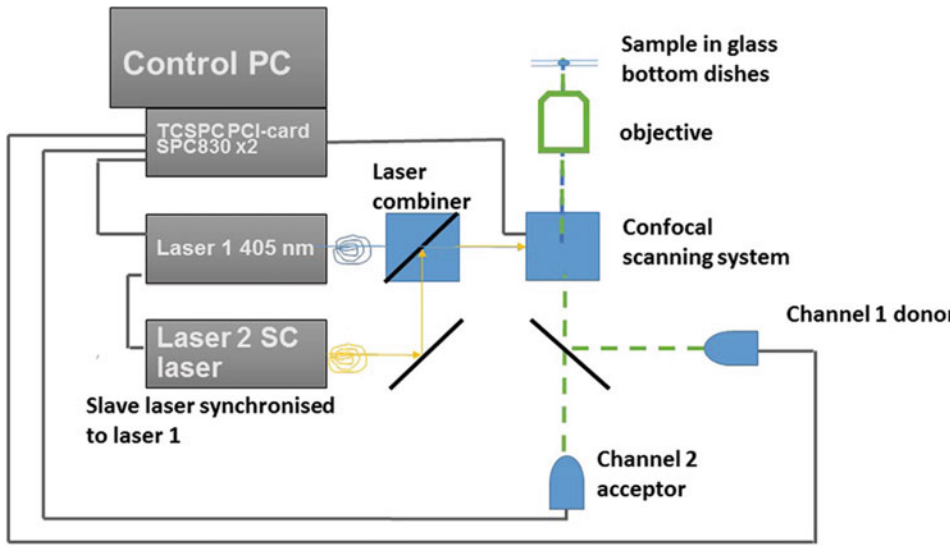


Fig. 3 Schematics of two-channel pulsed interleave excitation fluorescence lifetime imaging microscopy

2 Materials

General laboratory reagents were purchased as analytical grade from Sigma-Aldrich (USA), Millipore (UK), Fisher Scientific (UK) and Thermo Fisher Scientific (UK). Other chemical reagents and kits (all chemicals included) for the molecular biology were obtained from Qiagen (Germany), Machinery-Nagel (USA), and Agilent (UK) unless otherwise stated. All primers were purchased from Integrated DNA Technologies (IDT, Europe) at a 25 nmole (0.27 mg) DNA oligo concentration with standard desalting. FuGENE HD Transfection reagent was purchased from Promega. All cell culture reagents were bought from Gibco (Thermo Fisher Scientific).

1. All primers were designed with a Tm (melting temperature) close to 60 °C, although this varied slightly depending on the GC content. Gene-specific primers with 15 bp extension homologous to vector ends were designed for In-Fusion™ cloning [28], using an automated primer design tool: https://www.oppf.rc-harwell.ac.uk/Opiner/, or for fusing two inserts together using SnapGene® software (from GSL Biotech; available at snapgene.com). Lyophilized primer stocks (100 μM) were resuspended with nuclease-free water. Stocks were diluted 1:10 to make up 10 μM working stocks that were stored at −20 °C.
2. Buffers and media for cloning the fluorescent protein tagged proteins were prepared at room temperature using ultrapure water (deionized water, at 18 MΩ-cm at 25 °C). The S6K1-GFPSpark construct was obtained from Sino Biological (China) and mCherry-raptor and YFP-mTOR from Addgene. Miniprep and Maxiprep kits were purchased from Qiagen (UK).
3. HEK293 cell line was purchased from ATCC® to make sure of origin and purity and free from mycoplasma contamination as well as passage number. Cell culture dishes with glass bottom (35 mm) were bought from MatTek.

3 Methods

3.1 Amplification of cDNA by PCR

PCRs were set up on ice in 0.2 ml PCR tubes or in an eight-strip PCR tube (Star Lab) containing both forward (Fwd) and reverse (Rev) primers at a 10 μM working concentration, Phusion Flash Mix (Thermo Fisher Scientific), template plasmid DNA (100–200 ng) containing cDNA gene to be amplified and sterile water as shown in Table 1. Constituents were gently mixed by

Table 1
PCR volumes

	Amount for one reaction
Phusion flash master mix (2×)	25 μl
Forward primer (10 μM) Reverse primer (10 μM)	1.5 μl 1.5 μl
Template DNA plasmid	1 μl
Sterile water	21 μl
Total volume	50 μl

Table 2
Parameters of thermal cycles used for PCR

Step	Temperature	Time	Cycle
Initial denaturation	98 °C	10 s	1
Denaturation Annealing Extension	98 °C 60 °C 72 °C	1 s 5 s 15 s/kbp	29
Final extension	72 °C	2 min	1
Hold	4 °C	Hold	

pipetting up and down. The PCR tubes were placed into a Veriti™ 96-well thermal cycler (Thermo Fisher Scientific) running a program as listed in Table 2.

3.2 Agarose Gel Electrophoresis of PCR Products

Upon completion of the PCR, 10 μl of filtered DNA loading buffer (0.25% w/v bromophenol in 30% v/v glycerol) was added to each PCR tube and mixed gently by pipetting up and down. A volume of 40 μl of the PCR–dye mix was then loaded onto a pre-cast 1.0% agarose/Tris-borate-EDTA (TBE) gel, commonly used for DNA gel electrophoresis, containing the SYBRSafe stain (Thermo Fisher) as listed in Table 3. Adjacent to the samples, 5 μl of HyperladderTM 1 kb (BioLine) was loaded and run in a 10 × 8 gel tank (MiniRapide) filled with TBE for 120 min at 70 V using a Power Pac 300 (Bio-Rad) power supply until the markers were separated sufficiently. After marker bands were separated, the gel was viewed with a blue light illuminator (Jencons-PLS) to both visualize and verify PCR bands at their respected correct base pair lengths.

3.3 DNA PCR Product Extraction from Agarose and Clean Up

PCR products were visualized using blue light illumination, excised from the gel with a sterile surgical scalpel and transferred to a 1.5-ml tube. The PCR-amplified gene was purified from the gel using a commercially supplied kit (i.e., Machinery-Nagel).

Table 3
Parameters of thermal cycles used for PCR

Composition	Amount
Agarose (sigma-Aldrich)	0.5 g
TBE (1×) (sigma-Aldrich)	50 ml
SYBRSafe (Thermo fisher scientific)	5 μl

Composition of 1.0% TBE agarose gel

Table 4
In-fusion reaction for a two-way fusion

Components	Amount for one reaction
Linearized vector (20–100 ng)	1 μl
PCR product insert (10–100 ng)	2 μl
Fusion enzyme	1 μl
5× fusion buffer	2 μl
Sterile water	4 μl
Total	10 μl

3.4 In-Fusion Reaction with pOPIN Vectors

Ligation of the PCR products with linearized pOPIN vectors were performed according to the manufacturer's recommendations using a Quick-Fusion cloning kit (Biotool). A two-way fusion cloning reaction involving one PCR DNA insert and vector was set up as described in Table 4. For a three-way fusion cloning reaction involving two PCR DNA inserts and vector, 2 μl of each insert with 2 μl of vector was added and the final volume adjusted to 10 μl with sterile water.

1. Transform fusion products into competent *E. coli* (i.e., RecA strain DH5α) by electroporation or heat shock.
2. Incubate bacteria at 37 °C for 30–60 min.
3. Plate different volumes of the bacterial suspension on LB plates supplemented with appropriate antibiotics for selection.
4. Incubate plates at 37 °C overnight.
5. Pick single colonies and grow LB overnight cultures.
6. Purify the recombinant plasmid using a commercially supplied kit, i.e., the QIAprep Spin Miniprep Kit.
7. Optionally, determine the DNA concentration and purity using a spectrophotometer, i.e., NanoDrop Microvolume Spectrophotometer (Thermo Fisher Scientific).

3.5 Construction of S6K1-mTurq2 and Raptor-YFP Plasmid Constructs

1. S6K1-mTurqouise2 construct was generated in a similar manner by insertion into a pOPINE-3C-mTurq2 vector, also provided by the Protein Production UK (PPUK) using the primers below:

   S6K1-mTurq2 Fwd AGGAGATATACCATGAGGCGACG AAGGAGGCGG.

   S6K1-mTurq2 Rev CAGAACTTCCAGTTTTAGATTCA-TACGCAGGTGCTCTG.

2. Raptor-YFP was generated by in-fusion cloning the full length raptor ORF from mDsRed-raptor into the pOPINE-3C-YFP vector, using the primers below:

   raptor-YFP Fwd AGGAGATATACCATGGAGTCCGAAAT GCTGCAATCG.

   raptor-YFP Rev CAGAACTTCCAGTTTTCTGACACGCT TCTCCACCG.

3. Plasmid constructs were verified by reverse PCR screens and further validated by Sanger sequencing using T7F primers and appropriate reverse primers at Source Bioscience (UK).

3.6 Cell Lines

HEK293 cells were cultured in Minimum Essential Media (MEM) supplemented with 10% (w/v) fetal bovine serum (FBS), 2 mM L-glutamine, and 1% (w/v) penicillin–streptomycin. All culture reagents were acquired from Life Technologies. Cells were incubated at 37 °C with 5% CO_2 humidified air in T75 culture flasks (Thermo Scientific).

3.7 Cell Transfection

HEK293 was seeded for 24 h at a density of 1×10^5 or 1.5×10^5 cells per ml on uncoated 35 mm no 1.5 glass bottom dishes (MatTek). Cells were transfected with 0.5 μg of plasmid DNA using FuGENE HD (Promega) transfection reagent.

3.8 Equipment for Real-Time Imaging of Protein–Protein Interactions Using a Confocal and FRET-FLIM Setup

Our two-channel confocal and two-channel FLIM setup was fitted with time-correlated single-photon counting cards (TCSPC) from Becker & Hickl (BH SPC 830 or SPC150), and data were analyzed using the BH SPCImage analysis software v 5.1.

1. Although the two-photon microscopy technique has several advantages over a one-photon system, it is best to use a one-photon confocal microscope for the multicolor FRET-FLIM technique. This is because the ultrafast femtosecond pulsed laser is likely to excite both the donor(s) and acceptor (s) at the same time. This needs to be avoided. For the PIE operation which is essential for three protein interaction studies, two separate pulsed laser sources are required. It is more convenient to use two lasers that (1) can vary their repetition rates, (2) can be synchronized, and (3) can be triggered externally (Fig. 3).

2. We have used a BH BDL-SMC-405 pulsed laser with 40 ps pulse-width synchronized with the NKT supercontinuum laser operating at 80 MHz. The two lasers were combined using a fiber combiner (OzOptics, Canada). The supercontinuum laser operated at the full 80 MHz as the "master" and was used to trigger the BDL-SMC-405 in a "slave" mode with a 10-ns delay. It is recommended to run the lasers at 40 MHz, i.e., 25 ns between pulses to allow a full decay of the fluorescent protein excited state before the next excitation pulse arrives. Fluorescence emission was filtered using a 460/60 nm band-pass filter for CFP or mTurquoise2, or 515/30 nm filter for GFP (Thorlabs) and focused onto the two hybrid detectors.
3. The laser beams are focused to a diffraction-limited spot using a high numerical aperture oil or water-immersion objective (e.g., Nikon x60 VC NA 1.2) to illuminate specimens on the microscope stage. Fluorescence emission is collected by the same objective in a standard confocal mode. Line, frame, and pixel clock signals are generated and synchronized with two external detectors in the form of two fast hybrid photomultiplier tubes (PMT) or microchannel plate PMTs (i.e., Hamamatsu R3809U or HPM-100). The raw FLIM data are generated by linking these PMTs via two TCSPC PC modules. We used BH SPC830 or SPC150 cards. A single TCSPC may be used together with routing electronics, but this would lead to only half of the intensity from each channel recorded.
4. Prior to FLIM data collection, expression levels of the fluorescent protein-tagged proteins of interest are verified by imaging transfected cells using a confocal microscope. Here, an inverted Nikon TE2000-U or Ti-E confocal microscope attached to a Nikon C1 or C2 scanning unit was used with filter sets for GFP (488 nm excitation) or mDsRed/mCherry/Alexa 555 (543 or 561 nm excitation) and mTurqouise2/Alexa405 (405 nm excitation). A 633-nm interference filter is used in the red channel to significantly minimize the bleed-through of any GFP or autofluorescence emission that would otherwise obscure the mCherry emission. Ideally, equal expression levels for all proteins under investigation are selected to obtain optimal FRET conditions (Fig. 4). Alternatively, higher levels of the acceptor protein are desirable.
5. Prior to FLIM image data acquisition, the system performance needs to be checked and verified. Standard dyes with well-characterized excitation and emission spectra and excited-state lifetimes are commonly used (Table 5). We have routinely used aqueous solutions (~50 μM) of 7-hydroxy-coumarin carboxylic acid (7-OH-CCA, 360–405 nm excitation, 420 nm

emission), fluorescein at pH 7–10 (488 nm excitation, 515 nm emission), and rhodamine B (543–561 nm excitation, 590 nm emission).

6. FLIM images were acquired at 256 × 256 pixels, although megapixel sizes (2048 × 2048) can now be achieved due to recent advances and improvements in the FLIM acquisition PC cards and software compared to previous 128 × 128 pixels (Fig. 4). Supercontinuum laser coupled with AOTF may be used to select any excitation and emission wavelengths for both the confocal and one-photon FLIM operations.

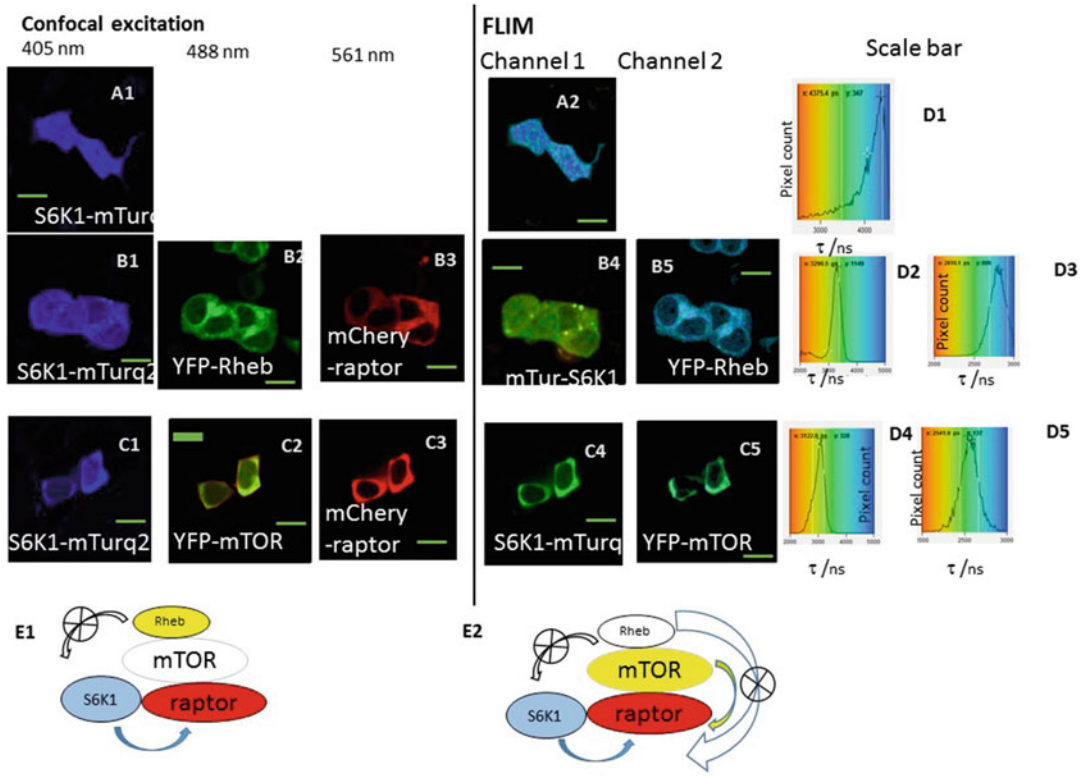


Fig. 4 Live HEK293 cell localization and interaction between the mTORC1-expressed proteins observed by three-channel confocal and PIE two-channel FLIM. A1: EGPF-S6K1 alone as control with observed lifetime. B1: The distribution of all lifetime in the image pixel is given in D1. B1–3 Confocal image of cell expression of mTurg2-S6K1, FYP-Rheb, and mCherry-raptor. B4 and B5 are the two-channel PIE FLIM images of mTurg2–S6K1 and YFP-Rheb. D2 and D3 are the respective lifetime distributions, indicating a reduction in lifetime of the mTurg2-S6K1 but not the YFP-Rheb. C1–3 Confocal image of cell expression of mTurg2-S6K1, FYP-mTOR, and mCherry-raptor. C4 and C5 are the two-channel PIE FLIM images of mTurg2-S6K1 and YFP-mTOR. D4 and D5 are the respective lifetime distributions, indicating a reduction in lifetime of the mTurg2-S6K1 but not the YFP-mTOR. E1 and E2 are schematic representation of the physical interaction between the different proteins within the complex. FLIM of the red channel is not measured but could be performed in a three-channel PIE setup. Scale bar is 10 μm

Table 5
Standard solution lifetime calibration for three-colour FRET-FLIM

Standard dye	%	Lifetime	Excitation (in nm)	Energy transfer
7-OH-CCA	100	3.7 ± 0.1 ns	405	N/A
Fluorescein	100	4.0 ± 0.1 ns	488	N/A
Rhodamine b	100	1.7 ± 0.05 ns	543 or 561	N/A
7-OH-CCA + fluorescein	50/50	3.3 ± 0.05 ns	405	Yes
Fluorescein + rhodamine b	50/50	3.1 ± 0.05 ns	488	Yes
7-OH-CCA + fluorescein + rhodamine b	33/33/33	3.3 ± 0.05 ns (Ch1), 3.1 ± 0.05 ns (Ch2)	404	Yes

Raw FLIM data are analyzed by obtaining excited-state lifetime values first on a pixel-by-pixel basis, then of a region of interest, and calculations are made using SPCImage analysis software.

7. The distribution of lifetime values within the region of interest is generated and displayed as a curve. Only chi-squared (χ^2) values between 0.9 and 1.3 are considered a good fit of the data points. For a pure sample such as the test solutions and donor FPs (mTurq2 or YFP), it is best to fit the data to a single-exponential decay. The median, minimum, and maximum values of the lifetime values from the curve are taken to generate the range of lifetimes per sample. At least three areas of a sample dish and at least three independent biological samples per protein–protein combination are analyzed and the average of the ranges is taken. The percentage efficiency of energy transfer (E) from one protein to the other may be determined using Eqs. (1) and (2):

$$E\% = \left[1 - \left(\frac{\tau_{DA}}{\tau_D}\right)\right] \times 100 \quad (5)$$

where τ_D and τ_{DA} are the measured excited-state lifetime of the donor and acceptor, respectively.

3.9 Typical Results

Typical multichannel FLIM results of multiple protein interactions are shown in Fig. 4 and changes in lifetime values are given in Table 6.

Table 6
Three-color FRET-FLIM results for multiple mTORC1 interactions

mTORC1 proteins	Lifetime (ns)	Comment
S6K1-mTurq2	4.3 ± 0.1	Control lifetime
YFP-Rheb	2.8 ± 0.1	Control lifetime
YFP-mTOR	~2.8 ± 0.1	Control lifetime
mCherry-raptor	N/A	τ not measured
S6K1-mTurq2 + mCherry-raptor	3.3 ± 0.1	Interaction
YFP-Rheb + mCherry-raptor	2.8 ± 0.1	No interaction
YFP-mTOR + mCherry-raptor	2.5 ± 0.1	Interaction
S6K1-mTurq2 + YFP-Rheb + mCherry-raptor	3.2 (ch 1) + 2.8 (ch 2) ± 0.1	Interaction between S6K1 and raptor, but no interaction between S6K1 and Rheb or raptor and Rheb. No direct three-way interaction between mTOR, raptor, and S6K1
S6K1-mTurq2 + YFP-mTOR + mCherry-raptor	3.1 (ch1) + 2.5 (ch 2) ± 0.1	Strong interaction between S6K1 and raptor strong interaction between raptor and mTOR. Possible slight direct three-way interaction between mTOR, raptor, and S6K1

4 Notes

Full personal protective equipment including gloves has to be worn for the whole procedure to reduce contamination throughout the preparation. Sterile conditions are required particularly for the cell culture.

Acknowledgments

This work was supported by a BBSRC iCASE PhD studentship (BB/L016052/1) to AA. We thank STFC for funding access to the Central Laser Facility. This work also acknowledges Professor Raymond Owens (PP-UK) for providing support and plasmid vectors for the cloning work.

References

1. Beretta L, Gingras AC, Svitkin YV et al (1996) Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. EMBO J 15(3):658–664
2. Brunn GJ, Hudson CC, Sekulić A et al (1997) Phosphorylation of the translational repressor PHAS-I by the mammalian target of rapamycin. Science 277(5322):99–101

3. Gingras AC, Kennedy SG, O'Leary MA et al (1998) 4E-BP1, a repressor of mRNA translation, is phosphorylated and inactivated by the Akt(PKB) signaling pathway. Genes Dev 12 (4):502–513
4. Holz MK, Ballif BA, Gygi SP et al (2005) mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. Cell 123(4):569–580
5. Arif A, Terenzi F, Potdar AA et al (2017) EPRS is a critical mTORC1–S6K1 effector that influences adiposity in mice. Nature 542:357–361
6. Caccamo A, Branca C, Talboom JS et al (2015) Reducing ribosomal protein S6 kinase 1 expression improves spatial memory and synaptic plasticity in a mouse model of Alzheimer 's disease. J Neurosci 35:14042–14056
7. Kim SY, Baik KH, Baek KH et al (2014) S6K1 negatively regulates TAK1 activity in the toll-like receptor. Mol Cell Biol 34:510–521
8. Marabita M, Baraldo M, Solagna F et al (2016) S6K1 is required for increasing skeletal muscle force during hypertrophy article S6K1 is required for increasing skeletal muscle force during hypertrophy. Cell Rep 17:501–513
9. Choo AY, Yoon SO, Kim SG et al (2008) Rapamycin differentially inhibits S6Ks and 4E-BP1 to mediate cell-type-specific repression of mRNA translation. Proc Natl Acad Sci U S A 105(45):17414–17419. https://doi.org/10.1073/pnas.0809136105
10. Choo AY, Blenis J (2009) Not all substrates are treated equally implications for mTOR, rapamycin-resistance and cancer therapy. Cell Cycle 8:567–572. https://doi.org/10.4161/cc.8.4.7659
11. Förster T (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz [Intermolecular energy migration and fluorescence]. Annalen der Physik (in German). 437: 55–75. B För s t e r T (1946) Energiewanderung und Fluoreszenz. Naturwissenschaften 33:166–175? (Translation: by Klaus Suhling). J Biomed Opt 17:011002
12. Pawley JB (2006) Handbook of biological confocal microscopy, 3rd edn. Springer, New York, NY, pp 1–985. https://doi.org/10.1007/978-0-387-45524-2_8. ISBN 978-0-387-25921-5
13. Fábián ÁI, Rente T, Szöllősi J et al (2010) Strength in numbers: effects of acceptor abundance on FRET efficiency. ChemPhysChem 11:3713–3721. https://doi.org/10.1002/cphc.201000568
14. Maliwal BP, Raut S, Fudala R et al (2012) Extending Förster resonance energy transfer measurements beyond 100 Å using common organic fluorophores: enhanced transfer in the presence of multiple acceptors. J Biomed Opt 17:11006. https://doi.org/10.1117/1.JBO.17.1.011006
15. Walczewska-Szewc K, Bojarski P, d'Auria S (2013) Extending the range of FRET—the Monte Carlo study of the antenna effect. J Mol Model 19:4195–4201. https://doi.org/10.1007/s00894-013-1810-3
16. Koushik SV, Blank PS, Vogel SS (2009) Anomalous surplus energy transfer observed with multiple FRET acceptors. PLoS One 4(11): e8031. https://doi.org/10.1371/journal.pone.0008031
17. Bunt G, Wouters FS (2017) FRET from single to multiplexed signaling events. Biophys Rev 9:119–129. https://doi.org/10.1007/s12551-017-0252-z
18. Wolfgang B (2015) Advanced single photon counting applications. Springer, New York, pp 1–624. https://doi.org/10.1007/978-3-319-14929-5
19. Botchway SW, Scherer KM, Hook S et al (2015) A series of flexible design adaptations to the Nikon E-C1 and E-C2 confocal microscope systems for UV, multiphoton and FLIM imaging. J Microsc 258(1):68–78. https://doi.org/10.1111/jmi.12218
20. Scully AD, Macrobert AJ, Botchway S et al (1996) Development of a laser-based fluorescence microscope with subnanosecond time resolution. J Fluoresc 6(2):119–125. https://doi.org/10.1007/BF00732051
21. Suhling K, Siegel J, Phillips D et al (2002) Imaging the environment of green fluorescent protein. Biophys J 83(3589–3595):2002
22. Marcu L, French PMW, Elson DS (2014) Fluorescence lifetime spectroscopy and imaging—principles and applications in biomedical diagnostics. CRC Press, Boca Raton, Florida, pp 1–570. SBN 9781439861677—CAT# K12851
23. Estandarte AK, Botchway S, Lynch C et al (2016) The use of DAPI fluorescence lifetime imaging for investigating chromatin condensation in human chromosomes. Sci Rep 6:31417. https://doi.org/10.1038/srep31417
24. Yadav RB, Burgos P, Parker AW et al (2013) mTOR direct interactions with Rheb-GTPase and raptor: sub-cellular localization using fluorescence lifetime imaging. BMC Cell Biol 14:3. https://doi.org/10.1186/1471-2121-14-3
25. Stubbs CD, Botchway SW, Slater SJ et al (2005) The use of time-resolved fluorescence imaging in the study of protein kinase C localisation in cells. BMC Cell Biol 6(1):22

26. Jeshtadi A, Burgos P, Stubbs CD et al (2010) Interaction of poxvirus intracellular mature virion proteins with the TPR domain of kinesin light chain in live infected cells revealed by two-photon-induced fluorescence resonance energy transfer fluorescence lifetime imaging microscopy. J Virol 84(24):12886–12894. https://doi.org/10.1128/JVI.01395-10
27. Osterrieder A, Carvalho CM, Latijnhouwers M et al (2009) Fluorescence lifetime imaging of interactions between Golgi tethering factors and small GTPases in plants. Traffic 10 (8):1034–1046. https://doi.org/10.1111/j.1600-0854.2009.00930.x
28. Nettleship JE, Assenberg R, Diprose JM et al (2010) Recent advances in the production of proteins in insect and mammalian cells for structural biology. J Struct Biol 172 (1):55–65. https://doi.org/10.1016/j.jsb.2010.02.006

Chapter 17

Context-Specific and Proximity-Dependent Labeling for the Proteomic Analysis of Spatiotemporally Defined Protein Complexes with Split-BioID

Cinthia Amaya Ramirez, Stefanie Egetemaier, and Julien Béthune

Abstract

Proximity-dependent labeling techniques such as BioID and APEX2 allow the biotinylation of proteins proximal to a protein of interest in living cells. Following streptavidin pulldown and mass spectrometry analysis, this enables the identification of native protein–protein interactions. Here we describe split-BioID, a protein-fragment complementation assay that increases the resolution of BioID. Using this technique, context-specific protein complexes can be resolved.

Key words BioID, Protein-fragment complementation assay, Protein–protein interactions, Biotin, Proteomics

1 Introduction

In all living organisms, thousands of proteins mediate most cellular functions. To do so, proteins usually do not act alone. Rather, they assemble with other proteins to build dynamic macromolecular complexes that can remodel according to the exact functions that need to be exerted. The protein–protein interactions (PPI) involved in such complexes are often deregulated in disease and are thus promising targets for therapeutics [1]. Identifying and characterizing PPI networks are hence of prime importance when trying to understand how a protein of interest (POI) works. To complement classical pulldown approaches in which the POI is isolated by affinity purification (AP) and co-purifying proteins identified by mass spectrometry (MS) analysis, proximity-dependent biotinylation techniques were recently introduced that allow the labeling of proteins vicinal to the POI in living cells. Two such techniques are currently available: APEX2- and BioID-mediated labeling [2, 3]. The former relies on an engineered peroxidase that activates a biotin-phenol substrate that is

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5_17,

fed to the cells, whereas the latter uses an abortive variant of a protein biotin ligase (BPL) that activates native biotin. In both the cases, the activated substrate diffuses around the enzyme and leads to the biotinylation of proximal proteins in an estimated range of approximately 10 nm for BioID [4]. When fused to a POI, both enzymes thus mediate the biotinylation of nearby factors including interacting proteins. These biotinylated proteins can then be efficiently isolated by streptavidin pulldown and identified by MS. As opposed to AP-MS, APEX2- and BioID-MS do not aim at purifying assembled protein complexes by pulling one of its components but rather identify proteins that were marked within cells because they were in close proximity to the POI. In proximity-dependent labeling approaches, it thus does not matter whether interacting proteins are still associated during the pulldown procedure, making these techniques more powerful than AP-MS for detecting transient interactions [5] or PPI that depend on intact or poorly soluble cellular structures [6]. Yet, a limitation to both approaches is that they cannot easily resolve the remodeling and/or maturation of protein complexes. In the common situation in which a POI is part of several distinct complexes depending on the cellular context, both AP- and APEX2/BioID-MS will identify all possible PPI but will not assign individual interactions to specific context-dependent complexes. We and others have recently introduced split-BioID assays [7, 8] in which BirA*, the BPL of BioID, is split into two inactive and poorly interacting fragments that can reassemble into an active enzyme when fused with two interacting proteins. Provided a POI is known to interact with a given factor within a specific protein complex, applying split-BioID to these two proteins allows the labeling of additional components of that specific assembly, ignoring any additional PPI the POI may have within other protein complexes (Fig. 1). The general strategy of a split-BioID experiment is outlined in Fig. 2. Briefly, the coding sequences for two putative interacting POI are cloned in frame with the two BirA* fragments, termed NBirA* and CBirA*, and transfected in mammalian cells. Biotinylation of proximal proteins is induced by the addition of excess biotin in the growth medium. Biotinylated proteins are then isolated by streptavidin pulldown and can be analyzed either by immunoblotting or MS. Of note the protocol described here is virtually identical to a classical BioID procedure, the only difference being that in BioID the POI is fused to the full BirA* enzyme.

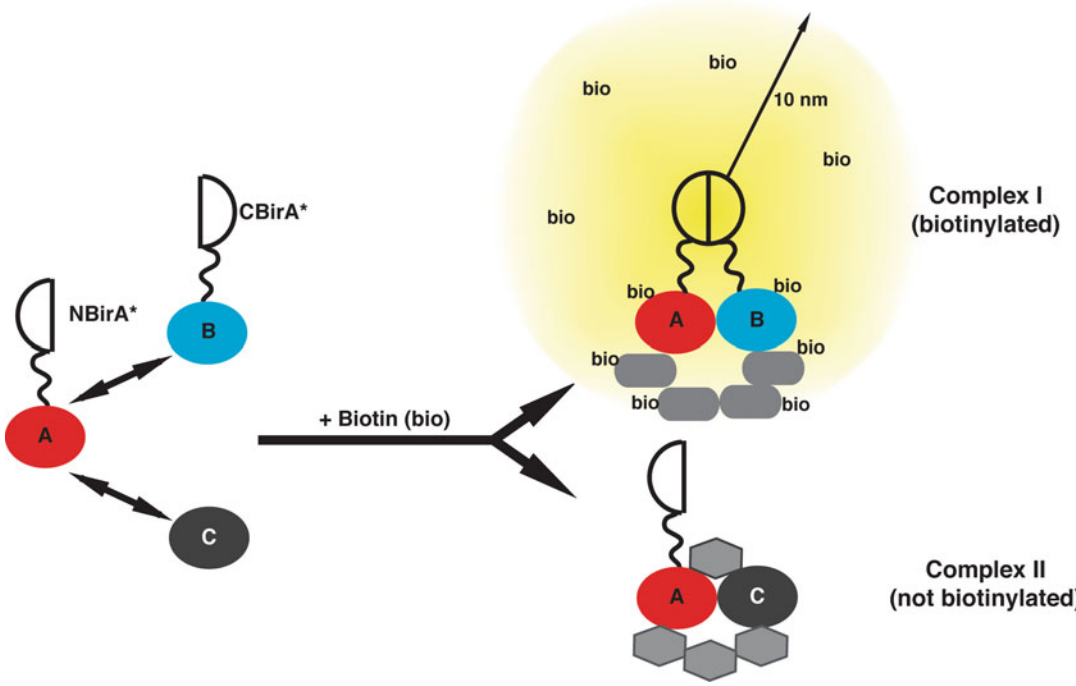


Fig. 1 Principle of split-BioID. Protein A interacts with protein B within Complex I or with protein C within Complex II. Fusing the BirA* fragments NBirA* and CBirA* to protein A and protein B, respectively, leads to the reassembly of an active enzyme upon dimerization of these two proteins and the specific labeling of further components of complex

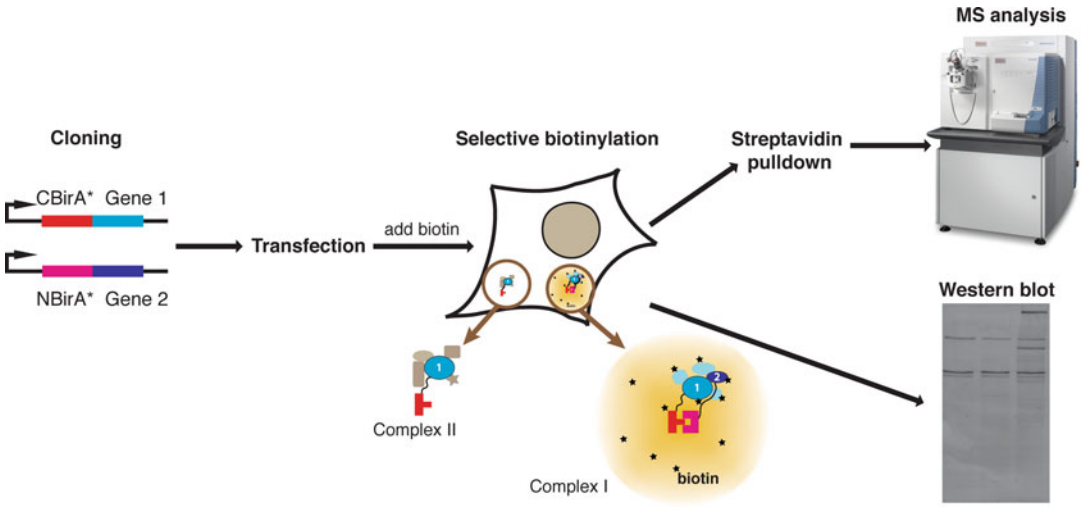


Fig. 2 Outline of the main steps in a typical split-BioID experiment

2 Materials

Prepare all solutions using ultrapure water (such as mQ water from Millipore or equivalent) and analytical grade reagents. When preparing stock solutions and reagents, care should be taken to avoid keratin contamination (essentially: always use gloves, single-use plastic ware, and do not touch yourself).

2.1 Preparation of Plasmids

1. DNA restriction enzymes.
2. T4 DNA ligase or DNA ligation kit.
3. Polymerase chain reaction (PCR) reagents.
4. Miniprep/Midiprep kit for plasmid DNA purification.
5. Transformation competent *E. coli* cells.

2.2 Transfection of Mammalian Cells

1. Cell line of interest (*see* **Note 1**).
2. Cell culture medium (*see* **Note 2**).
3. Transfection reagent (*see* **Note 3**).

2.3 Proximity-Dependent Biotinylation

1. Biotin stock solution: 50 mM in neutralized ammonium hydroxide (*see* **Note 4**). Store at −20 °C.
2. Doxycycline stock solution: 10 mg mL^{-1} in 70% ethanol (*see* **Note 5**). Store at −20 °C.

2.4 Streptavidin Pulldown

1. Streptavidin-coupled magnetic beads (*see* **Note 6**). Store at 4 °C.
2. Magnet for 1.5-mL tubes.
3. Lysis buffer II: 50 mM Tris–HCl, 500 mM NaCl, 5 mM EDTA, 0.4% SDS, 1 mM DTT, protease inhibitor cocktail, pH adjusted to 7.4. Prepare freshly (*see* **Note 7**).
4. Equilibration buffer: 50 mM Tris–HCl, 150 mM NaCl, 0.05% Triton X-100, 1 mM DTT, adjusted to pH 7.4.
5. Wash buffer 1: 2% SDS in water. Prepare freshly (*see* **Note 7**).
6. Wash buffer 2: 50 mM Hepes, 1 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% sodium-deoxycholate, pH adjusted to 7.4. Prepare freshly (*see* **Note 7**).
7. Wash buffer 3: 10 mM Tris–HCl, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% sodium-deoxycholate, pH adjusted to 8. Prepare freshly (*see* **Note 7**).
8. Wash buffer 4: 50 mM Tris–HCl, 50 mM NaCl, 0.1% NP-40, pH adjusted to 7.4. Prepare freshly (*see* **Note 7**).
9. Elution buffer: 10 mM Tris–HCl, 2% SDS, 5% β-mercaptoethanol, 2 mM biotin, pH adjusted to 7.4. Store at −20 °C.

10. Bradford reagent.
11. 25G needles.
12. 1-mL syringes.
13. Rotating wheel for 1.5-mL tubes.

2.5 SDS-Polyacrylamide Gel Electrophoresis and Western Blotting

1. Lysis buffer I: 50 mM Tris–HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 0.5% NP-40, 0.5 mM DTT, protease inhibitor cocktail.
2. Stacking gel buffer: 1 M Tris–HCl, pH 6.8.
3. Resolving gel buffer: 1.5 M Tris–HCl, pH 8.8.
4. Thirty percent acrylamide/bisacrylamide solution (37.5:1).
5. Running buffer: 25 mM Tris–HCl, 192 mM glycine, 0.1% SDS.
6. Sample loading dye (5×): 250 mM Tris–HCl, pH 6.8, 5% β-mercaptoethanol, 0.02% bromophenol blue, 10% SDS, 30% glycerol.
7. Low-fluorescence PVDF membranes (*see* **Note 8**).
8. Trans-Blot Turbo Transfer System and reagents (*see* **Note 9**).
9. Infrared dye-conjugated streptavidin (*see* **Note 8**). Store at 4 °C.
10. Precast protein gels and reagents (*see* **Note 10**).
11. Coomassie staining solution: 0.02% (w/v) Coomassie blue stain G-250, 5% (w/v) aluminum sulfate-(14–18)-hydrate, 10% (v/v) ethanol (96%), 2% (v/v) orthophosphoric acid (85%). Store at room temperature.
12. Destaining solution: 10% (v/v) ethanol (96%), 2% (v/v) orthophosphoric acid (85%).
13. Sterile 15-cm bacterial dishes.
14. Sterile scalpels.

3 Methods

3.1 Cloning of Two Genes of Interests in Split-BioID Plasmid

1. Design primers to amplify the two genes to be tested, so that restriction sites are included that allow subcloning in split-BioID plasmids (*see* **Note 11**).
2. Perform PCR to amplify both genes, run the reactions on a 1% agarose-TAE gel and cut slices of agarose gel that encompass the amplified DNA. Extract and purify the amplified DNA using a standard DNA gel extraction kit.
3. Digest amplified DNA and acceptor plasmid(s) with the relevant restriction enzymes. Typically 2 μg of plasmid is mixed with 1 μL of each restriction enzyme in a total volume reaction of 20 μL, whereas 25 μL of the purified PCR products is mixed

with 1 μL of each restriction enzyme in a total volume reaction of 30 μL. DNA digestion with restriction enzymes is typically performed at 37 °C for at least 1 h (*see* **Note 12**).

4. Run the plasmid restriction reactions on a 1% agarose-TAE gel and cut slices of agarose gel around the digested plasmid backbone bands. Extract and purify the digested DNA using a standard DNA gel extraction kit. Digested PCR products can be directly purified using a PCR or reaction clean up protocol usually included in gel extraction kits.
5. Ligate digested plasmids and PCR products using T4 DNA ligase. Typically, a total of 50–100 ng DNA (plasmid + insert with a three- to fivefold molar excess insert/plasmid) is mixed with ligation buffer and DNA ligase in a total of 10 μL and incubated at room temperature for 5–10 min.
6. Use 3 μL of the ligation reaction to transform 50 μL of competent *E. coli* cells. Typically bacteria and DNA are gently mixed and incubated on ice for 30 min. Transformation is then induced by heat shock at 42 °C for 30–45 s. The cells are then placed back on ice. After 2 min, 250 μL of pre-warmed (37 °C) LB medium is added and the cells incubated on a shaking block at 37 °C for 1 h at 800 rpm (*see* **Note 13**). Thereafter, 100 μL of the incubation is spread on a 10 cm bacterial LB-agar plate containing the appropriate antibiotic selection. The plates are incubated overnight at 37 °C.
7. On the next day, two to four antibiotic-resistant colonies are picked to inoculate 3 mL of LB medium containing the appropriate antibiotic in 15 mL tubes. The tubes are then incubated for 12–16 h at 37 °C, 180 rpm in a shaker-incubator.
8. On the next day, plasmids are isolated by Miniprep and their sequences verified by Sanger sequencing analysis.
9. Correct plasmids can be re-amplified and purified with a Midiprep protocol to generate a larger stock of sequence-verified split-BioID plasmids.

3.2 Small-Scale Testing of Proximity-Dependent Biotinylation in Living Cell

1. Seed a six-well cell culture plate with the cells to transfect so that they reach 50–70% confluence on the following day. Incubate overnight at 37 °C, 5% CO_2. For HeLa cells, typically 80–150 × 10^3 cells are seeded per well in a volume of 2 mL (*see* **Note 14**).
2. On the next day, remove the medium from the six-well plate and replace with 2 mL of fresh medium per well.
3. Prepare the transfection mixes for all conditions to be tested, usually the pair of proteins of interest and a negative control (*see* **Note 15**). When using PEI as a transfection reagent for HeLa cells, we typically use 6 μg PEI diluted in 100 μL DMEM

(without serum) mixed with 3 μg plasmid diluted in 100 μL DMEM (without serum) for each well to transfect. The transfection mixes (200 μL per well) are incubated for 5–15 min at room temperature (RT) and then added drop wise to their corresponding wells. The cells are then incubated overnight at 37 °C, 5% CO_2.

4. On the next day, remove the medium from each well and replace by 2 mL of DMEM containing 50 μM biotin and 200 ng mL^{-1} doxycycline if using inducible plasmids under the control of a tet-responsive element (*see* **Note 16**). Incubate the cells for at least 16 h at 37 °C, 5% CO_2.
5. On the following day, wash cells in each well with 1 mL ice-cold PBS and then scrap the cells in 100 μL of lysis buffer I with disposable cell scrappers.
6. Transfer each sample of scrapped cells to 1.5 mL tubes, resuspend briefly by pipetting and centrifuge at 14,000 × *g* for 10 min at 4 °C to clear the lysates.
7. Transfer the supernatants to fresh 1.5-mL tubes and place on ice.
8. Determine the protein concentration of each sample using a standard Bradford assay.
9. Prepare samples for SDS-PAGE by mixing an equal amount (15–30 μg protein content) of each lysate with 5× SDS-loading buffer to yield a total of 30 μL SDS-PAGE sample in 1× SDS-loading buffer (volume are adjusted with lysis buffer I). Denature the samples at 95 °C for 5 min (*see* **Note 17**).
10. Load the samples (20 μL/lane) and molecular weight markers on an SDS-polyacrylamide gel with a percentage allowing the appropriate resolution of the different fusion proteins.
11. After electrophoresis, transfer the fractionated proteins to an immunoblotting membrane (*see* **Note 9**).
12. Following protein transfer, block the membrane with 5% dry milk in PBS for 30 min at RT.
13. To detect biotinylated proteins, incubate the membrane in a 50-mL falcon tube with 5 mL of conjugated streptavidin diluted 1:15,000 in PBS containing 2% bovine serum albumin (BSA) and 0.1% Tween-20 for 30 min at RT on a roller-shaker.
14. Wash the membrane three times with PBS containing 0.1% Tween-20 for 10 min at RT (*see* **Note 18**) and then proceed with detection of chemoluminescence or fluorescence signals (Fig. 3).
15. If detection of the fusion proteins is desired, incubate the membrane in a 50-mL falcon tube with 5 mL of anti-Myc (to detect the fusion to NBirA*) and anti-FLAG (to detect

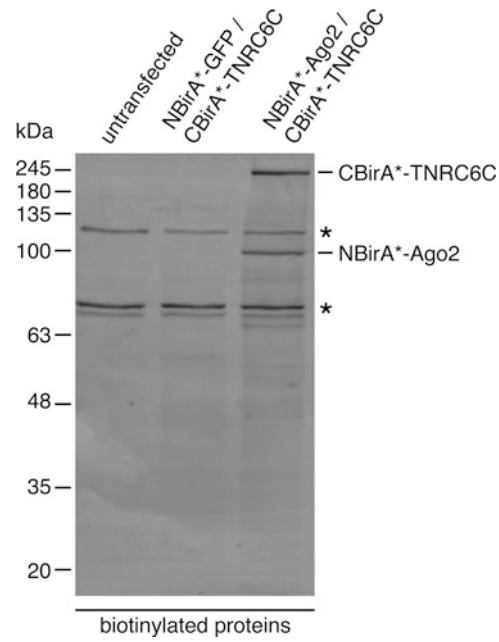


Fig. 3 Typical Western blot for a small-scale split-BioID validation experiment. Biotinylated proteins were detected with conjugated streptavidin. Endogenous biotinylated proteins are detected in non-transfected cells. A similar pattern is observed for the GFP/TNRC6C negative control. Additional biotinylated proteins (mainly the fusion proteins themselves) are detected when split-BioID is applied to the Ago2 and TNRC6C, which are known to directly interact

the fusion to CBirA*) diluted at their optimal concentrations (*see* **Note 19**) in PBS containing 2% bovine serum albumin (BSA) and 0.1% Tween-20 for 1 h at RT on a roller-shaker (*see* **Note 20**).

16. Wash the membrane three times with PBS containing 0.1% Tween-20 for 10 min at RT.
17. Incubate the membrane in a 50-mL falcon tube with 5 mL of conjugated secondary antibodies directed against the anti-Myc and -FLAG primary antibodies diluted at their optimal concentrations (*see* **Note 21**) in PBS containing 2% bovine serum albumin (BSA) and 0.1% Tween-20 for 30 min at RT on a roller-shaker.
18. Wash the membrane three times with PBS containing 0.1% Tween-20 for 10 min at RT (*see* **Note 18**) and then proceed with detection of chemoluminescence or fluorescence signals.

3.3 Large-Scale Proximity-Dependent Biotinylation for Proteomics Studies

1. For each condition to be tested, seed three to four 10-cm cell culture dishes with the cells to transfect, so that they reach approximately 70% confluence on the following day. Incubate overnight at 37 °C, 5% CO_2. For HeLa cells, typically 8×10^5 cells are seeded per dish in a volume of 10 mL (*see* **Note 14**).

2. On the next day, remove the medium from the dishes and replace with 10 mL of fresh medium.
3. Prepare the transfection mixes for all conditions to be tested, usually the pair of proteins of interest and a negative control (*see* **Note 15**). When using PEI as a transfection reagent for HeLa cells, we typically use 12 μg PEI diluted in 150 μL DMEM medium (without serum) mixed with 6 μg plasmid diluted in 150 μL DMEM medium (without serum) per plate. The transfection mixes are incubated for 5–15 min at room temperature (RT) and then dispatched drop wise to their corresponding dishes (300 μL per 10 cm plate). The cells are then incubated overnight at 37 °C, 5% CO_2.
4. On the next day, transfer the cells from each dish to 15 cm plates. To do so, the cells are first washed with 7 mL PBS and then incubated with 1.5 mL of a trypsin-EDTA solution for 5 min at RT. DMEM (3.5 mL) is then added and the cells resuspended by pipetting and transferred to 15 cm plates filled with 20 mL DMEM containing enough biotin to reach a final concentration of 50 μM for a total volume of 25 mL, and enough doxycycline for a concentration of 200 ng mL^{-1} doxycycline if using inducible plasmids under the control of a tet-responsive element (*see* **Note 16**). Incubate the cells for at least 16 h at 37 °C, 5% CO_2.
5. On the following day, cells in each 15 cm dish are washed twice with 20 mL PBS and then scrapped in 1.5 mL PBS using disposable cell scrappers.
6. Scrapped cells are then transferred to 1.5 mL tubes and harvested by centrifugation at 1200 × *g* for 5 min at 4 °C.
7. Cell pellets can then be snap frozen in liquid nitrogen and if desired stored at −80 °C until further processing.
8. To prepare cell lysates, the frozen cell pellets are first resuspended in 1 mL lysis buffer II at room temperature and passed ten times (five strokes) through 25G needle attached to a 1-mL syringe (*see* **Note 21**).
9. The lysates are then sonicated in a cold ultrasonic bath for 4 cycles of 30 s at high intensity with 30 s pause time between each cycle (*see* **Note 22**).
10. Add Triton X-100 to the sonicated lysates to a final concentration of 2% (Typically 100 μL of a 20% Triton X-100 solution is added to 900 μL of lysate) and then enough volume of 50 mM Tris–HCl, pH 7.4 solution, so that the final NaCl concentration is adjusted to 150 mM (typically 2.3 mL is added to 1 mL of Triton X-100 adjusted lysates) (*see* **Note 23**).

11. Distribute the adjusted lysates (total volume approximately 3.3 mL) in three 1.5-mL tubes and clear by centrifugation at 16,000 × *g* for 10 min at 4 °C.
12. Pool the supernatants from each condition in 15-mL tubes.
13. Determine the protein concentration of each sample using a standard Bradford assay. Adjust volumes of each sample, so that they all have equal concentration and keep 10% aside as Input sample for later analysis. Typically 3–3.5 mg of protein content is used for a streptavidin pulldown.

3.4 Streptavidin Pulldown for Proteomics Studies

1. Pre-equilibrate the total amount of streptavidin-coupled beads, typically 200 μL per pulldown (*see* **Note 24**) by transferring the beads to a 1.5-mL tube and wash them once with 1 mL equilibration buffer.
2. Dispatch the equilibrated beads in enough 1.5-mL tubes to accommodate the starting amount of 3–3.5 mg protein sample per conditions (usually two to four tubes are needed).
3. Place the tubes on a magnetic rack, remove the supernatant, and for each condition, resuspend the beads with equal amounts of cleared lysates.
4. Incubate the bead/lysate mixture overnight at 4 °C on a rotating wheel.
5. On the next day, place the tubes on a magnetic rack, wait until the beads have stuck to the side of the tubes and gently aspirate the supernatants using a capillary tip connected to a vacuum pump. From now on, the following steps are performed at RT.
6. In each tube, resuspend the beads with 200 μL of wash buffer 1 and pool beads from the same conditions in a single 1.5-mL tube.
7. Incubate for 8 min on a rotating wheel.
8. Place the tubes on a magnetic rack, wait until the beads have stuck to the side of the tubes and gently aspirate the supernatants using a capillary tip connected to a vacuum pump or manually with a pipette.
9. Wash the beads once more with 1 mL of wash buffer 1 as in **steps 7** and **8**.
10. Following the same procedure as in **step 9**, wash the beads twice with 1 mL of wash buffer 2, then wash buffer 3, then transfer to fresh 1.5-mL tubes and proceed with wash buffer 4.
11. After the final wash, briefly centrifuge the samples on a table micro-centrifuge, place back on the magnetic rack, and remove any remaining buffer with a pipette.
12. Elute captured biotinylated proteins (*see* **Note 25**) by adding 30 μL of elution buffer to each sample. Incubate at 98 °C,

800 rpm for 15 min on a thermomixer, then immediately remove the beads on the magnetic rack and transfer the eluted material in a fresh 1.5-mL tube.

13. Eluted samples can be stored at −20 °C until further processing.

3.5 SDS-PAGE and Sample Preparation for MS Analysis

1. Prior to sample preparation for MS analysis, assess the success of the split-BioID experiment by Western blot. To do so, prepare input and elution blot samples as follows: equal protein amounts of input samples mixed with 5× SDS loading buffer in a total volume of 28 µL (load 25 µL per lane), 5 µL of each elution sample mixed with 1.25 µL of 5× SDS loading buffer (load 6 µL per lane). Western blot analysis is performed as described in Subheading 3.2, **steps 10–18** (Fig. 4a).
2. To prepare sample for MS analysis, we recommend using commercial precast gels and reagents (*see* **Note 10**). Prepare samples for SDS-PAGE as follows: 18.75 µL elution samples mixed with 4× sample buffer. Load the samples on a 4–20% SDS-polyacrylamide gel and run the electrophoresis until the running front migrate 2–4 cm into the gel.

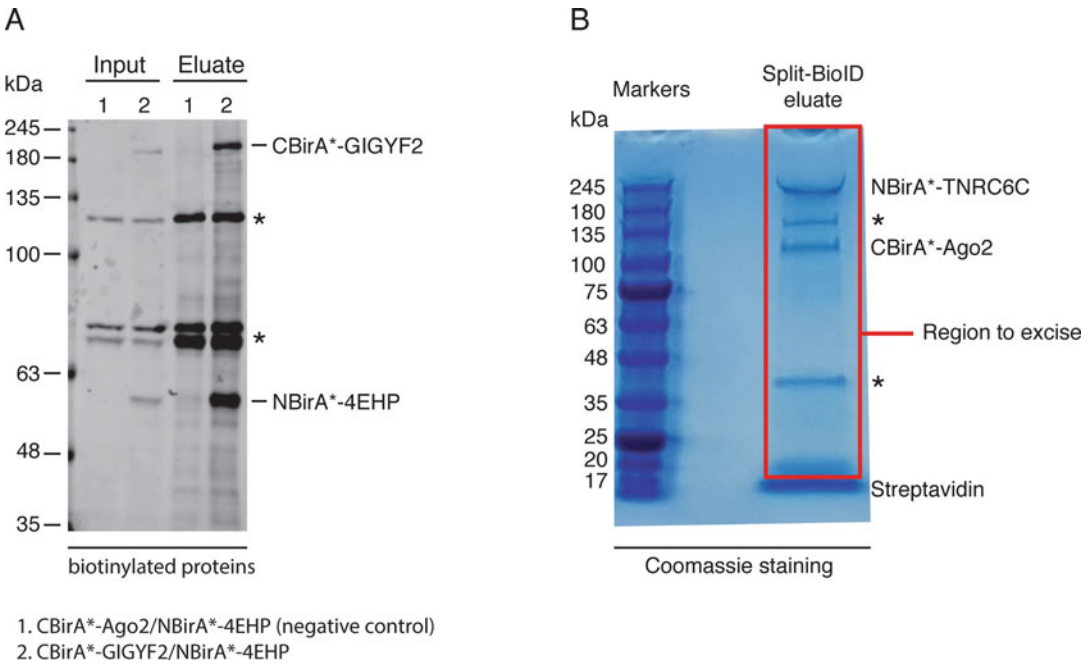


Fig. 4 Typical Western blot and protein gel for a split-BioID experiment. (**a**) Biotinylated proteins are shown for input and eluted samples from a streptavidin pulldown experiment. GIGYF2 and 4EHP are direct binding partners. Ago2 and 4EHP do not interact directly and serve as a negative control pair. (**b**) The elution sample of a split-BioID experiment was run on a protein gel. The region to be excised and processed for MS analysis is indicated. (Panel (**b**) is adapted from Fig. 6 of Schopp and Béthune 2018 [15] under a creative commons license(CRC 3.0, by,nc,nd)). Asterisks indicate endogenous biotinylated proteins

3. Stain the gel with colloidal Coomassie Brillant Blue G250 staining as follows (*see* **Note 26**): wash the gel three times for 10 min with ultrapure water in a 15-cm bacterial dish. Incubate the gel with Coomassie staining solution for 2–12 h (until bands for streptavidin and some biotinylated proteins are clearly visible) at RT. Optionally destain for 1 h at RT with destaining solution. Rinse the gel twice briefly with ultrapure water.
4. For each sample, excise the whole lane with a clean scalpel, excluding the streptavidin band (typically the main visible band running at approximately 17 kDa), and transfer to 1.5 mL tube (Fig. 4b).
5. The excised lanes can be sent to a proteomics facility for in-gel trypsin digestion and MS analysis (*see* **Note 27**).

4 Notes

1. We have tested split-BioID in HeLa cells, in principle it should work in any mammalian cell type and more generally in any cell type in which BioID has been demonstrated to work.
2. Any standard cell culture medium appropriate for the chosen cell line is fine. Most media will already contain biotin as it is an essential vitamin. However, split-BioID like BioID requires the addition of an excess (50 μM) of exogenous biotin to work.
3. Any transfection reagent known to promote high transfection efficiency and low toxicity for the chosen cell line is fine.
4. Biotin has a good solubility in ammonium hydroxide or NaOH. To make a stock solution of biotin, dissolve it at 50 mg mL^{-1} (corresponds to approximately 200 mM) in 2 M ammonium hydroxide or 2 M NaOH. Once biotin is completely dissolved, dilute this solution to 50 mM biotin in 500 mM Hepes, pH 7.4, then adjust the pH to 7.4 with HCl. Sterilize by passing through a 0.22-μM filter, then prepare aliquots in 1.5-mL safe lock tubes and store the resulting 1000× stock solution at −20 °C. We have used up to 2-year-old stock solutions prepared and stored this way without apparent loss of activity.
5. Doxycycline can be dissolved at 10 mg mL^{-1} in 70% ethanol and stored in a screw cap microtube at −20 °C in the dark. The high percentage alcohol ensures that the stock solution is sterile. We have used up to 1-year-old stock solutions prepared and stored this way without apparent loss of activity.
6. We routinely perform streptavidin pulldowns with magnetic Dynabeads MyOne Streptavidin C1 beads. Many other types of beads, including sepharose beads [5] and NeutrAvidin- [7]

or StrepTactin- [9]coupled beads, have also been used successfully by others. Some beads are provided as pre-coated with BSA to decrease nonspecific protein binding. We recommend not using pre-coated beads, as BSA will later be a major source of contamination for the MS analysis.

7. We usually prepare lysis and wash buffers freshly from stock solutions (1 M Tris–HCl, pH 7.4 or 8, 5 M NaCl, 500 mM EDTA pH 8, 20% SDS, 20% Triton X-100, 20% NP-40, 10% sodium deoxycholate, 5 M LiCl, 1 M DTT).
8. Low-fluorescence PVDF membranes are required for fluorescence-based detection of IR dyes-conjugated secondary antibody. From all the PVDF membranes we tested, only Immobilon®-FL PVDF membrane from Millipore gave satisfactory results. Alternatively nitrocellulose membranes typically yield low autofluorescence. If using chemoluminescence-based detection, any immunoblotting membrane is fine. Detection of biotinylated proteins with IR dye- or HRP-conjugated streptavidin works equally well.
9. Ultrafast transfer protocols (7 min) allowed by the Transblot Turbo system from Bio-Rad or equivalent devices is sufficient for the detection of biotinylated proteins with conjugated streptavidin. Any other transfer method is also fine and may yield better results when the detection of the fusion proteins is needed, especially for large-size proteins.
10. The use of precast gels and commercial reagents such as loading buffer is optional but recommended to avoid excessive keratin contamination when running SDS-PAGE for MS samples.
11. We have provided split-BioID plasmids (*see* https://www.addgene.org/Julien_Bethune/) that allow the bidirectional expression of the NBirA* and CBirA* fusions in various orientations and under the control of the same tet-responsive element [8]. Of important note, these plasmids should be used in tet-on or tet-off cell lines. The two fusion proteins can be expressed from the same plasmid or from two separate plasmids. When testing a pair of protein for the first time, we recommend trying all possible iterations of the fusion proteins (NBirA* and CBirA* appended at the N- or C- terminus of each protein of interest, and protein 1 appended to NBirA* or CBirA*). We did observe that certain combination work significantly better than others.
12. If later many of the resulting isolated plasmids prove to be the wrong constructs, a dephosphorylation can be included. Typically we digest the plasmids as indicated, and after 1 h, add alkaline phosphatase buffer and enzyme and let the reaction

proceed for another hour at 37 °C. Alkaline phosphate and restriction enzymes are removed in the following gel extraction step.

13. When using ampicillin selection, the 1 h incubation step at 37 °C can be omitted and the cells directly spread on selection plates.
14. These numbers are indicative and must be adapted to each cell line, including HeLa cells from different lab origin.
15. As a negative control, we recommend pairing each the two tested proteins with an unrelated protein. Whereas using NBirA*-GFP as an unrelated protein worked fine for us, we recommend not using CBirA*-GFP as an additional control protein as it always yielded considerable background biotinylation in our hands. This is likely due to the very high expression level of CBirA*-GFP that probably leads to spontaneous re-association of CBirA* with NBirA* [8]. Other unrelated CBirA* fusions worked fine for us. A powerful negative control for MS analysis is a split-BioID or BioID experiment (or a pool of experiments) applied to two interacting proteins that are unrelated to the proteins of interest.
16. Doxycycline is necessary if using tet-responsive element–regulated plasmids in tet-on cell lines.
17. Samples may be stored at −20 °C at that point.
18. If using fluorescence-based detection and Tween-20 as a source of autofluorescence, a fourth short wash of the membrane with plain PBS is recommended before detection.
19. We had good experience with the 9E10 (dilution 1:1000) and M2 (dilution 1:500) mouse monoclonal antibodies directed against myc and FLAG, respectively.
20. If detection of the fusion proteins proves to be challenging, incubation of the primary antibodies overnight at 4 °C may yield stronger signals.
21. Cell lysates can be very viscous. If needed the syringe can be filled the first few times without the needle.
22. These parameters worked well for a Bioruptor plus sonification device from Diagenode and are given as guidance when using another sonication bath.
23. We have tended to observe less efficient binding to the streptavidin-coupled beads at much higher salt concentrations.
24. These numbers are for Dynabeads MyOne Streptavidin C1 beads and may need to be adjusted when using other types of beads.
25. We chose to elute biotinylated proteins from the beads rather than performing on-bead trypsin digestion. The latter is the

most widely used in published BioID experiments but digestion of coupled streptavidin, which is in large excess to the captured biotinylated proteins, yields a large amount of Streptavidin peptides that might affect the sensitivity of the MS analysis. On the other hand, elution of biotinylated proteins requires very harsh denaturing conditions that induce significant release of streptavidin from the beads. We address this by running the eluted sample on a gel and cutting the lane above the streptavidin bands (running at 17 kDa). Whereas this procedure allows minimizing streptavidin contamination it has the disadvantage of excluding small molecular weight proteins (MW < approximately 20 kDa) from the MS analysis. StrepTactin-coupled beads were described to allow efficient capture of biotinylated proteins and elution in mild conditions using an excess of free biotin [9, 10]. This procedure may thus be the method of choice to minimize streptavidin contamination and allow analyzing the full range of captured biotinylated proteins.

26. In our hands this procedure that has been previously described by Dyballa and Metzger [11] consistently produces very sensitive staining with a fast and easy protocol. Other sensitive and MS-compatible staining procedures can also be applied.
27. Standard MS analysis protocols similar to those used to analyze sample from AP approaches are fine. More elaborate protocols that also aim at enriching and identifying biotinylated peptides were also described [9, 12–14].

Acknowledgments

This work was supported by the excellence initiative by the German federal and state governments: DFG-EXC81, cluster of excellence CellNetworks.

References

1. Scott DE, Bayly AR, Abell C et al (2016) Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. Nat Rev Drug Discov 15(8):533–550. https://doi.org/10.1038/nrd.2016.29
2. Lam SS, Martell JD, Kamer KJ et al (2015) Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat Methods 12(1):51–54. https://doi.org/10.1038/nmeth.3179
3. Roux KJ, Kim DI, Raida M et al (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol 196(6):801–810. https://doi.org/10.1083/jcb.201112098
4. Kim DI, Birendra KC, Zhu W et al (2014) Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc Natl Acad Sci U S A 111(24):E2453–E2461. https://doi.org/10.1073/pnas.1406459111
5. Lambert JP, Tucholska M, Go C et al (2015) Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. J Proteome 118:81–94. https://doi.org/10.1016/j.jprot.2014.09.011

6. Morriswood B, Havlicek K, Demmel L et al (2013) Novel bilobe components in Trypanosoma brucei identified using proximity-dependent biotinylation. Eukaryot Cell 12 (2):356–367. https://doi.org/10.1128/EC.00326-12
7. De Munter S, Gornemann J, Derua R et al (2017) Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions. FEBS Lett 591(2):415–424. https://doi.org/10.1002/1873-3468.12548
8. Schopp IM, Amaya Ramirez CC et al (2017) Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. Nat Commun 8:15690. https://doi.org/10.1038/ncomms15690
9. Opitz N, Schmitt K, Hofer-Pretz V et al (2017) Capturing the Asc1p/receptor for activated C kinase 1 (RACK1) microenvironment at the head region of the 40S ribosome with quantitative BioID in yeast. Mol Cell Proteomics 16 (12):2199–2218. https://doi.org/10.1074/mcp.M116.066654
10. Liu X, Salokas K, Tamene F et al (2018) An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. Nat Commun 9(1):1188. https://doi.org/10.1038/s41467-018-03523-2
11. Dyballa N, Metzger S (2009) Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels. J Vis Exp 30. https://doi.org/10.3791/1431
12. Kim DI, Cutler JA, Na CH et al (2018) BioSITe: a method for direct detection and quantitation of site-specific biotinylation. J Proteome Res 17(2):759–769. https://doi.org/10.1021/acs.jproteome.7b00775
13. Mackmull MT, Klaus B, Heinze I et al (2017) Landscape of nuclear transport receptor cargo specificity. Mol Syst Biol 13(12):962. https://doi.org/10.15252/msb.20177608
14. Udeshi ND, Pedram K, Svinkina T et al (2017) Antibodies to biotin enable large-scale detection of biotinylation sites on proteins. Nat Methods 14(12):1167–1170. https://doi.org/10.1038/nmeth.4465
15. Schopp IM, Béthune J (2018) Split-BioID—proteomic analysis of context-specific protein complexes in their native cellular environment. J Vis Exp 134:57479. https://doi.org/10.3791/57479

Chapter 18

A Directed Evolution System for Lysine Deacetylases

Martin Spinck, Maria Ecke, Damian Schiller, and Heinz Neumann

Abstract

Lysine acetylation is a ubiquitous modification permeating the proteomes of organisms from all domains of life. Lysine deacetylases (KDACs) reverse this modification by following two fundamentally different enzymatic mechanisms, which differ mainly by the need for NAD^+ as stoichiometric co-substrate. KDACs are often found as catalytic subunit in protein complexes involved in cell cycle regulation, chromatin organization and transcription. Their promiscuity with respect to sequence context and type of lysine acylation convolutes the network of functional and physical connections.

Here we present an efficient selection method for KDACs in *E. coli*, which allows for the creation of acyl-type specific KDAC variants, which greatly facilitate the investigation of their physiological function. The selection system builds on the incorporation of acylated lysines by genetic code expansion in reporter enzymes with essential lysine residues. We describe the creation of KDAC mutant libraries by saturation mutagenesis of active site residues, the isolation of individual mutants from this library using the selection system, and their biochemical characterization with acylated firefly luciferase.

Key words Genetic code expansion, Non-canonical amino acids, Lysine deacetylases, Directed evolution

1 Introduction

Lysine acetylation was first discovered on histone proteins and linked to transcriptional regulation by neutralizing the positive charge of lysine, thereby loosening the DNA–histone interaction [1]. The reversible nature of this process was first suggested in 1977 due to a spike in histone acetylation observed upon *n*-butyrate treatment [2]. Twenty years later, the responsible enzyme, histone deacetylase 1 (HDAC1), was isolated by affinity chromatography, and HDAC2 was independently identified by sequence homology to yeast Rpd3 as a transcription factor [3, 4]. This elicited interest in this enzyme family and led to the discovery of 16 further human KDACs [5].

KDACs are divided into two families (histone deacetylases and sirtuins) and four classes (I–IV) based on their reaction mechanism and sequence homology. Classes I, II, and IV consist of histone

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247,
https://doi.org/10.1007/978-1-0716-1126-5_18,

deacetylases that use a catalytic zinc ion to hydrolyze the peptide bond, while class III enzymes, the sirtuins, use NAD^+ in a unique mechanism for cleavage [6, 7]. While histones were the first reported substrates of KDACs, thousands of other proteins and several new modifications have now been identified in mammals [8], plants [9, 10], fungi [11] and bacteria [12]. Among the identified proteins are important cellular components, for example, p53 as substrate of Sirt1 [13] and HDAC1 [14] and α-tubulin of HDAC6 [15].

In addition to the large number of protein substrates, it was found that various lysine modifications (butyryl, crotonyl, malonyl, etc.) can be reversed by KDACs [8]. While the function of many of these modifications is still unclear, it has been shown that they possess their own set of effector proteins such as readers [16], writers [17], and eraser [18].

Due to the promiscuous activity of KDACs to various substrates, regulation plays an important role, which in the case of KDACs often means that they are part of larger protein complexes, for example the transcriptional regulators CoREST [19], mSIN3 [20], NuRD [21], or NCoR/SMRT [22]. While most substrate proteins have long been known, the role of individual modifications and their connection to particular KDACs are still unclear. Problems arise from the high substrate overlap of the different KDACs, the vast number of modified proteins and the questionable significance of many acylations, since the modification reaction can also take place non-enzymatically on proteins [23].

Here we describe a powerful bacterial selection system to isolate KDACs by their activity and selectivity. Using genetic code expansion, we create an acylation-responsive bifunctional auxotrophic marker, Ura3 K93ac, to couple cell growth to KDAC activity. Applying the same general concept, we develop a straightforward luciferase assay to detect various deacylase activities of KDACs. The created system is exemplified by the identification and characterization of acyl-type selective *E. coli* CobB variants from a library of 30 million mutants. This approach allows access to various new biochemical tools, for example, KDACs with specific deacylase activity. This may open new, systematic approaches to deconvolute the complex interactions of KDACs and lysine acylation.

2 Materials

2.1 Strains and Plasmids (Table 1)

2.2 Creation of KDAC Mutant Libraries by Inverse PCR

1. Expand High Fidelity Polymerase (11732650001, 3.5 U μL^{-1}, Sigma Aldrich).
2. 3 M NaOAc pH 5.2 (ITW Reagents).

Table 1
Strains and plasmids

Cell line	Manufacturer	Genotype
Efficiency™ DH10B™ competent cells	Thermo fisher Scientific	F$^-$ *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 *recA1 endA1 ara*D139 Δ(*ara, leu*)7697 *galU galK* λ$^-$ *rpsL nupG*/pMON14272/pMON7124
DH10B ΔcobB ΔpyrF	Spinck et al. 2020 [24]	[DH10B] *pyrF cobB*
BL21 (DE3)	Novagen	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*
pPylT-URA3	Spinck et al. 2020 [24]	scURA3 cloned in pPylT
pPylT-URA3K93TAG-PylS	Spinck et al. 2020 [24]	PylS cloned cistronically in pPylT-URA3K93TAG
pPylT-URA3K93TAG-AcKRS3	Spinck et al. 2020 [24]	pPylT-URA3K93TAG-AcKRS3
pBK-His_6-CobB	Spinck et al. 2018 [25]	pBK-His_6-CobB
pFluc	Spinck et al. 2018 [25]	FlucK529TAG-His_6 cloned in pCDF-PylT
pBK-AcKRS3opt	Spinck et al. 2018 [25]	Encodes AcKRS3 with mutations for improved PylT binding
pBK-PylS	Neumann et al. 2008 [26]	Encodes *M. barkeri* pyrrolysyl-tRNA synthetase

3. 70% ethanol (Thermo Fisher Scientific).
4. 100% ethanol.
5. Primers (100 μM, PAGE purified, Sigma Aldrich, **Note 1**).
6. dNTPs (10 mM, Thermo Fisher Scientific).
7. QIA Quick PCR Purification Kit.
8. Bsa I-HF (20 U μL^{-1}, NEB).
9. Dpn I (20 U μL^{-1}, NEB).
10. T4 DNA Ligase (5 U μL^{-1}, Thermo Fisher Scientific).
11. 10× T4-Ligase-Buffer (Thermo Fisher Scientific).
12. Hind III (10 U μL^{-1}, Thermo Fisher Scientific).
13. Electroporation Cuvette (2 mm, sterile, VWR).
14. Recovery Medium (Invitrogen).
15. Thermocycler (SensoQuest).
16. Centrifuge 5424R (Eppendorf).
17. Vortex Genie 2 (Scientific Industries).
18. Thermomixer comfort (Eppendorf).
19. NanoDrop (ND-1000, Peqlab).
20. Electroporation System GenePulser Xcell (BioRad).

2.3 Selection of Lysine Deacetylases

1. M9 Medium: 42.5 g/l $Na_2HPO_4{\cdot}2H_2O$, 15.0 g/l KH_2PO_4, 2.5 g/l NaCl, 5.0 g/l NH_4Cl, 1 mM $MgSO_4$, 0.4% glucose, 0.1 mM $CaCl_2$ in sterile ddH_2O.
2. LB medium: 1% tryptone, 0.5% yeast extract, 1% NaCl.
3. LB agar plates: LB medium, 2% agar, 15 μg/mL tetracycline (tet), and 50 μg/mL kanamycin (kan).
4. Positive selection agar plates: M9 medium, 2% agar, 0.2% arabinose, 1% glycerol, 4.1 mg/mL tryptophan, 16 mg/mL valine, 16 mg/mL leucine, 16 mg/mL isoleucine, 50 μg/mL kan, 15 μg/mL tet, 10 mM *N*(ε)-acetyl-L-lysine (or 1 mM of any other modified lysine).
5. Negative selection agar plates: M9 minimal medium, 2% agar, 0.2% arabinose, 1% glycerol, 4.1 mg/L tryptophan, 16 mg/L valine, 16 mg/L leucine, 16 mg/L isoleucine, 50 μg/mL kan, 15 μg/mL tet, 10 mM *N*(ε)-acetyl-L-lysine (or 1 mM of any other modified lysine), 0.1% 5-fluoroorotic acid, 0.1 mM uracil.
6. Phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4.
7. Petri dishes (150 × 20 mm, Greiner Bio-One).
8. 1 M in H_2O. H-Lys(Ac)-OH (BACHEM).
9. 50 mg/mL kanamycin (Applichem).

10. 15 mg/mL tetracyclin (Sigma Aldrich).
11. 10% in DMSO 5-Fluoroorotic acid (Zymo Research).
12. 25 mM in H_2O Uracil (Applichem).
13. 96 deep well block (AXYGEN).
14. 10 mm cuvettes (Sarstedt).
15. Plasmid Isolation Kit (Roti-Prep Plasmid MINI, ROTH).
16. Incubator IPP55 (Memmert).
17. Multichannel pipette (10–100 µL Research plus, Eppendorf).
18. Photometer (Biowave CO8000 Cell Density Meter, WPA).
19. Centrifuge 5810 R (Eppendorf).

2.4 KDAC Assay Based on Acylated Firefly Luciferase

1. 20% L(+)-Arabinose (Roth).
2. 50 mg/mL spectinomycin.
3. 1 M IPTG (Roth).
4. 2 M Nicotinamide (Sigma Aldrich).
5. Pierce™ Protease Inhibitor Mini Tablets, EDTA Free (Thermo Fisher Scientific).
6. 200 mM Phenylmethylsulfonylfluorid (PMSF, Roth) in isopropanol.
7. 1 M Dithiothreitol (DTT, AppliChem).
8. Lysozyme (SERVA).
9. 1 mg per purification DNase I (Roche).
10. HisPur™ Ni-NTA Resin (Thermo Fisher Scientific).
11. CobB wash buffer: 20 mM HEPES, pH 7.5, 200 mM NaCl, 20 mM DTT, 1 mM PMSF supplemented with lysozyme (0.1 mg/mL), and DNAse (~0.1 mg).
12. Elution buffer (CobB wash buffer supplemented with 200 mM imidazole).
13. Gel filtration buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM DTT).
14. KDAC reaction buffer: 25 mM Tris, pH 8, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 1 mg/mL BSA.
15. 2× Luciferase Buffer: 40 mM Tricine, pH 7.8, 200 µM EDTA, 7.4 mM $MgSO_4$, 2 mM $NaHCO_3$, 34 mM DTT, 0.5 mM ATP, 0.5 mM luciferin (Thermo Fisher Scientific).
16. 96-well PCR plates (clear, semi-skirted, Thermo Fisher Scientific).
17. 80 mM NAD^+, in 200 mM Tris pH 8 (Sigma Aldrich).
18. 96-well plate (Microfluor® 2, black, flat bottom, Thermo Fisher Scientific).

19. FLuc wash buffer: 20 mM Tris–HCl, 10 mM imidazole, 200 mM NaCl, 10 mM DTT, 2 mM PMSF, 0.5× Roche Protease Inhibitor cocktail, 20 mM nicotinamide, pH 8.0, 0.1 mg/mL Lysozyme, DNase I.
20. FLuc-store buffer: 20 mM Tris pH 8, 50 mM NaCl, 10 mM DTT.
21. Microfluidizer (110S, Hyland SCIENTIFIC).
22. Poly-Prep® Chromatography Columns (9 cm, 2 mL bed volume, Bio-Rad).
23. Amicon Ultra-15 (10 kDa, 30 kDa, Merck).
24. Superdex 200 column (HiLoad 26/600, GE Healthcare).
25. NanoDrop (ND-1000, Peqlab).
26. Microtiter plate reader (FLUOstar Omega, BMG Labtech).

3 Methods

The following three sections outline the necessary steps for the creation, selection, and characterization of a KDAC mutant library. We are using *E. coli* KDAC CobB as an example. CobB is a multifunctional enzyme able to reverse various lysine modifications such as acetylation, crotonylation, propionylation or butyrylation. In this example, we decided to remove its decrotonylation activity by directed evolution.

3.1 Creation of KDAC Mutant Libraries by Inverse PCR

The active site residues of sirtuins were randomized by three rounds of enzymatic inverse PCR [27]. Positions in this example were chosen based on the crystal structure of CobB (pdb-file: 1S5P [28]) close to the bound acetyl-lysine head group (~7 Å radius, *see* Fig. 1).

3.1.1 Inverse PCR

Primers contain randomized bases (NNK) at the sites of interest and recognition sites for Bsa I-HF to allow scar less ligation for plasmid circularization (Fig. 2).

1. Set up several PCRs in parallel as indicated in Table 2 (*see* **Notes 2** and **3**):
2. Run PCR according to the program detailed in Table 3.
3. Add 1 µL of Dpn I (20 U μL^{-1}, NEB) to each 50 µL PCR. Incubate for 1 h, 37 °C.
4. Purify PCR product by QIA Quick PCR Purification Kit or by Gel purification kit (*see* Fig. 3).
5. Bsa I-HF digestion of the PCR products for 2 h at 37 °C (Table 4).
6. Purify PCR product by QIA Quick PCR Purification Kit and elute in 50 µL.

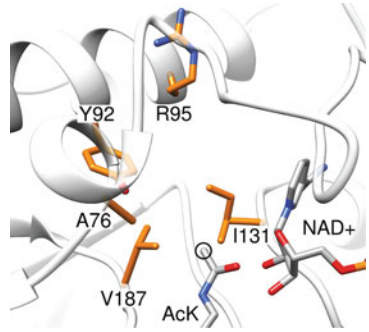


Fig. 1 Active site of the CobB crystal structure (1S5P [28]) with bound peptide. The acetyl-lysine (AcK) and NAD^+ (placed with chimera from 2H4F [29]) are shown, the selected atom to place the sphere for the selection of the active site residues is indicated (black circle). Five amino acids (orange) were chosen because of their distance to the indicated atom and to the NAD^+. The codons of all five amino acids were replaced by NNK codons in three rounds of mutagenesis

3.1.2 Ligation

1. Digested PCR product is pooled and ligated overnight at 16 °C (*see* **Note 4**).

	Ligation(scale to adjust DNA concentration <10 ng/μl)
ddH_2O	325 μL
Ligation buffer (10×)	50 μL
Digested PCR product	100 μL
T4 ligase (5 U μL^{-1})	25 μL
Total volume	500 μL

2. Precipitate ligation reaction with ethanol by addition of 1/10 vol. of 3 M NaOAc, pH 5.2–500 μL of ligation reaction and 2 vol. of 100% ethanol.
3. Let the solution stand on ice for 15 min.
4. Centrifuge at 19,000 × *g* for 45 min at 4 °C (*see* **Note 5**).
5. Remove the supernatant.
6. Add 1 mL of 70% ethanol.
7. Vortex shortly and let stand on ice for 10 min.
8. Centrifuge at 19,000 × *g* for 45 min at 4 °C.
9. Repeat the last two steps one more time.
10. Remove the ethanol.

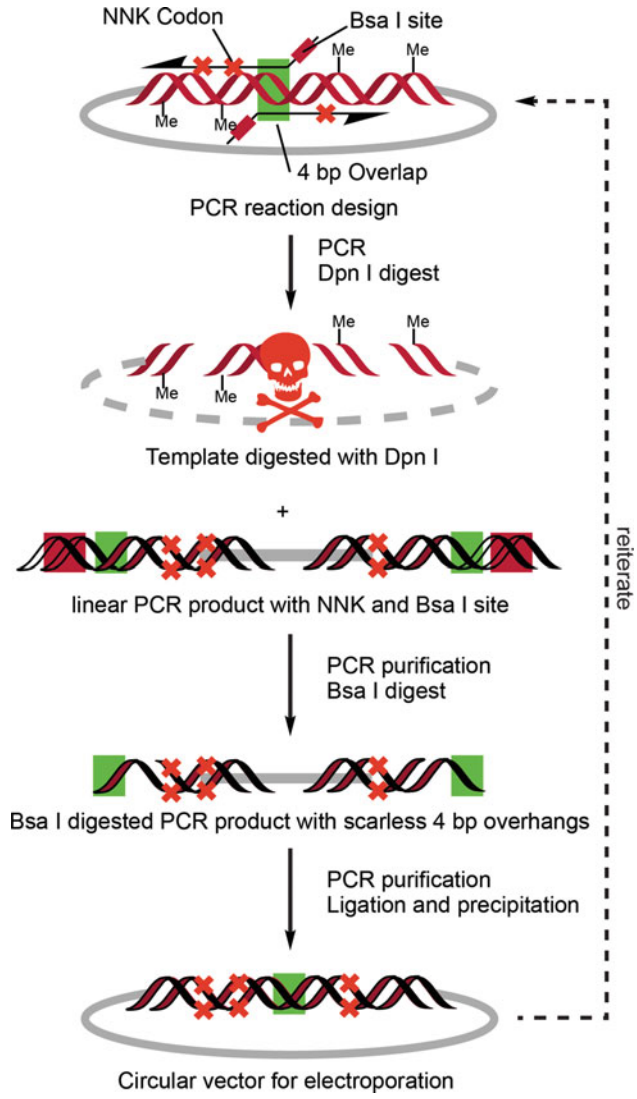


Fig. 2 Workflow of library creation. Primers are designed to overlap with 4 bp (green box) flanked by Bsa I sites (dark red box) on the 5′-side, so that the overlap region will become single-stranded DNA after digestion. Several NNK codons (red cross) can be introduced by a single primer pair. Amplifying the target plasmid in a PCR yields a linear product (black) containing the mutations, the Bsa I site on both sides of the duplicated 4 bp overlap. The template is destroyed using Dpn I, single-strand overhangs created by Bsa I digest and the product ligated

11. Evaporate leftover ethanol (37 °C or RT) (*see* **Note 6**).
12. Dissolve all pellets in 10–40 μL ddH_2O (pooling all ligation reactions).
    (a) First, pipette the droplet above the DNA pellet (the pellet may be invisible).

Table 2
Composition of the inverse PCR mix

ddH_2O	41 μL
Buffer expand high fidelity (10×)	5 μL
Volume of each primer (100 μM)	0.5 μL
dNTP mix (10 mM)	1 μL
Template (~400 ng μL^{-1})	1 μL
Expand polymerase (3.5 U μL^{-1})	1 μL
Total volume	50 μL

Table 3
Cycles for the inverse PCR

	Time (min:sec)	Temp.	# Passes	
Denat.	1:00	95 °C	1	
Denat.	0:20	95 °C	10	
Anneal.	0:30	65 °C	10	−1 °C/cycle
Elong.	5:00	68 °C	10	
Denat.	0:20	95 °C	25	
Anneal.	0:30	55 °C	25	
Elong.	5:00	68 °C	25	

(b) Then slowly pull down the droplet with the pipette tip.

(c) When the round droplet touches the DNA pellet, it will spread and run to the bottom of the tube slowly.

(d) Repeat this process several times until there is no DNA left (droplet stays round).

(e) Transfer the solution to the next tube.

13. Determine DNA concentration by NanoDrop.
14. Test ligation efficiency by Hind III digest of 100 ng of DNA in a 10-μL reaction (*see* **Note 7** and Fig. 3).

3.1.3 Electroporation

1. *E. coli* MAX Efficiency™ DH10B™ Competent Cells (Invitrogen) are thawed on ice.
2. 4 μL ligated PCR product are mixed with 40 μL Competent Cells in a 1.5-mL reaction tube.
3. Incubate 1 min on ice.
4. Transfer 40 μL to the bottom of a 0.2 cm electroporation cuvette (BioRad).

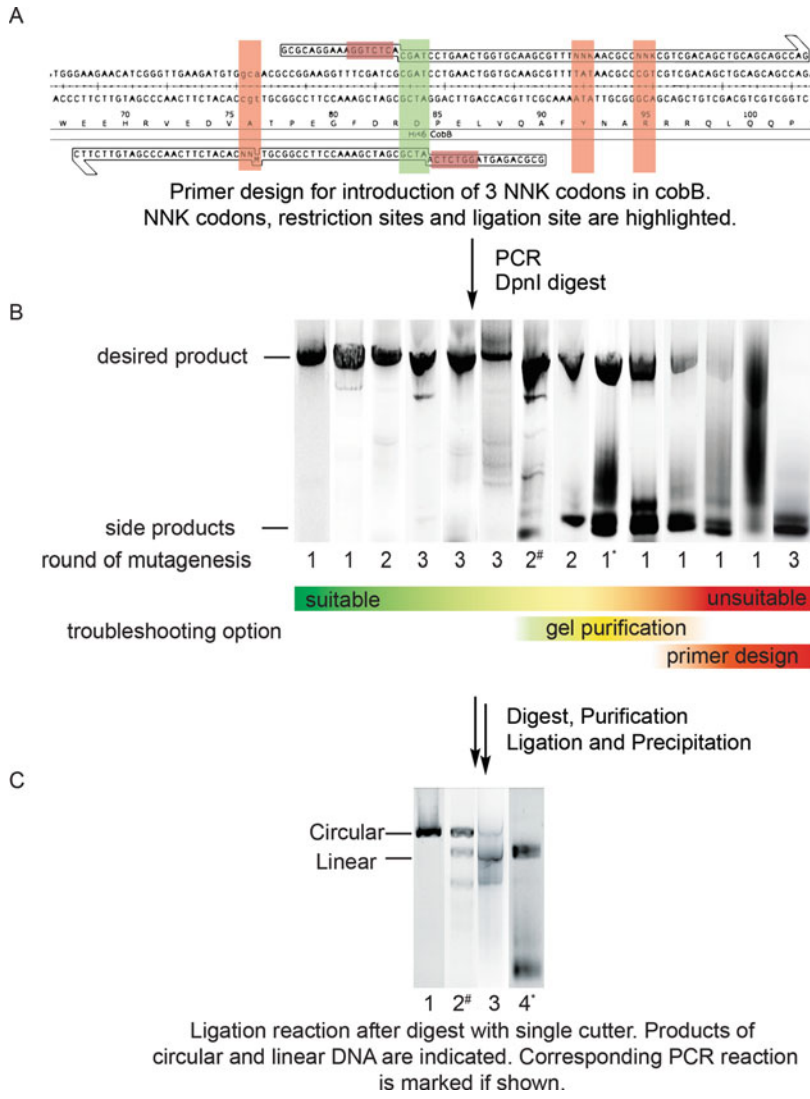


Fig. 3 Examples for each step are from the creation of various libraries. (**a**) Example for a primer pair, important sequence features are highlighted. (**b**) Examples for PCR products ranked from suitable to unsuitable for subsequent steps. Troubleshooting recommendations are indicated. Start a library with the least efficient PCR, because the required number of transformants increases while performance usually declines in successive rounds of mutagenesis. (**c**) Efficiency of the ligation should be tested before electroporation. A circular vector will give a single band (1) when digested with a single cutting restriction enzyme, if the ligation is incomplete, it will result in two extra bands with different intensities depending on the ligation efficiency (2, 3). No ligation results in just a double band (4). Reason for incomplete ligation can be the presence of side products capping the linear DNA (compare PCRs labeled # and * in **b**). Transformation of 3 or 4 did not yield any or sufficient numbers of colonies, while transformation of 2 was still sufficient

Table 4
Restriction digest of PCR product (scale to match volume of PCR product)

PCR product	30 μL
ddH_2O	11 μL
Cut smart buffer (10×)	5 μL
Bsa I-HF (20 U μL^{-1})	4 μL
Total volume	50 μL

5. Dry the cuvette with a paper towel and place it in the electroporation system.
6. Electroporate using *E. coli* electroporation settings: 2500 V, 25 μF, 200 Ω.
7. Immediately add 1 mL recovery medium (RT).
8. Transfer the reaction into a 1.5 mL reaction tube.
9. Incubate at 37 °C for 1 h.
10. Pool the transformations in a volume of 250 mL LB medium containing 50 μg/mL kan and plate a dilution series (10^{-4} to 10^{-7}) on LB-kan plates. Incubate both overnight at 37 °C, the flask shaking at 200 rpm.
11. Estimate transformation efficiency from the colony count of the dilution series.
12. Isolate plasmid DNA from the pool and from at least 12 individual clones and sequence the KDAC ORF.
13. If transformation efficiency exceeds three times the diversity of the library and the majority of individual sequences are unique and without frameshifts continue to the next step.
14. Use the plasmid preparation isolated from the pool as template in the next round of mutagenesis.

3.2 Selection of Lysine Deacetylases

The mutant libraries generated above contain a large number of different mutants due to the numerous possible combinations of randomized positions (20^5 different proteins coded for by 32^5 codon combinations if five positions are mutated). Most of the mutations will be detrimental to protein function or folding, rendering the mutants inactive. Hence, an efficient selection system is required to identify the mutants of interest in a large background of inactive ones. Therefore, we designed a bacterial selection system that allows parallel screening of each mutant in a single step [24]. We created an *E. coli* strain without deacetylase activity and impaired uracil biosynthesis by inactivating the genes for CobB and PyrF using CRISPR/Cas9 and Lambda Red Recombineering [30]. We complemented the uracil auxotrophic strain with Ura3 (orotidine 5′-phosphate decarboxylase), which contains an amber codon in

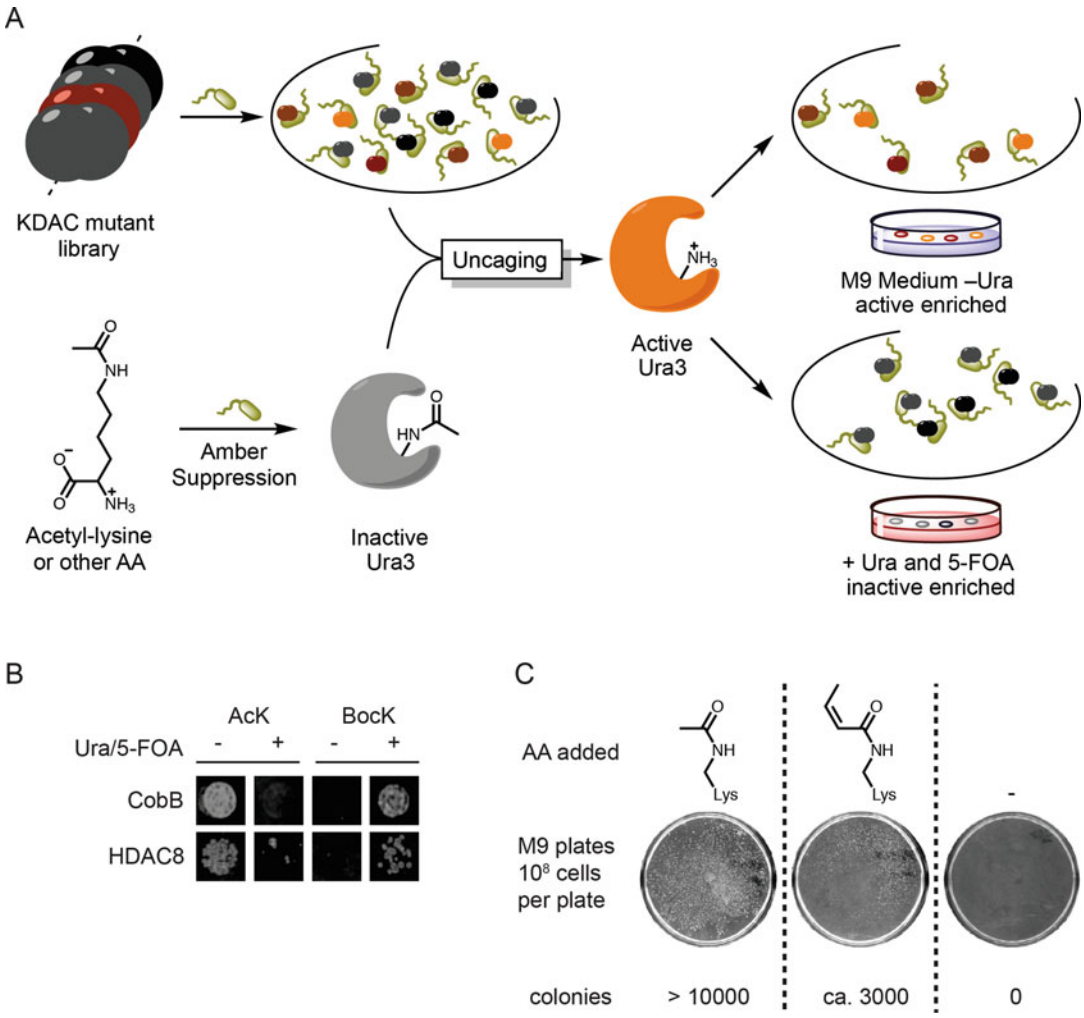


Fig. 4 Outline of the bacterial KDAC selection system. (**a**) Specific mutants can be enriched by transforming *E. coli* expressing acylated Ura3 with the library. Each individual cell will then perform the demodification reaction and only produce active Ura3 if the encoded KDAC mutant has activity for the attached modification. The active Ura3 can then be used to either remove or enrich the active mutants depending on the selection plate composition. (**b**) Phenotyping. DH10B *ΔcobB ΔpyrF* expressing Ura3 K93ac or K93boc and either CobB or HDAC8 were grown on positive (–Ura) or negative (+5-FOA/Ura) selection conditions. Both KDACs are able to cleave AcK and hence grow only under positive selection. The situation is inversed when BocK is incorporated which neither KDAC can cleave. (**c**) Example for a positive selection of the CobB active site mutant library in the presence or absence of AcK or CrK. Approximately 10^8 cells were plated in each case (library diversity approximately 3.4×10^7). Number of colonies gives an estimate of surviving mutants and is used as a reference for the required transformation efficiency in subsequent selection rounds

place of the codon for an essential lysine in its active site (K93) [31]. Suppression of this codon with an acylated lysine by genetic code expansion [26] produces an enzyme that can be activated by co-expression of a KDAC. This allows for positive (in the absence of uracil) and negative (in the presence of 5-fluoroorotic acid [32] and uracil) selection of its activity (Fig. 4).

3.2.1 Preparation of Electrocompetent Cells (E. coli ΔpyrF ΔcobB pURA3K93TAG-PylS/AcKRS3)

1. Grow an overnight culture of DH10B Δ*cobB* Δ*pyrF* pPylT-URA3K93TAG-PylS or DH10B Δ*cobB* Δ*pyrF* pPylT-URA3K93TAG-AcKRS3 (5 mL LB medium containing 15 μg/mL tet).
2. Inoculate 100 mL LB with 1 mL overnight culture and incubate at 37 °C for 2.5–3 h until the OD_{600} reaches 0.3.
3. Harvest the cells by centrifugation at 2500 × *g* for 5 min at 4 °C and remove the medium.
4. Suspend the cells with 30–40 mL ice-cold, autoclaved, particle free ddH_2O.
5. Centrifuge at 2000 × *g* for 5 min at 4 °C and discard the supernatant.
6. Repeat **steps 4** and **5** once again (*see* **Note 8**).
7. Centrifuge at 2000 × *g*, remove remaining water.
8. Suspend cells in 200 μL ddH_2O.
9. Keep the cells on ice until further use.

3.2.2 Selection of KDACs

1. Electroporate freshly prepared electrocompetent DH10B Δ*cobB* Δ*pyrF* pPylT-URA3K93TAG-PylS or DH10B Δ*cobB* Δ*pyrF* pPylT-URA3K93TAG-AcKRS3 with 1 μg of the CobB mutant library.
2. Follow **steps 3**–**7** of Subheading 3.1.3.
3. After the recovery time, add all transformed cells to 100 mL LB medium containing 50 μg/mL tet and 50 μg/mL kan. Prepare a small-scale serial dilution down to a factor of 10^{-7} and plate them on appropriate LB agar plates. Incubate media and plates at 37 °C overnight.
4. Estimate the number of transformants from the colony count of plates. It should exceed three times the number of all possible mutants in the library.
5. Inoculate 50 mL LB containing 50 μg/mL tet and 50 μg/mL kan with 1 mL of overnight culture and incubate at 37 °C until the OD_{600} exceeds 0.3 (*see* **Note 9**).
6. Harvest the cells by centrifugation at 500 × *g* for 5 min at 4 °C and remove the medium.
7. Resuspend the cells in sterile PBS. For first positive selection, plate out aliquots of 8×10^8 cells on each positive selection agar plate and incubate at 37 °C for 2–3 days. For negative selection and all successive selections, plate aliquots of 8×10^5 cells on each negative selection plate and incubate at 37 °C for 16 h (*see* **Note 10**).
8. Add ~3 mL sterile PBS onto the plate and scrape the colonies from the agar with the help of a sterile cell spreader. Distribute

the cells into aliquots of 500 µL and isolate the plasmid DNA using a commercial kit. Pool plasmid DNA for the next round of selection.

9. Repeat **steps 1–8** until all selection rounds are completed.
10. Transform DH10B cells with the plasmid DNA obtained by the selection via heat shock. Plate the cells on agar plates containing 50 µg/mL kan and incubate them overnight at 37 °C.
11. On the next day, prepare a 96 deep well block by filling each well with 1 mL LB medium containing 50 µg/mL kan. Inoculate each well by picking single colonies of the transformed DH10B cells. Incubate the block overnight at 37 °C.
12. Harvest the cells by centrifuging the 96 well block for 10 min at 2700 × *g* and remove the media.
13. Isolate the plasmid DNA of each well and sequence the KDAC ORF (*see* **Note 11**).

3.2.3 6-AU Assay (Optional)

This step is optional, but can help filter out the most active variants from a pool of mutants. Therefore, repeat the last selection step of Subheading 3.2.2 with the addition of 6-azauridine (6-AU) to the agar plates (0.01–2 mM, at least 10 different concentrations). This can be done by plating the pool of mutants or by replicating single colonies from a 96-well plate onto the selection plates. 6-AU is an inhibitor of Ura3, which requires cells to produce more active Ura3, increasing the stringency of the positive selection assay.

3.3 KDAC Assay Based on Acylated Firefly Luciferase

Similar to Ura3, the firefly luciferase (FLuc) also contains an essential lysine (K529) in its active site. When modified, the luciferase is completely inactive and therefore can be used to monitor the activity of enzymes capable of reversing the modification [25]. We use this to assay the activity and selectivity of the isolated mutants.

3.3.1 Preparation of $FLuc_{mod}$

1. Transform *E. coli* BL21 DE3 with plasmids pFluc and pBK-AcKRS3opt or pBK-PylS depending on the desired modification.
2. Grow the cells in LB medium (50 µg/mL spectinomycin and 50 µg/mL kan, 5 mM *N*(ε)-acetyl-lysine or 1 mM other AS and 20 mM nicotinamide) at 37 °C to and OD_{600} of 1.0.
3. Shift cells to 30 °C and induce protein expression by the addition of 1 mM IPTG.
4. After 4 h at 30 °C harvest cells by centrifugation, wash with PBS, and suspend in Ni-wash buffer.
5. Lyse cells by passing them three times through a microfluidizer.
6. Centrifuge lysate at 20,000 × *g* for 20 min and transfer the supernatant to a falcon tube.

7. Supplement the supernatant with 500 µL Ni-NTA-beads.
8. Incubate the beads for 2 h with agitation at 4 °C.
9. Wash the beads with 50 mL FLuc-wash buffer and elute bound proteins in 2 mL FLuc-wash buffer supplemented with 200 mM imidazole.
10. Exchange buffer to FLuc-store buffer using a centrifugal filter.
11. Use the Fluc as deacetylase substrate (*see* **Note 12**).

3.3.2 Purification of CobB

1. Transform *E. coli* Δ*pyrF* Δ*cobB* with pBK-His_6-CobB and pPylT-Ura3 by a method of choice (*see* **Note 13**).
2. Setup 20 mL overnight culture in LB medium (containing 15 µg/mL tet, 50 µg/mL kan) from a single colony.
3. Inoculate 1L LB medium (containing 15 µg/mL tet, 50 µg/mL kan) with 10 mL overnight culture.
4. Grow to OD 0.3–0.5 at 37 °C and 200 rpm.
5. Decrease temperature to 30 °C and induce protein expression by addition of arabinose (0.2% final).
6. Harvest cells by centrifugation (6000 × *g*, 20 min) after 16 h of expression.
7. Wash the cell pellet with PBS and store at −20 °C.
8. Suspend the cell pellet in 25 mL CobB wash buffer supplemented with lysozyme (0.1 mg/mL) and DNase (~0.1 mg).
9. Lyse cells by passing through a microfluidizer 2–3 times.
10. Pellet cell debris by centrifugation and decant supernatant into a Falcon tube.
11. Add 500 µL Ni-NTA resin to the supernatant, incubate for 1 h with agitation.
12. Pour the suspension into a plastic column with frit.
13. Wash with 50 mL CobB wash buffer.
14. Elute protein in 2 mL Elution buffer.
15. Concentrate fractions using a centrifugal filter and exchange buffer by diluting twice with gel filtration buffer.
16. Apply the concentrated fractions to a Superdex 200 column equilibrated with gel filtration buffer (2.5 mL/min, 4 °C).
17. Collect 4 mL fractions while monitoring protein concentration at 280 nm.
18. Analyze protein containing fractions by SDS-PAGE, pool, and concentrate by ultrafiltration.
19. Flash freeze protein in liquid nitrogen and store at −80 °C.

3.3.3 Characterization of Isolated Mutants

For testing a large number of mutants for a desired property, KDACs partially purified by Ni-NTA (Subheading 3.3.2, **steps 1–14**) or even cell lysates can be used if mutants can be expressed in an *E. coli* Δ*cobB* strain (e.g., DH10B Δ*cobB* Δ*pyrF*). The assay is exceptionally sensitive, so that even small contaminations with KDACs such as *E. coli* CobB will produce high background signals. All isolated mutants are expressed, and activity for different modifications is assessed. Eventually, the desired KDACs should be purified to homogeneity by Ni-NTA affinity and gel filtration chromatography for full characterization.

1. Prepare cell lysate (Subheading 3.3.2, **steps 1–10**), Ni-Eluate (**steps 1–14**), or fully purified KDACs (**steps 1–18**) of all mutants and a wild-type control.
2. Estimate protein concentration by SDS-PAGE or nano drop.
3. Set up reactions in 96-well PCR plates in triplicate for every lysine modification as following: 0.2 mg/mL KDAC, 2 mM NAD^+, 10 μL Luciferase eluate (1:125 to 1:8000 dilution, depending on the signal strength, at least 100-fold over background) in KDAC reaction buffer to a final volume of 50 μL.
4. Incubate at 30 °C for 30 min with 50 °C lid temperature in a thermocycler.
5. Transfer 40 μL of the reaction mixture to a 96-well microtiter plate.
6. Add 40 μL of 2× Luciferase Buffer.
7. Measure luminescence using a Microplate reader.
8. Normalize activity to the signal of the wild-type.
9. Mutants with enhanced selectivity and/or activity are selected for further analysis.
10. Perform **steps 1–8** with various concentrations of mutant and wild-type enzyme (between 1 nM and 16 μM).
11. Activity can be compared in the linear region of the obtained curve for each individual substrate.

4 Notes

1. The primer needs to be as defined as possible. Prime fragments can lead to PCR artifacts and increase side products. Ultimately this decreases library quality and increases the effort in creating it.
2. Total volume is adjusted to required diversity. For example, 4×10^7 mutants approximately 10–16 PCRs are required, depending on the yield and ligation efficiency.

3. Side product formation severely decreases the efficiency of ligation, requiring more PCRs for sufficient DNA amounts after gel purification.
4. High concentrations of DNA will lead to the formation of concatemers, scale volume to adjust DNA concentration <10 ng/μl.
5. Longer centrifugation times improve DNA recovery.
6. Do not overly dry the DNA.
7. More than 50% of the DNA should be circular for efficient electroporation.
8. If the cells continue to stick to the wall of the tube and do not get resuspended easily, add one more washing step.
9. The cells tend to grow very slowly at this step. An OD600 of 1 would be ideal, but values as low as 0.3 are also sufficient if the growth takes too much time.
10. The default procedure for screening is one round of positive selection in the beginning, then one round of negative selection and finally one more round of positive selection with the same conditions as in the first round. If there are only very few colonies growing on the selection plates, the screening can be stopped after one or two rounds. In this case, the single colonies can be picked directly for plasmid DNA isolation, and **step 7** can be skipped.
11. The 96-well block can be stored at −20 °C, and the isolation of plasmid DNA can be continued on another day.
12. Titrate dilution series to find the optimal working concentration for your deacetylase.
13. Due to low transformation efficiency, it is recommended to successively transform the plasmids. A stock of *E. coli* DH10B ΔcobB ΔpyrF competent cells with pPylT-Ura3 plasmid can be stored at −80 °C. The pPylT-Ura3 plasmid is required to provide sufficient AraC protein for the induction of CobB expression. Since CobB isolates are initially encoded on minimal pBK plasmids, this method avoids an intermediate recloning.

Acknowledgments

The authors are grateful for the financial support by the Max-Planck Institute of Molecular Physiology and the German Research Foundation (DFG) [NE1589/5-1 to H.N.].

References

1. Allfrey VG, Faulkner R, Mirsky AE (1964) Acetylation and methylation of histones and their possible role in the regulation of rna synthesis. Proc Natl Acad Sci U S A 51:786–794
2. Riggs MG, Whittaker RG, Neumann JR, Ingram VM (1977) *n*-Butyrate causes histone modification in HeLa and friend erythroleukaemia cells. Nature 268:462. https://doi.org/10.1038/268462a0
3. Taunton J, Hassig CA, Schreiber SL (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science 272:408
4. Yang WM, Inouye C, Zeng Y, Bearss D, Seto E (1996) Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. Proc Natl Acad Sci U S A 93:12845–12850
5. Seto E, Yoshida M (2014) Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol 6:a018713. https://doi.org/10.1101/cshperspect.a018713
6. Sauve AA (2010) Sirtuin chemical mechanisms. Biochim Biophys Acta 1804:1591–1603. https://doi.org/10.1016/j.bbapap.2010.01.021
7. Lombardi PM, Cole KE, Dowling DP, Christianson DW (2011) Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. Curr Opin Struct Biol 21:735–743. https://doi.org/10.1016/j.sbi.2011.08.004
8. Tan M et al (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell 146:1016–1028. https://doi.org/10.1016/j.cell.2011.08.008
9. Finkemeier I, Laxa M, Miguet L, Howden AJ, Sweetlove LJ (2011) Proteins of diverse function and subcellular location are lysine acetylated in Arabidopsis. Plant Physiol 155:1779–1790. https://doi.org/10.1104/pp.110.171595
10. Wu X et al (2011) Lysine acetylation is a widespread protein modification for diverse proteins in Arabidopsis. Plant Physiol 155:1769–1778. https://doi.org/10.1104/pp.110.165852
11. Henriksen P et al (2012) Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in *Saccharomyces cerevisiae*. Mol Cell Proteomics 11:1510–1522. https://doi.org/10.1074/mcp.M112.017251
12. Colak G et al (2013) Identification of lysine succinylation substrates and the succinylation regulatory enzyme CobB in *Escherichia coli*. Mol Cell Proteomics 12:3509–3520. https://doi.org/10.1074/mcp.M113.031567
13. Vaziri H et al (2001) hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. Cell 107:149–159
14. Luo J, Su F, Chen D, Shiloh A, Gu W (2000) Deacetylation of p53 modulates its effect on cell growth and apoptosis. Nature 408:377–381. https://doi.org/10.1038/35042612
15. Hubbert C et al (2002) HDAC6 is a microtubule-associated deacetylase. Nature 417:455–458. https://doi.org/10.1038/417455a
16. Andrews FH et al (2016) The Taf14 YEATS domain is a reader of histone crotonylation. Nat Chem Biol 12:396–398. https://doi.org/10.1038/nchembio.2065
17. Sabari BR et al (2015) Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. Mol Cell 58:203–215. https://doi.org/10.1016/j.molcel.2015.02.029
18. Wei W et al (2017) Class I histone deacetylases are major histone decrotonylases: evidence for critical and broad function of histone crotonylation in transcription. Cell Res 27:898–915. https://doi.org/10.1038/cr.2017.68
19. Meier K, Brehm A (2014) Chromatin regulation: how complex does it get? Epigenetics 9:1485–1495. https://doi.org/10.4161/15592294.2014.971580
20. Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P, Reinberg D (1997) Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell 89:357–364. https://doi.org/10.1016/S0092-8674(00)80216-0
21. Xue Y et al (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell 2:851–861
22. Wen YD et al (2000) The histone deacetylase-3 complex contains nuclear receptor corepressors. Proc Natl Acad Sci U S A 97:7202–7207
23. Simithy J et al (2017) Characterization of histone acylations links chromatin modifications with metabolism. Nat Commun 8:1141. https://doi.org/10.1038/s41467-017-01384-9
24. Spinck M, Neumann-Staubitz P, Ecke M, Gasper R, Neumann H (2020) Evolved, selective erasers of distinct lysine acylations. Angew

Chem Int Ed Engl 59:11142–11149. https://doi.org/10.1002/anie.202002899

25. Spinck M, Ecke M, Sievers S, Neumann H (2018) Highly sensitive lysine deacetylase assay based on acetylated firefly luciferase. Biochemistry 57:3552–3555. https://doi.org/10.1021/acs.biochem.8b00483
26. Neumann H, Peak-Chew SY, Chin JW (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. Nat Chem Biol 4:232–234. https://doi.org/10.1038/nchembio.73
27. Stemmer WP, Morris SK (1992) Enzymatic inverse PCR: a restriction site independent, single-fragment method for high-efficiency, site-directed mutagenesis. BioTechniques 13:214–220
28. Zhao K, Chai X, Marmorstein R (2004) Structure and substrate binding properties of cobB, a Sir2 homolog protein deacetylase from Escherichia coli. J Mol Biol 337:731–741. https://doi.org/10.1016/j.jmb.2004.01.060
29. Hoff KG, Avalos JL, Sens K, Wolberger C (2006) Insights into the Sirtuin mechanism from ternary complexes containing NAD+ and acetylated peptide. Structure 14 (8):1231–1240
30. Pyne ME, Moo-Young M, Chung DA, Chou CP (2015) Coupling the CRISPR/Cas9 system with lambda red Recombineering enables simplified chromosomal gene replacement in Escherichia coli. Appl Environ Microbiol 81:5103–5114. https://doi.org/10.1128/aem.01248-15
31. Appleby TC, Kinsland C, Begley TP, Ealick SE (2000) The crystal structure and mechanism of orotidine 5′-monophosphate decarboxylase. Proc Natl Acad Sci U S A 97:2005–2010. https://doi.org/10.1073/pnas.259441296
32. Boeke JD, Trueheart J, Natsoulis G, Fink GR (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol 154:164–175

INDEX

Arnaud Poterszman (ed.), *Multiprotein Complexes: Methods and Protocols*, Methods in Molecular Biology, vol. 2247, https://doi.org/10.1007/978-1-0716-1126-5,