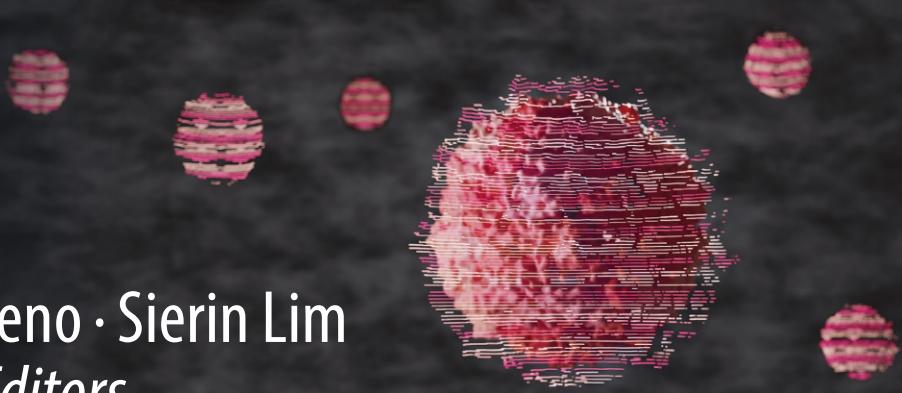


Takafumi Ueno · Sierin Lim
Kelin Xia *Editors*

Protein Cages

Design, Structure, and
Applications



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.

Protein Cages

Design, Structure, and Applications

Edited by

Takafumi Ueno

Tokyo Institute Of Technology, Yokohama, Japan

Sierin Lim

Chemical & Biomedical Engineering, Nanyang Technological University, Singapore, Singapore

Kelin Xia

Nanyang Technological University, Singapore, Singapore



Editors

Takafumi Ueno
Tokyo Institute Of Technology
Yokohama, Japan

Sierin Lim
Chemical & Biomedical Engineering
Nanyang Technological University
Singapore, Singapore

Kelin Xia
Nanyang Technological University
Singapore, Singapore

ISSN 1064-3745

ISSN 1940-6029 (electronic)

Methods in Molecular Biology

ISBN 978-1-0716-3221-5

ISBN 978-1-0716-3222-2 (eBook)

<https://doi.org/10.1007/978-1-0716-3222-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

Protein cages are one of the vital protein assemblies that play important roles in cellular activities. In nature, their functions range from storage molecules and biocatalysts to carriers of genetic materials and scaffolds for multiple displays of enzyme complexes. Their unique properties, such as assembly/disassembly ability, symmetrical structures, and sizes, as well as the amenability to chemical and biological modifications, make them attractive for applications beyond their natural functions. The versatility of protein cages has landed them as exciting targets in various fields, such as cell biology, drug delivery, structural biology, and material science. Recent development of protein cage research and advancement in interdisciplinary science leads to exciting translation for protein cages.

This book is an update of the 2015 *Protein Cages: Methods and Protocols* under the same series, Methods in Molecular Biology. It highlights state-of-the-art methodology to design, prepare, functionalize, characterize, and analyze protein cages in solution and at interfaces. The chapters are categorized into four parts. Part I details the design and construction of artificial cages. Part II provides information on functionalization and modifications of protein cages. As new protein cage systems emerge and their application expands, Part III highlights new ways to characterize the structures and functions of protein cages, notably the emulsion system and electronic properties, as well as simulation and modeling. The drive towards translation of protein cage systems requires extensive characterizations not only *in vitro* but also *in vivo*. Part IV includes methods to evaluate protein cage for applications in health and materials science and concludes with ways to produce protein cages beyond the convention.

The book has been a pleasure to put together where we collate contributions from the experts in protein cage from a range of research fields. The 23 chapters cover fundamentals of protein cage to the latest methods for translation. We hope this book offers comprehensive guidance and inspiration to all protein cage researchers, from beginners to experts, and helps create new ideas for next-generation applications.

We thank all authors for contributing their expertise. We appreciate the assistance in editorial processing by Ms. Anna Rakovsky and Mr. Patrick Marton. Finally, we are grateful to the series editor, Prof. John M. Walker, for the opportunity and guidance throughout the editing process.

*Yokohama, Japan
Singapore, Singapore
Singapore, Singapore*

*Takafumi Ueno
Sierin Lim
Kelin Xia*

Contents

Preface	v
Contributors	xi

PART I DESIGN AND CONSTRUCTION OF ARTIFICIAL CAGES

1 Coiled-Coil Protein Origami: Design, Isolation, and Characterization	3
<i>Žiga Strmšek, Jaka Snoj, Tadej Satler, and Roman Jerala</i>	
2 Artificial Protein Cages Assembled via Gold Coordination	49
<i>Karolina Majsterkiewicz, Izabela Stupka, Kinga Borzęcka-Solarz, Artur Biela, Szymon Gaweł, Monika Pasternak, and Jonathan Heddle</i>	
3 Reassembly Design of Ferritin Cages	69
<i>Yu Liu and Guanghua Zhao</i>	
4 Protein Cages and Nanostructures Constructed from Protein Nanobuilding Blocks	79
<i>Naoya Kobayashi and Ryoichi Arai</i>	
5 Preparation of Cage-Like Micellar Assemblies of Engineered Hemoproteins	95
<i>Koiji Oohora and Takashi Hayashi</i>	

PART II FUNCTIONALIZATION OF PROTEIN CAGES

6 Fabrication of Protein Macromolecular Frameworks (PMFs) and Their Application in Catalytic Materials	111
<i>Masaki Uchida, Ekaterina Selivanovitch, Kimberly McCoy, and Trevor Douglas</i>	
7 Iron Accumulation in Ferritin	121
<i>Akankshika Parida and Rabindra K. Behera</i>	
8 A Generalized Method for Metal Fixation in Horse Spleen L-Ferritin Cage	135
<i>Basudev Maity and Takafumi Ueno</i>	
9 Dual Modification of Artificial Protein Cage	147
<i>Norifumi Kawakami, Erika Nasu, and Kenji Miyamoto</i>	
10 Modification and Production of Encapsulin	157
<i>Sandra Michel-Souzy and Jeroen J. L. M. Cornelissen</i>	

PART III CHARACTERIZATION AND SIMULATION OF PROTEIN CAGE SYSTEMS

11 Cryo-electron Microscopy of Protein Cages	173
<i>Raymond N. Burton-Smith and Kazuyoshi Murata</i>	

12	Time-Resolved Small-Angle X-Ray Scattering of Protein Cage Assembly	211
	<i>Daisuke Sato, Takaaki Hikima, and Masamichi Ikeguchi</i>	
13	Protein Cage-Stabilized Emulsions: Formulation and Characterization	219
	<i>Mridul Sarker, Samuel Watts, Stefan Salentinig, and Sierin Lim</i>	
14	A Method to Investigate the Mechanism of Charge Transport Across Bio-Molecular Junctions with Ferritin.....	241
	<i>Senthil Kumar Karuppannan and Christian A. Nijhuis</i>	
15	Potato Virus X Inactivation and Characterization	257
	<i>Yifeng Ma and Nicole F. Steinmetz</i>	
16	Molecular Dynamics Simulation of Protein Cages.....	273
	<i>Chenlin Lu, Xue Peng, and Diannan Lu</i>	
17	Coarse-Grained Models for Vault Normal Model Analysis	307
	<i>D. Vijay Anand, Ronald Koh Joon Wei, and Kelin Xia</i>	

PART IV TRANSLATION OF PROTEIN CAGES

18	Evaluating Anti-tumor Immune Responses of Protein Nanoparticle-Based Cancer Vaccines	321
	<i>Enya Li, Nina Butkovich, Jo A. Tucker, Edward L. Nelson, and Szu-Wen Wang</i>	
19	Construction Protocol of Drug-Protein Cage Complexes for Drug Delivery System.....	335
	<i>Yuichi Nakahara, Yuta Endo, and Ipppei Inoue</i>	
20	Protein Cage Relaxivity Measurement for Magnetic Resonance Imaging Contrast Agents	349
	<i>Ambrish Kumar, Bhargy Sharma, and Sierin Lim</i>	
21	Crystalline Biohybrid Materials Based on Protein Cages	361
	<i>Hendrik Böhler, Michael Rütten, Laurin Lang, and Tobias Beck</i>	
22	Production and Purification of Virus-Like Particles by Transient Expression in Plants	387
	<i>Lygie Esquirol, Donna McNeale, Micol Venturi, and Frank Sainsbury</i>	
23	Laboratory Scale Production of Complex Proteins Using Charge Complimentary Nanoenvironments	403
	<i>Girish Vallerinteavide Mavelli, Samira Sadeghi, and Chester Lee Drum</i>	
	<i>Index</i>	419

Contributors

D. VIJAY ANAND • *School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore*

RYOICHI ARAI • *Department of Biomolecular Innovation, Institute for Biomedical Sciences, Shinshu University, Ueda, Nagano, Japan; Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda, Nagano, Japan*

TOBIAS BECK • *Universität Hamburg, Department of Chemistry, Institute of Physical Chemistry, Hamburg, Germany; The Hamburg Centre for Ultrafast Imaging, Hamburg, Germany*

RABINDRA K. BEHERA • *Department of Chemistry, National Institute of Technology, Rourkela, Odisha, India*

ARTUR BIELA • *Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland*

HENDRIK BÖHLER • *Universität Hamburg, Department of Chemistry, Institute of Physical Chemistry, Hamburg, Germany*

KINGA BORZECKA-SOLARZ • *Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland*

RAYMOND N. BURTON-SMITH • *Exploratory Research Center on Life and Living Systems (ExCELLS), National Institute for Natural Sciences, Okazaki, Aichi, Japan; National Institute for Physiological Sciences (NIPS), National Institute for Natural Sciences, Okazaki, Aichi, Japan*

NINA BUTKOVICH • *Department of Chemical and Biomolecular Engineering, University of California, Irvine, Irvine, CA, USA*

JEROEN J. L. M. CORNELISSEN • *Department of Molecules and Materials, MESA+ Institute for Nanotechnology, University of Twente, Enschede, The Netherlands*

TREVOR DOUGLAS • *Department of Chemistry, Indiana University, Bloomington, IN, USA*

CHESTER LEE DRUM • *Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore*

YUTA ENDO • *Research Institute for Bioscience Products & Fine Chemicals, Ajinomoto Co., Inc., Kawasaki, Kanagawa, Japan*

LYGIE ESQUIROL • *Centre for Cell Factories and Biopolymers, Griffith Institute for Drug Discovery, Griffith University, Nathan, QLD, Australia*

SZYMON GAWEL • *Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland*

TAKASHI HAYASHI • *Department of Applied Chemistry, Graduate School of Engineering, Osaka University, Yamadaoka, Suita, Japan*

JONATHAN HEDDLE • *Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland*

TAKAAKI HIKIMA • *RIKEN SPring-8 Center, Sayo, Hyogo, Japan*

MASAMICHI IKEGUCHI • *Department of Bioinformatics, Soka University, Hachioji, Tokyo, Japan; Department of Biosciences, Soka University, Hachioji, Tokyo, Japan*

IPPEI INOUE • *Research Institute for Bioscience Products & Fine Chemicals, Ajinomoto Co., Inc., Kawasaki, Kanagawa, Japan*

ROMAN JERALA • *Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia*

- SENTHIL KUMAR KARUPPANAN • *National Quantum Fables Foundry (NQFF), Institute of Materials Research and Engineering, Singapore, Singapore*
- NORIFUMI KAWAKAMI • *Department of Bioscience and Informatics, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Japan*
- NAOYA KOBAYASHI • *Division of Materials Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Nara, Japan*
- AMBRISH KUMAR • *Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA*
- LAURIN LANG • *Universität Hamburg, Department of Chemistry, Institute of Physical Chemistry, Hamburg, Germany; The Hamburg Centre for Ultrafast Imaging, Hamburg, Germany*
- ENYA LI • *Department of Chemical and Biomolecular Engineering, University of California, Irvine, Irvine, CA, USA*
- SIERIN LIM • *School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, Singapore, Singapore*
- YU LIU • *College of Food Science & Nutritional Engineering, China Agricultural University, Beijing, China*
- CHENLIN LU • *Department of Chemical Engineering, Tsinghua University, Beijing, China*
- DIANNAN LU • *Department of Chemical Engineering, Tsinghua University, Beijing, China*
- YIFENG MA • *Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA*
- BASUDEV MAITY • *School of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Japan*
- KAROLINA MAJSTERKIEWICZ • *Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland*
- GIRISH VALLERINTEAVIDE MAVELLI • *Tong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore*
- KIMBERLY MCCOY • *Department of Chemistry and Biochemistry, California State University, Fresno, Fresno, CA, USA*
- DONNA MCNEALE • *Centre for Cell Factories and Biopolymers, Griffith Institute for Drug Discovery, Griffith University, Nathan, QLD, Australia*
- SANDRA MICHEL-SOUZY • *Department of Molecules and Materials, MESA+ Institute for Nanotechnology, University of Twente, Enschede, The Netherlands*
- KENJI MIYAMOTO • *Department of Bioscience and Informatics, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Japan*
- KAZUYOSHI MURATA • *Exploratory Research Center on Life and Living Systems (ExCELLS), National Institute for Natural Sciences, Okazaki, Aichi, Japan; National Institute for Physiological Sciences (NIPS), National Institute for Natural Sciences, Okazaki, Aichi, Japan; Department of Physiological Sciences, School of Life Science, The Graduate University for Advanced Studies (SOKENDAI), Okazaki, Aichi, Japan*
- YUICHI NAKAHARA • *Research Institute for Bioscience Products & Fine Chemicals, Ajinomoto Co., Inc., Kawasaki, Kanagawa, Japan*
- ERIKA NASU • *Department of Bioscience and Informatics, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Japan*
- EDWARD L. NELSON • *Department of Medicine, University of California, Irvine, Irvine, CA, USA; Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, CA, USA*

- CHRISTIAN A. NIJHUIS • *Hybrid Materials for Opto-Electronics Group, Department of Molecules and Materials, MESA+ Institute for Nanotechnology, Molecules Center and Center for Brain-Inspired Nano Systems, Faculty of Science and Technology, University of Twente, Enschede, The Netherlands*
- KOJI OOHORA • *Department of Applied Chemistry, Graduate School of Engineering, Osaka University, Yamadaoka, Suita, Japan*
- AKANKSHIKA PARIDA • *Department of Chemistry, National Institute of Technology, Rourkela, Odisha, India*
- MONIKA PASTERNAK • *Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland*
- XUE PENG • *Department of Chemical Engineering, Tsinghua University, Beijing, China*
- MICHAEL RÜTTEN • *Universität Hamburg, Department of Chemistry, Institute of Physical Chemistry, Hamburg, Germany*
- SAMIRA SADEGHI • *Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore*
- FRANK SAINSBURY • *Centre for Cell Factories and Biopolymers, Griffith Institute for Drug Discovery, Griffith University, Nathan, QLD, Australia*
- STEFAN SALENTINIG • *Department of Chemistry, University of Fribourg, Fribourg, Switzerland*
- MRIDUL SARKER • *School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, Singapore, Singapore*
- TADEJ SATLER • *Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia*
- DAISUKE SATO • *Department of Bioinformatics, Soka University, Hachioji, Tokyo, Japan*
- EKATERINA SELIVANOVITCH • *Department of Chemistry, Indiana University, Bloomington, IN, USA*
- BHARGY SHARMA • *School of Materials Science and Engineering, Nanyang Technological University, Singapore, Singapore*
- JAKA SNOJ • *Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia*
- NICOLE F. STEINMETZ • *Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA; Department of Bioengineering, University of California San Diego, La Jolla, CA, USA; Department of Radiology, University of California San Diego, La Jolla, CA, USA; Center for Nano-ImmunoEngineering, University of California San Diego, La Jolla, CA, USA; Institute for Materials Discovery and Design, University of California San Diego, La Jolla, CA, USA; Moores Cancer Center, University of California San Diego, La Jolla, CA, USA; Center for Engineering in Cancer, Institute for Engineering in Medicine, University of California San Diego, La Jolla, CA, USA*
- ŽIGA STRMŠEK • *Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia*
- IZABELA STUPKA • *Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland*
- JO A. TUCKER • *Department of Medicine, University of California, Irvine, Irvine, CA, USA*
- MASAKI UCHIDA • *Department of Chemistry and Biochemistry, California State University, Fresno, Fresno, CA, USA*
- TAKAFUMI UENO • *School of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Japan; Living Systems Materialogy (LiSM) Research Group, International Research Frontiers Initiative (IRFI), Tokyo Institute of Technology, Yokohama, Japan*

MICOL VENTURI • *Centre for Cell Factories and Biopolymers, Griffith Institute for Drug Discovery, Griffith University, Nathan, QLD, Australia*

SZU-WEN WANG • *Department of Chemical and Biomolecular Engineering, University of California, Irvine, Irvine, CA, USA; Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, CA, USA; Department of Biomedical Engineering, University of California, Irvine, Irvine, CA, USA*

SAMUEL WATTS • *School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, Singapore, Singapore; Department of Chemistry, University of Fribourg, Fribourg, Switzerland*

RONALD KOH JOON WEI • *School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore*

KELIN XIA • *School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore*

GUANGHUA ZHAO • *College of Food Science & Nutritional Engineering, China Agricultural University, Beijing, China*

Part I

Design and Construction of Artificial Cages



Chapter 1

Coiled-Coil Protein Origami: Design, Isolation, and Characterization

Žiga Strmšek, Jaka Snoj, Tadej Satler, and Roman Jerala

Abstract

Coiled-coil protein origami (CCPO) is a rationally designed de novo protein fold, constructed by concatenating coiled-coil forming segments into a polypeptide chain, that folds into polyhedral nano-cages. To date, nanocages in the shape of a tetrahedron, square pyramid, trigonal prism, and trigonal bipyramide have been successfully designed and extensively characterized following the design principles of CCPO. These designed protein scaffolds and their favorable biophysical properties are suitable for functionalization and other various biotechnological applications. To further facilitate the development, we are presenting a detailed guide to the world of CCPO, starting from design (CoCoPOD, an integrated platform for designing CCPO strictures) and cloning (modified Golden-gate assembly) to fermentation and isolation (NiNTA, Strep-trap, IEX, and SEC) concluding with standard characterization techniques (CD, SEC-MALS, and SAXS).

Key words CCPO, Protein origami, Coiled-coil, CC, Protein, CoCoPOD, SAXS, SEC-MALS, CD, FPLC

1 Introduction

Polypeptides/proteins present the most structurally and chemically versatile group of biopolymers, due to the high chemical diversity of amino acids as polypeptide building blocks. Although their structure is mostly defined by the linear amino acid sequence, a reliable design of desired 3D structures still poses a great challenge. There are several strategies to design and explore new protein folds. One of the modular design strategies is coiled-coil protein origami (CCPO). CCPO has been inspired by the principles of DNA origami as DNA and RNA are highly malleable and designable. These highly desirable characteristics, combined with the ease of synthesis and clear interaction principles, facilitated the design of highly

Žiga Strmšek, Jaka Snoj and Tadej Satler contributed equally with all other contributors.

complex nanostructures. With CCPO, we successfully combined high designability and modularity following the example of DNA origami, with chemical and functional versatility of polypeptides. Analogous to the complementary nucleotide pairing used in DNA origami, coiled-coils, which form the basic building modules of CCPO, pair in similar, pair-wise fashion—moreover, the interaction between CC in designable.

Furthermore, coiled-coils (CC) are a favorable protein domain of choice due to their periodicity, rigidity, and orthogonality. Based on these principles, a set of dimeric and orthogonal CCs was designed, characterized [1], and further developed. The first successfully designed and characterized CCPO was a tetrahedral nanocage composed of 6 CC-forming pairs [2]. The principle was further developed with a successful design and characterization of several improved tetrahedral nano-cages, with the addition of square pyramid and trigonal prism [3] and with single and multi-chain bipyramidal nanocages [4]. Moreover, the possibility to bind nanobodies to the edges of CCPO has also been successfully investigated [5] and a tetrahedral protein origami that halves the required CC pairs needed has also been successfully designed and characterized, thus paving a road toward bigger and more complex designs [6]. CCPO describes a design principle for a protein fold, which is unique and not present in the nature. CCPOs are highly modular, robust, and versatile but are as a group of unnatural proteins particular enough as to allow us developing a more or less universal approach to working with them. In this chapter, we will discuss design, cloning, fermentation, and isolation and characterization of CCPO in greater detail. To date, numerous CCPO with different shapes, topologies, and basic biophysical properties have been successfully designed and characterized; therefore, a great set of protocols and empirical comments have been developed. To ease the entry into the CCPO world, a schematic of order of steps is shown in Fig. 1.

2 Design and Modelling (CoCoPOD)

The modelling process of CCPO will be described through the CoCoPOD platform, which automates all the steps of the design process. Coiled-coil protein-origami design platform (CoCoPOD) [3] is a python module used for the design and generation of polyhedral protein models based on the pairing of coiled-coil building blocks from a single polypeptide chain. The module is mostly automated and uses Modeller [7], Chimera [8], and MDtraj [9] dependencies. Source code can be found in the GitHub repository (https://github.com/NIC-SBI/CC_protein_origami.git). A comprehensive description of CoCoPOD computational methodology can be found in [3].

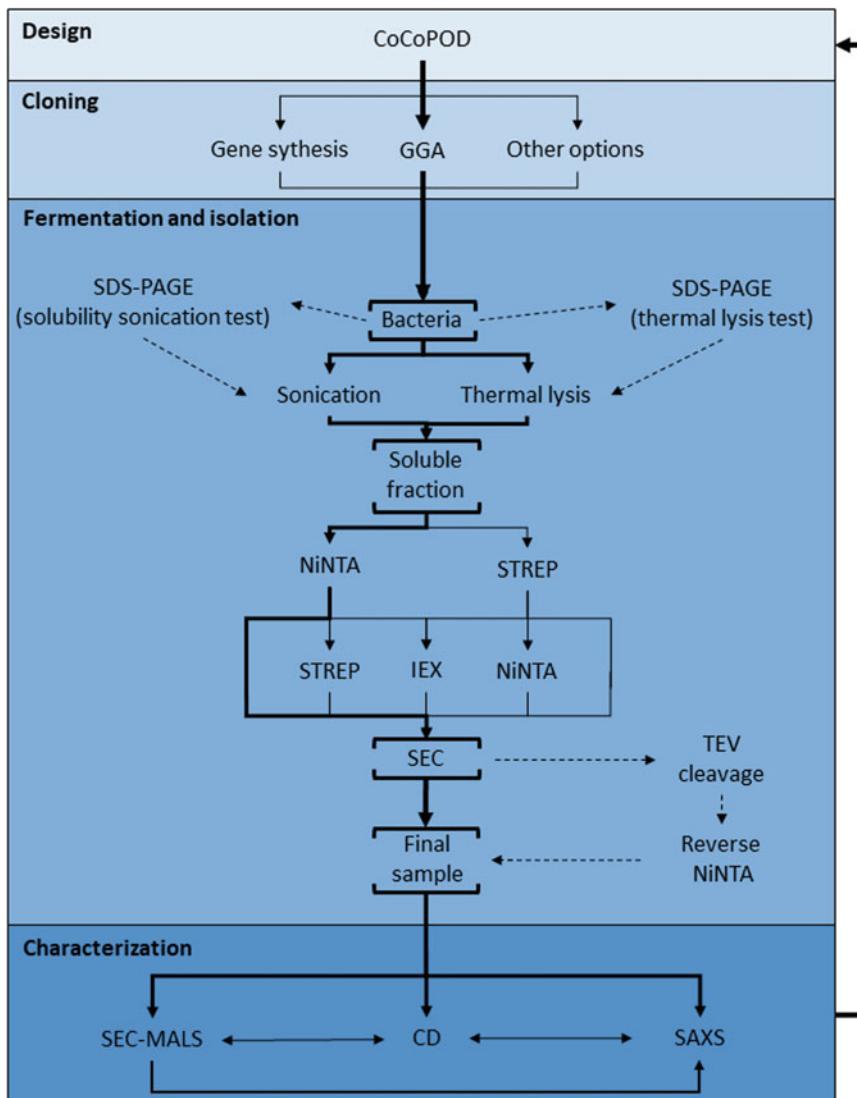


Fig. 1 Workflow for CCPO design, cloning, isolation, and characterization. The width of lines between steps represents the preferred use

2.1 CoCoPOD Installation (See Note 1)

1. Installation of dependencies can be simplified by using python distribution platforms such as Anaconda or Miniconda. Use the following code:

```
conda install numpy scipy pandas ipython Notebook ipywidgets
pyyaml xlrd biopython
```

```
export KEY_MODELLER=XXXX #set the modeller licence key (see
Note 2)
```

```
conda install -c salilab modeller
conda install -c omnia mdtraj
pip install plyfile doit #not available in conda (see Note 3).
```

Also, ensure to enter the Modeller license key, which is available free of charge for academic non-profit institutions (<https://salilab.org/modeller/registration.html>)

2. Install Chimera separately and make sure it is available on the system path.
3. CoCoPOD files can be cloned and installed from the GitHub repository by running:

```
git clone https://github.com/NIC-SBI/protein_origami.git copod
cd copod
python setup.py develop
```

4. Add the location of the CoCoPOD folder to the system path.

2.2 Polyhedral Designs with CoCoPOD (See Note 4)

CoCoPOD design process can be split into several steps—(i) selection of polyhedral geometry, (ii) selection of optimal topology and circular permutation, (iii) coiled-coil building block assignment, and (iv) 3D protein model construction (Fig. 2).

1. Import All the necessary python modules:

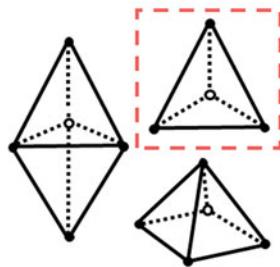
```
from __future__ import print_function, division, absolute_import
import sys
%load_ext autoreload
%autoreload 2

import cocopod
import cocopod.segment_assignment as sa
import cocopod.topology as t
```

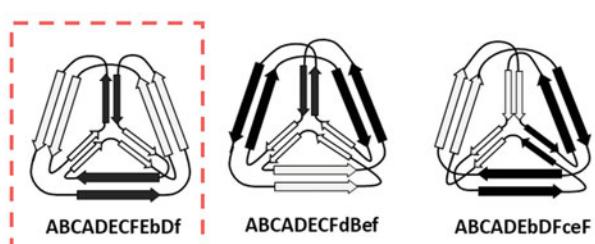
2. Polyhedral geometry selection

Select the desired shape of polyhedral protein by providing geometry information in Stanford PLY ASCII file format (Geometry files for tetrahedron, pyramid, and triangular prism can be found in CoCoPOD folder).
(see Note 5)

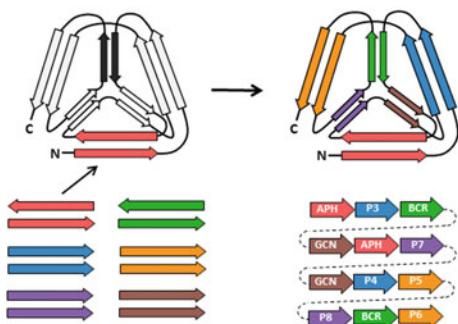
1 Geometry selection



2 Topology and circular permutation selection



3 Building block assignment



4 Building 3D protein model

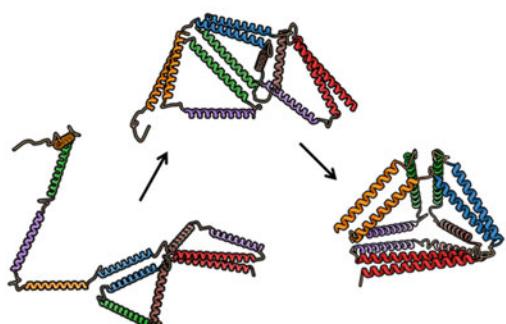


Fig. 2 CoCoPOD polyhedral design and modelling process. Start with the selection of polyhedral geometry (1), then optimal topology and circular permutation (2). Assign coiled-coil modules to selected topology (3) and begin the assembly process of a protein model (4)

```
#load ply file and convert it to edge notation
edge_topology = t.convert_vface_to_efaces(t.load_vfaces
("ply/01_tetrahedron.ply"))
#Explore all topologies
tops = t.explore(edge_topology, True)
#enumerate and name them
tops = t.name_topologies_and_permutations(tops)
#calculate the topological contact order
df = t.calculate_TCO(tops)
df.sort_values(by= ["num_AP", "num_cross", "TCO", "stdTCO", "segments" ])
```

CoCoPOD will generate all the possible topologies as a string of letters, where each letter represents one coiled-coil segment and matching letters signify matching coiled-coil pairs (e.g., ABCADECFCFeDf). Same cased letters indicate pairs with parallel orientation (e.g., AA), while mixed cased letters denote pairs with antiparallel orientation (e.g., Bb). Each

topology can have several different circular permutations that define the vertex in polyhedral shape, in which the polypeptide chain will start and end.

(*see Note 6*)

3. Topology and circular permutation selection

Choose optimal topology and circular permutation based on the lowest topological contact order (TCO), standard deviation of distances between pairs (stdTCO), and number of coiled coils with parallel/antiparallel orientation (*see Note 7*).

```
#Optimal topologies are 1.10 or 1.1R
topology = df.loc['1.10']['segments'] #This is 1.10 or 1.1R
print(topology)
topology = t.standard(topology)
print(topology)
```

4. Building block assignment

Assign topological modules with coiled-coil forming peptide segments from the CC toolbox by connecting the topology letters to the coiled-coil peptide pairs (*see Note 8*).

Choose orthogonal coiled-coil pairs with desired properties and orientation. Several different peptide pairs are already provided with the CoCoPOD module (*see Note 9*).

Determine the amino acid sequence of linkers between peptide segments and specify C- and N-terminal peptide tags (*see Note 10*).

```
seg_assignment_str = """
A->APHshSN:APHshSN
B->P3SN:P4SN
C->BCRSN:BCRSN
D->P9SN:P10SN
E->P1SN:P2SN
F->P11SN:P12SN
"""

model_name = "TET12SN"
linker = "GSGPG"

seg_assignment = sa.segment_assignments_to_dict(seg_assignment_str)
real_segments = sa.do_assignment_replacements(topology, seg_assignment)

#one can also manually edit the sequence
annotated_seq = sa.get_annotated_sequence(real_segments,
'segments.xlsx', N_tag="M", C_tag="LEHHHHHHHH", linkers=linker)
```

```

annotated_seq_edit = sa.sequence_edit_gui(annotated_seq,
caption="--".join(real_segments), model_name=model_name)

included_pairs = [": ".join(s) for s in seg_assignment.values()
()]
pairs_info = sa.get_included_pairs_info('segments.xlsx',
sheetname='pairs', included_pairs=included_pairs)
sa.write_make_config(model_name, annotated_seq_edit.value,
pairs_info, out_name='make_config.py')

```

All the steps until now are to prepare make_config.py file, which contains all the necessary information for model building step—model_name, annotated_sequence, and pairs_info (see **Note 11**).

5. Building a 3D protein model

Specify the number of independent folding simulations with N_fold and the number of independent homology refinements for each folding simulation with the N_homology variable (see **Note 12**).

Start the assembly of polyhedral protein shape by executing the following command:

```
%%bash
doit -n 3 N_fold=1 N_homology=3
```

Modelling starts with Chimera creating a straight helix with provided amino acid sequence. Modeller then sequentially brings coiled-coil forming segments together (Fig. 3a) and adds super-helical twist via homology modelling (Fig. 3b). Generated models are then evaluated and used to fit experimental data (see **Note 13**).

3 Cloning

3.1 Materials

3.1.1 Golden Gate Assembly

1. T4 DNA ligase (NEB, MA, USA).
2. 10× T4 DNA ligation buffer (NEB, MA, USA).
3. BsaI-HF v2 (NEB, MA, USA).
4. PCR tubes.
5. Thermocycler.
6. Competent *E. coli* (see **Note 14**).
7. Liquid media of choice (such as LB and TEB).
8. Agar plates (such as LB and TEB).
9. Antibiotics (should match resistances on used vectors).

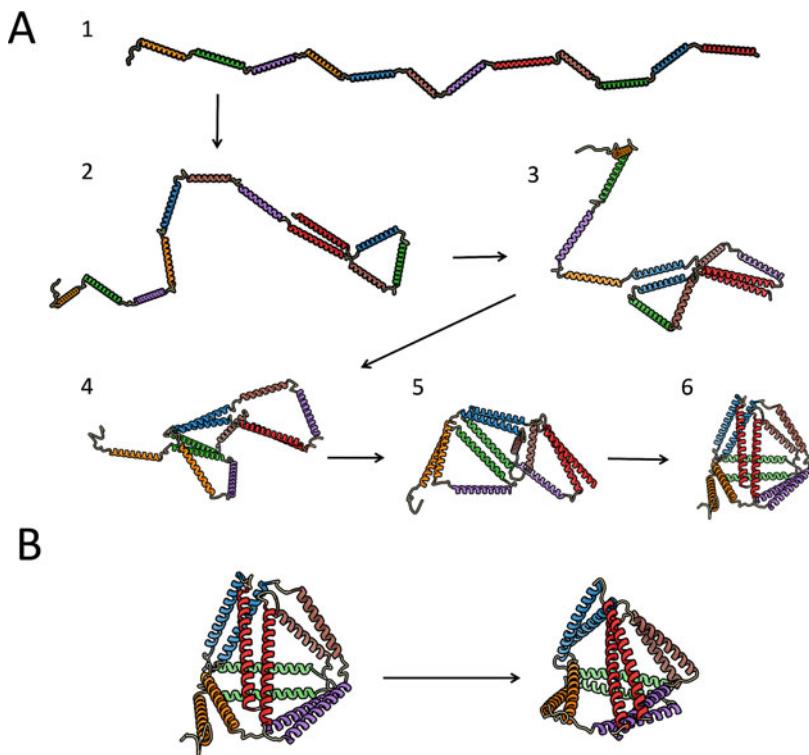


Fig. 3 Schematic representation of automated model building with Modeller. Consecutive folding of coiled-coil forming segments (a) and twisting coiled-coil pairs based on homology models (b)

3.1.2 Colony PCR and Screening

1. DNA isolation kit.
2. Tabletop centrifuge.
3. Tabletop heating block with shaker.
4. PCR master mix, suitable for colony PCR (such as 2× DreamTaq PCR Master Mix (NEB), but others can be used).
5. Capillary electrophoresis or DNA agarose electrophoresis assembly.

3.2 Methods

3.2.1 Golden Gate

Due to the highly modular nature of CCPO structures and a limited number of basic building blocks, Golden gate assembly (GGA) has been adopted (Fig. 4). In short, GGA enables fast assembly of CCPO genes from an existing library of basic building blocks using a designed system of unique DNA overhangs, present in the uniform linker present in all CCPO (linker sequence: GSGPG).

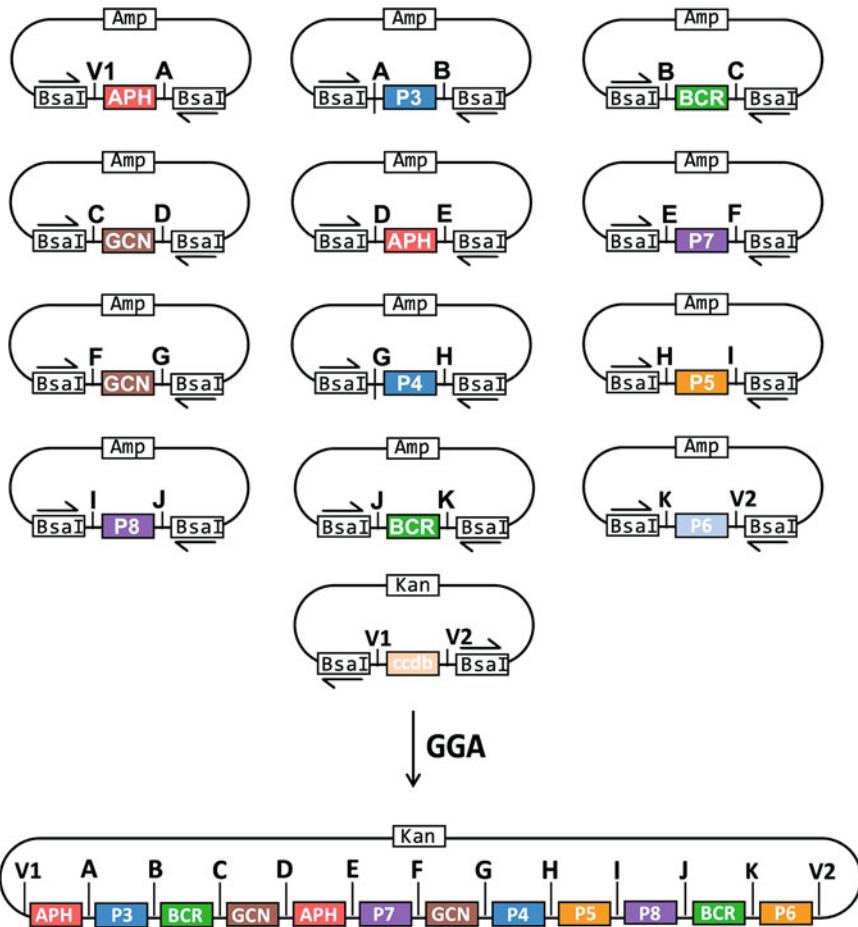


Fig. 4 Schematic representation of Golden gate assembly

1. The schematic design of a single building block for GGA is shown in Fig. 5. Each component of the building block has its purpose:
 - (i). BsaI restriction sites are directional; therefore, special care should be taken when designing DNA sequences.
 - (ii). Unique GGA overhangs have been designed in silico and are presented in Fig. 5 (*see Note 15*).
 - (iii). CC sequences should be designed without restriction sites that could interfere with further cloning steps.
2. A library of single building blocks, inserted into the vector with ampicillin resistance, has been created using a variety of different approaches: (i) ordered directly in the vector, (ii) cloned in the vector using DNA gBlocks or primers and Gibson assembly or classical cloning, and (iii) with modifying already existing building blocks using quick change or Gibson assembly (*see Note 16*).

BsaI	Left linker		Right linker	BsaI
NNNGGTCTCN	NNNNN	CC	NNNNNNGAGACCNNN	NNNNNNCTCTGGNNN
CCACGCATATG		Position 1	GGCTCAC	
CCTCAGGTCCGGGT		Position 2	GGTCAGC	
CTCAGGTCCGGGC		Position 3	GGGTC	
CGGGTCTGGCCCGGGC		Position 4	GGTAGCGGTCTGC	
CCCTGGC		Position 5	GGTCGCG	
CTTCGGTCCGGGT		Position 6	GGTAGTGC	
CACTGGTCCGGGT		Position 7	GGTAGC	
CGTAGCGGTCCGGGC		Position 8	GGCAGTC	
CCAGTGGCCGGGT		Position 9	GGTAGCGGCCCTC	
CCCTGGT		Position 10	GGTAGCGGACCGC	
CACCGGGC		Position 11	GGTAC	
CGGTAGCGGTCCCTGGT		Position 12	CTCGAGC	

Fig. 5 Unique Golden gate assembly overhangs, as designed and used in the assembly of CCPo

3. All needed GGA building blocks should be sequenced, transformed separately, amplified in bacteria, and isolated (*see Note 17*).
4. The accepting vector should have a different resistance vector to the one used in a library of GGA building blocks. Furthermore, the vector of choice should be further modified with the sequence, which codes for ccdb toxin, flanked with BsaI recognition sites, as schematically shown in Fig. 5 (*see Notes 18 and 19*).
5. **GG assembly** (*see Note 20*).
 - (i). 30 µL of DNA master mix should be prepared with the final concentration of each donor vector of 14 ng/µL (*see Note 21*).
 - (ii). To 7.25 µL of DNA master mix, add 1 µL of accepting vector (pET-41a-ccdb) (*see Note 22*), 1 µL 10× T4 ligase buffer (NEB) (*see Note 23*), 0.25 µL T4 ligase (NEB), and 0.5 µL BsaI-HF v2 (NEB) (*see Note 24*). Mix, do a quick down spin and transfer to a thermocycler. Incubate using the following parameters (Table 1):
6. Transfer GGA reaction into 200 µL of freshly thawed competent cells and perform a transformation with a heat shock (*see Note 25*). In short, after the addition of the DNA, mix by gently flicking and transfer back on ice. Incubate on ice for 20–30 min, do a quick heat shock for 1–2 min at 42 °C, followed by rapid cooling on ice for 3–5 min. Add 800 µL of liquid media and incubate for 40–50 min at 37 °C and approx. 300 rpm. Pellet the cells by centrifugation on a tabletop centrifuge at approx. 4000–5000 rpm for 2–3 min. Remove 800 µL of supernatant, resuspend the pellet, transfer to the agar plate, with kanamycin added (*see Note 26*), and incubate overnight at 37 °C or over the weekend at room temperature.

Table 1
Thermal cycling conditions used for Golden gate assembly reactions

	Temperature [°C]	Time [min]
1 ×	37	5
20–25 ×	37	5
	16	10
1 ×	16	20
	37	40
	75	6
	4	∞

3.2.2 Colony PCR and Screening

1. Perform colony PCR on the colonies that have grown on the plate (see Note 27). In short, prepare a fresh agar plate, with a matching antibiotic to be used as a reference plate later on. Mix 20 µL of PCR master mix per colony (see Note 28) and aliquot it into PCR strips. Transfer the small amount of bacterial colony into separate PCR tubes (see Note 29), and on the reference plate (see Note 30). Run the reaction in the thermocycler according to the PCR master mix manufacturer's recommendations and T_m calculation.
2. Analyze PCR reactions using capillary electrophoresis or DNA agarose gel. Isolate the vector from the bacteria that produced DNA bands of the correct size and send for sequencing (see Note 31).
3. For potential further cloning, additional general guidelines should also be considered (see Notes 32 and 33).

4 Fermentation and Isolation

Production of protein can be done one day up to a maximum of 2 weeks before the planned isolation since the pellets can be stored in the freezer. It should be considered that producing a protein takes roughly one whole day. Time permitting, it is recommended to perform the whole isolation (NiNTA and/or StrepTrap and SEC) in one day. Possible breakpoints are as follows: (i) after NiNTA (see Note 34), (ii) after StrepTrap, (iii) after SEC.

4.1 Materials

4.1.1 Bacterial Fermentation for Production of Proteins

1. Stock culture of *E. coli* strain NiCo21(DE3) (can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala Δ hsdS λ DE3 = λ sBamH1o Δ EcoRI-B int:(lacI::PlacUV5::T7 gene1) i21 Δ nin5) (NEB, MA, USA) transformed with expression vector pET41 containing desired genetic construct or some other equivalent expression strain.

2. Antibiotic Kanamycin (50 mg/mL stock): Prepare the desired amount of stock kanamycin in a beaker by mixing the appropriate volume of deionized water with 50 mg/mL of powdered Kanamycin. Mix to dissolve and filter sterilize using 0.22 µm sterile syringe filter unit and store in 1–2 mL aliquots at –20 °C for up to 1 year.
3. IPTG (1 M stock): Prepare a stock of 1 M IPTG in a beaker by dissolving 2.38 g of IPTG in deionized water with a final volume of 10 mL. Filter sterilize using 0.22 µm sterile syringe filter unit and store in 1–2 mL aliquots at –20 °C.
4. Lysogeny broth (LB) liquid media: Prepare according to the manufacturer’s recommendations. In short, prepare up to 1/5 of the total volume of media in Erlenmeyer flasks by mixing the appropriate volume of deionized water with 25 g/L of media. Mix to dissolve, cover with a cap, and autoclave.
5. Erlenmeyer flasks (0.5 L and 5 L).
6. 0.5 L centrifugation tubes.
7. Centrifuge with the rotor to spin 0.5 L–1 L tubes.
8. UV-VIS spectrophotometer with disposable cuvettes, both suitable for A_{600 nm} measurements.
9. –80 °C freezer.
10. Incubator shaker.

4.1.2 Lysis of Bacterial Cells

1. Pellets of bacteria.
2. Lysis buffer (20 mL per liter of culture): 50 mM Tris–HCl at pH 8.0, 150 mM NaCl, 10 mM imidazole, 18 U/mL Benzonase, 1 mM MgCl₂, 2 µL/mL CPI-his (Complete Protease Inhibitor, compatible with NiNTA).
3. 4× SDS sample buffer with reducing agent. To make 10 mL of 4× stock mix the following ingredients: 2.0 mL 1 M Tris–HCl pH 6.8, 0.8 g SDS, 4.0 mL 100% glycerol, 0.4 mL 14.7 M β-mercaptoethanol, 1.0 mL 0.5 M EDTA, 8 mg bromophenol Blue (*see Note 35*) and store in 1–2 mL aliquots at –20 °C.
4. 50 mL falcon tubes.
5. 11 mL centrifuge tubes.
6. Centrifuge.
7. Ice and styrofoam container.

4.1.3 Ultrasonication

1. Sonicator device Vibra-cell VCX (Sonics, CT, USA) with a large probe with replaceable tip (13 mm diameter) or equivalent and a microtip probe (3 mm diameter) or equivalent.
2. 1 L beaker.
3. Ice and styrofoam container.

4.1.4 Thermal Lysis

1. Water incubator (capable of reaching 95 °C).
2. Ice and styrofoam container.

4.1.5 NiNTA Affinity Chromatography (See Note 36)

1. NiNTA buffer A (50 mM Tris–HCl, 150 mM NaCl, 10 mM imidazole, pH 8).
2. NiNTA buffer B (50 mM Tris–HCl, 150 mM NaCl, 20 mM imidazole, pH 8).
3. NiNTA elution buffer (50 mM Tris–HCl, 150 mM NaCl, 250 mM imidazole, pH 8).
4. NiNTA agarose resin as 50% suspension.
5. Gravity column.
6. Column holder.
7. UV-VIS spectrophotometer.

4.1.6 Strep-Tag Affinity Chromatography (See Note 36)

1. FPLC system (with fraction collector, UV detector, and a way of loading the sample onto the column).
2. 5 mL StrepTrap column or equivalent Strep-trap binding resin (approx. 5 mL of resin per 1 L of bacterial culture).
3. 0.5 L MQ (filtered through 0.22 µm and degassed).
4. 0.5 L Strep buffer (50 mM Tris–HCl, 150 mM NaCl, 1 mM EDTA, pH 7.5; filtered through 0.22 µm and degassed).
5. 25 mL Strep elution buffer (50 mM Tris–HCl, 150 mM NaCl, 1 mM EDTA, 2.5 mM d-dethiobiotin, pH 7.5; filtered through 0.22 µm and degassed).
6. 25 mL HABA buffer (50 mM Tris–HCl, 150 mM NaCl, 1 mM EDTA, 1 mM HABA (2-[4'-hydroxy-benzeneazo] benzoic acid), pH 7.5; filtered through 0.22 µm and degassed).
7. 50 mL Strep regeneration buffer (0.5 M NaOH).

4.1.7 Ion-Exchange Chromatography (See Note 36)

1. 10 mL DEAE-Sepharose packed in 16/100 XK column (or equivalent) (see Note 37).
2. FPLC system (with fraction collector, UV detector, and a way of loading the sample onto the column, also must be able to make precise gradients).
3. 1 L IEX buffer A (50 mM Tris–HCl, 150 mM NaCl, pH 8.0).
4. 0.5 L IEX buffer B (50 mM Tris–HCl, 2 M NaCl, pH 8.0).
5. 0.5 L MQ.
6. 20% ethanol (see Note 38).

4.1.8 Size Exclusion Chromatography (See Note 36)

1. FPLC system.
2. Superdex 200 pg 26/600 column (or equivalent).
3. 1 L MQ.

4. 1 L SEC buffer (20 mM Tris–HCl, 150 mM NaCl, 10% (V/V) glycerol, pH 7.5) (*see Note 39* and **40**).
 5. Centricons (*see Note 41*).
 6. Low bind tubes.
- 4.1.9 TEV Protease Cleavage**
1. TEV protease.
 2. DTT (1 M stock) or TCEP (0.5 M stock) (*see Note 42*).
- 4.1.10 Reverse NiNTA**
1. NiNTA resin.
 2. NiNTA buffer A (50 mM Tris–HCl, 150 mM NaCl, 10 mM imidazole, pH 8).
 3. NiNTA agarose resin as 50% suspension.
 4. Gravity column.
 5. Column holder.
 6. UV-VIS spectrophotometer.

4.2 Methods

4.2.1 Bacterial Fermentation for Production of Proteins

Production of CCPo-based proteins is for practical and economic reasons optimized for bacterial expression (*see Note 43*). The bacteria are grown in LB media, and the expression of the protein is induced through lac operon by the addition of IPTG. Especially *E. coli* strain *NiCO* is suitable for our purposes since the majority of our constructs contain a His-tag.

1. Prepare a preculture by inoculating 100 mL (*see Note 44*) liquid LB media in a 0.5 L flask, supplemented with antibiotics (Kanamycin 50 µg/mL), with stock culture (*see Note 45*) containing desired expression vector and incubated at 37 °C, 160 rpm overnight.
2. Dilute the preculture to 0,1 OD_{600 nm} (*see Note 46*) in a larger (5 L) (*see Note 47*) Erlenmeyer flasks filled with 1 L of LB media supplemented with antibiotic Kanamycin in a final concentration of 50 µg/mL and leave growing at 37 °C before reaching a stationary phase where OD_{600 nm} values reach between 0.6 and 0.9 [[3–6](#), [10](#)] (*see Note 48*).
3. At OD_{600 nm} values of 0.6–0.9, induce (*see Note 49*) the cultures with 1 mM IPTG (*see Note 50*) and grow for 4 h at 30 °C and 160 rpm (*see Note 51*).
4. Pellet bacteria via centrifugation (5500 × g, 5 min, 5 °C) (*see Note 52*). Freeze the tubes with bacterial pellets at –20 °C or –80 °C for at least 4 h, preferably overnight (*see Note 53*).

4.2.2 Lysis of Bacterial Cells

Lysis of bacterial cells is performed either by sonication or by thermal lysis. Each has its benefits and drawbacks. Ultrasonication is a universal procedure for lysis and it works for different proteins. When using a standard single-probe sonicator, the main drawback

of this procedure is that it is time-consuming, especially when working with multiple proteins as only one can be sonicated at a time. Generally, it takes 15–30 min per protein.

Thermal lysis is a simple method that does not require a sonicator and is capable of removing the majority of bacterial proteins from the supernatant which makes subsequent affinity chromatography washing steps shorter. Most CCPO-based proteins are capable of spontaneous refolding after thermal denaturation which is necessary for thermal lysis. Thermal lysis is not an option when CCPO structures are fused to proteins, which do not refold upon thermal denaturation.

1. Perform a “thermal lysis test” to decide whether the protein is suitable for thermal lysis (*see Note 54*).
 - (i). Resuspend the pellet from 5 mL of bacterial culture in 1 mL of lysis buffer and transfer the suspension in a 1.5 mL tube.
 - (ii). Perform thermal lysis by incubating on a thermoblock for 15 min at 95 °C.
 - (iii). Spin the tube in a centrifuge at 12000 × g , 15 min, 5 °C.
 - (iv). Remove and store the supernatant (i.e., soluble fraction) and resuspend the pellet (i.e. insoluble fraction) in 300 µL of lysis buffer.
 - (v). Perform an SDS-PAGE gel loading both soluble and insoluble fractions (*see Note 55*) to determine if the protein is suitable for thermal lysis (Fig. 6) (*see Note 56*).
2. Perform a small batch sonication for “solubility test” to decide whether the protein can be lysed by sonication and subsequently isolated from soluble fraction.
 - (i). Resuspend the pellet from 10 mL of bacterial culture in 0.5 mL of lysis buffer and transfer the suspension in a 1.5 mL tube (*see Note 57*).
 - (ii). Position the tube with a sample on a stable rack (*see Note 58*) and adjust the microtip (3 mm) so that the end of the tip is 0.5 cm–1.5 cm above the tube’s bottom.
 - (iii). Sonicate for 1–2 cycles of 15 s of total pulse time, at intervals of 1 s pulse and 3 s pause (20–30% amplitude) which results in 1 min per cycle. Put the tube on ice to chill in between cycles.
 - (iv). Spin the tube in a centrifuge at 12000 × g , 15 min, 5 °C.
 - (v). Remove and store the supernatant (i.e., soluble fraction) and dilute it 5–10 times using lysis buffer and resuspend the pellet (i.e. insoluble fraction) in 300 µL of lysis buffer.

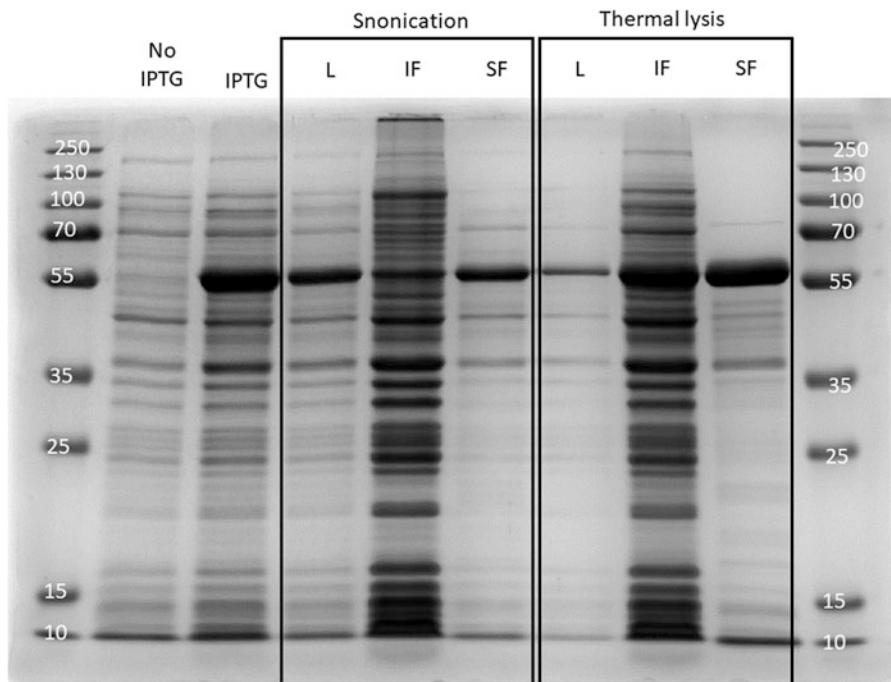


Fig. 6 SDS-PAGE protein TET12SN that was overexpressed in bacteria. The left side of the gel shows a comparison between lysate before expression induction (No IPTG) and 4 h after induction (IPTG). The right side of the gel shows lysate (L), insoluble fraction (IF), and soluble fraction (SF) of bacteria treated either with sonication or with thermal lysis

- (vi). Perform an SDS-PAGE gel loading both soluble and insoluble fractions (*see Notes 55 and 59*) to determine if the protein is suitable for sonication and isolation from soluble fraction (Fig. 6).
3. Add an appropriate amount of buffer (15 mL per 1 L of LB media used for the fermentation) to the still frozen bacterial pellet and resuspended on ice (*see Note 60*).
4. Proceed to lysate cells either by ultrasonication or thermal lysis, according to the properties of the protein.

4.2.3 Ultrasonication

1. Put a 50 mL falcon tube with lysate on wet ice (*see Note 61*) inside a 1 L beaker and adjust the probe (large probe (13 mm)) of the sonicator so half of it is submerged in the lysate (*see Note 62*).
2. Sonicate for 2–10 cycles (*see Note 63*) of 1 min of total pulse time, at intervals of 1 s pulse and 3 s pause (55–65% amplitude) which results in 4 min per cycle [3–6].
3. Aliquot the lysate in centrifugation tubes and spin them at 5 °C on 12,000 × g for 20–30 min.). As CCPO are normally present in the supernatant, save it and discard the pellets (Fig. 6).

4.2.4 Thermal Lysis

1. Loosely screw the cap (*see Note 64*) and place the tube with resuspended pellets on a polyfoam floating rack (or some other heat-resistant floating device) in a water incubator at 95 °C for 15 min.
2. Put the tube on ice to chill. After the tube is cold to the touch, add 18 U/mL Benzonase.
3. Aliquot the lysate in centrifugation tubes and spin them at 12000 × φ for 15–30 min on 5 °C. Save the supernatant and discard the pellets (Fig. 6).

4.2.5 NiNTA Affinity Chromatography

Gravity NiNTA presents the main method for CCPO isolation as it is fast, robust, reliable, and cost-effective. Moreover, it can easily be performed in parallel for many samples at the same time. However, gravity NiNTA can be substituted for NiNTA ran on FPLC.

1. Secure the column for gravity NiNTA in such a way that larger amounts of buffer can be flown through and collected (*see Note 65*). Transfer 6 mL of 50% NiNTA resin per 1 L of bacterial culture into the gravity column and rinse the resin with approx. 20 mL of MQ and 40 mL of NiNTA buffer A (*see Note 66*).
2. Filter the soluble fraction through a 0.45 µm syringe filter into a 50 mL falcon tube. Carefully transfer filtered soluble fraction onto the washed NiNTA column (*see Note 67*) and let it flow through the column (*see Note 68*). After all of the soluble fraction has flown through the resin, start washing with NiNTA buffer A (*see Notes 69 and 70*). Periodically collect wash fractions and measure $A_{280\text{ nm}}$. When $A_{280\text{ nm}}$ is below 0.1, switch to washing the resin with NiNTA buffer B until $A_{280\text{ nm}}$ is below 0.1 again (*see Note 71*).
3. Remove as much residual NiNTA buffer B from the column as possible by pushing through or sucking out the remainder of the liquid. Start with elution of the protein by adding 1 mL of elution buffer at the time and collecting each fraction in 2 mL tubes (*see Notes 72 and 73*). Using UV-VIS spectroscopy (or alternative concentration determination methods), identify fractions that contain the highest amount of protein. Merge the fractions depending on the next step of the isolation.

4.2.6 Strep-Tag Affinity Chromatography

For CCPO, that do not respond well to gravity NiNTA or when His-tag is not desired, Strep-tag can be used. CCPO structures, prone to degradation, respond well to double-tagging with both His- and Strep-tag.

1. Depending on the tags on the CCPO, the sample should be prepared differently before loading onto the column:
 - (i). *CCPO tagged only with Strep-tag*: after the soluble fraction is obtained, filter it through a 0.45 µm syringe filter. Volume permitting, dilute the sample with Strep buffer till 50 mL (*see Notes 74 and 75*).
 - (ii). *CCPO tagged with both His- and Strep-tag*: after elution from NiNTA, dilute the sample with Strep buffer A at least 5 times and filter it through 0.45 µm syringe filter (*see Note 76*).
2. Using an FPLC system, prepare the StrepTrap HP column for isolation following the manufacturer's recommendations. In short, after connecting the appropriate number of StrepTrap HP columns (*see Note 77*), wash out 20% EtOH with filtered and degassed MQ at up to 2.5 mL/min for at least 1 column volume (CV), followed by conditioning in Strep buffer at up to 5 mL/min for at least 2 CV.
3. Load the sample onto the StrepTrap column using a superloop or pump with the flow of 1.0–1.5 mL/min (*see Notes 78 and 79*). After the loading is complete, begin washing with Strep buffer A at 2.5–3.5 mL/min until UV signal at 280 nm stabilizes (*see Notes 80 and 81*).
4. Start the elution with Strep elution buffer and the flow of 1 mL/min. If possible, put the column in an up-flow position. Collect elution fractions of 1–2 mL. Based on the UV signal, merge appropriate fractions (*see Notes 82 and 83*).
5. After the elution, the column should be regenerated according to the manufacturer's recommendations. In short, wash the columns with 4 CV of MQ, followed by 1 CV of Strep regeneration buffer and at least 4 CV of MQ again (*see Note: 84*). After that, the columns are ready for the next isolation or to be washed and stored in 20% EtOH.

4.2.7 Ion-Exchange Chromatography

When double-tagging is not desired or does not offer adequate purity, an ion-exchange chromatography step could be attempted (*see Note 85*).

1. Prepare the IEX column, according to the manufacturer's recommendations. In short, wash the 20% ethanol out of the column with at least 1 CV of MQ (*see Note 86*), then equilibrate the column with either 5 CV of IEX buffer A or by 0.5 CV of IEX buffer B, followed by IEX buffer A (*see Note 87*).
2. Dilute the sample eluted from NiNTA at least 5× with IEX buffer A and load onto the column (*see Note 88*). Wash with IEX buffer A until UV signal stabilizes.

3. Perform the elution with a gradient of 0% to 15% of IEX buffer B in 6 CV, using 1 mL/min and, if available, a column in reverse flow configuration. Collect fractions of lower volumes, such as 1–2 mL per fraction. Analyze obtained fractions with SDS-PAGE to determine which fractions are suitable for further isolation steps (*see Note 89*).
4. Regenerate the resin with IEX buffer B and 1 M NaOH following the manufacturer's recommendations. In short, wash the column with 2 CV of IEX buffer B, followed by 2 CV of MQ and up to 5 CV of 1 M NaOH (*see Note 90*). Before storage, wash the resin with another 4 CV of MQ and 4 CV of 20% ethanol (*see Note 38*).

4.2.8 Size Exclusion Chromatography

Size exclusion chromatography is a crucial step in the isolation of CCPO as it separates higher aggregates from monomers.

1. Condition the Superdex 200 pg 26/600 column (*see Note 91*) in SEC buffer according to the manufacturer's recommendations. In short, after connecting the column stored in 20% EtOH, wash with at least 1 CV of MQ, followed by 1.5 CV of SEC buffer before injecting the sample.
2. Inject up to 12 mL of sample onto the conditioned column using the superloop (*see Notes 92 and 93*) and perform the separation at 2.6 mL/min. Most CCPO, regardless of shape and size, have a distinct elution profile, which comprises a small fraction of higher oligomers, distinct trimeric, and major monomeric fraction (Fig. 7) (*see Note 94*).
3. Based on elution time, calibration curve, and potential additional analysis (if previous experience is not available), such as SDS-PAGE (*see Note 95*), DLS, and/or SEC-MALS, merge the appropriate fractions of interest (*see Note 96*).
4. Prepare centricons (centrifugal filtration devices) with appropriately sized pores by first rinsing the membrane with MQ. Load the filtered MQ on the top portion of the centricon and spin for 1–3 min at $1000 \times g$ below the maximum rating. Discard the flow-through and the remaining MQ (*see Note 97*). Load the filtered sample on the top portion of the centricon and spin at the maximum rating for 5, 10, or 20 min for centricons with 30, 10, or 3-kDa cut-off, respectively. After each round of concentration, inspect the concentrate for visible aggregation (*see Note 98*) and mix by pipetting (*see Note 99*). Moreover, periodically check the absorbance of the flow-through for protein leakage (*see Note 100*). After the desired concentration has been reached, aliquot the sample in low bind tubes, shock freeze in liquid nitrogen, and store at -80°C (*see Note 101*).

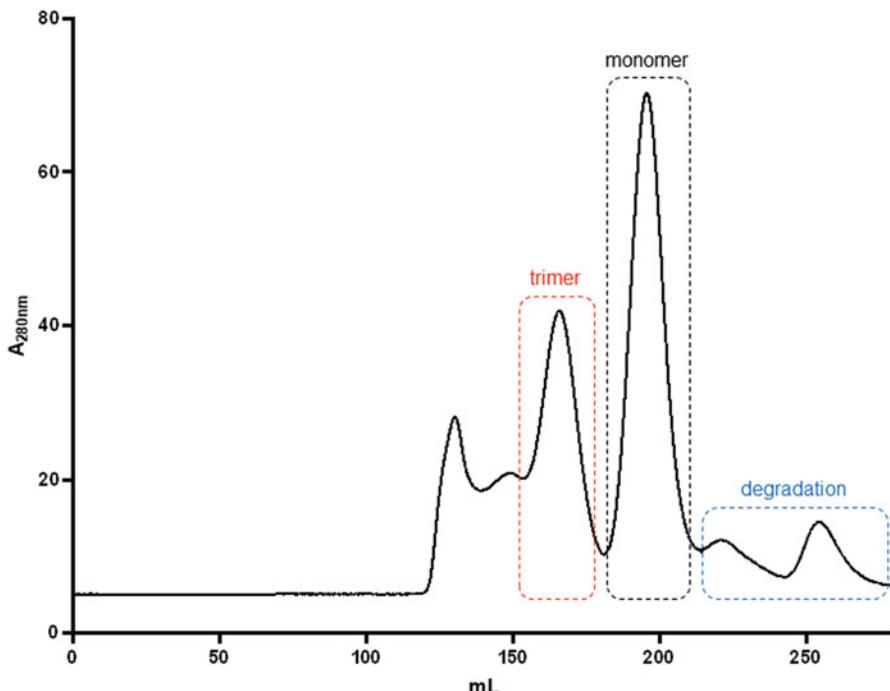


Fig. 7 An example of UV trace of a tetrahedral CCPo, separated with Superdex 200 pg 26/600. Monomeric fraction is marked with the black box, while distinct trimeric fraction is presented in a red box. A smaller fraction of degradation products (marked in a blue box) is rarely present. A similar shape of the UV trace is observed across most CCPo, regardless of shape, topology, and permutation

4.2.9 TEV Protease Cleavage and Reverse NiNTA (See Note 102)

Polypeptide cleavage with TEV protease is mostly used for the removal of affinity tags from CCPo after the purification but can also be utilized for mediated assembly of multi-chain CCPo structures [4].

1. TEV protease cleavage:

- Add reducing agent (DTT or TCEP) to the target protein solution to the final concentration of 1 mM (see Note 103).
- Add TEV protease to the protein solution and incubate overnight at 4 °C (see Note 104). The solution should have 50–200× molar excess of TEV recognition site to TEV protease (see Note 105).
- Analyze the reaction with SDS-PAGE and proceed with reverse NiNTA for the removal of uncleaved products, cleaved protein tags, and TEV protease.

2. Reverse NiNTA (see Note 106)

- Prepare the column for NiNTA, as described in subheading Sect. 4.2.5., step 1 (see Note 107).

- (ii). After thorough column equilibration with NiNTA buffer A, carefully transfer reaction solution on the column (*see Note 108*).
- (iii). Collect the flow-through in a falcon tube and start washing the column with NiNTA buffer A (*see Note 109*).
- (iv). Collect wash fractions of 0.5–1.5 mL in 1.5–2 mL tubes and measure protein concentration with $A_{280\text{ nm}}$.
- (v). Merge the desired fractions with the flow-through in a falcon tube and analyze the separation on SDS-PAGE.

5 Characterization

For analysis of CCPO samples, care should be taken to thaw the sample as fast as possible and to filter all samples with 0.1 μm centrifugal filters. Thawing should be performed using one of three methods: (i) under running water, (ii) put under the latex glove to be thawed by the warmth of the hand (only usable for smaller volumes for fear of frostbite), or (iii) put in the trousers pocket. However, if a fresh sample is available, it should be used preferentially. After analysis, the remaining sample could be refrozen, but it is not recommended, as the trimeric fraction increases with each subsequent freezing and thawing cycle.

5.1 Materials

5.1.1 Circular Dichroism Spectroscopy

1. Circular dichroism instrument ChiraScan (Applied Photophysics, UK), equipped with a Peltier thermal control block (Mellcor, NJ) or equivalent.
2. 1 mm quartz cuvette (Hellma, Germany).
3. SEC buffer (for blank): 20 mM Tris–HCl pH 7.5, 150 mM NaCl, 10% v/v Glycerol, or equivalent (*see Note 110*).

5.1.2 Size Exclusion Chromatography Coupled with Multiangle Light Scattering (SEC-MALS) (See Note 36)

1. HPLC system with an autoinjector, coupled to UV, MALS, and RI detectors (*see Note 111*). An example of the system is described in Note 138.
2. Superdex 200 Increase 10/300 column or equivalent (*see Note 112*).
3. 0.5 L MQ.
4. 1 L SEC-MALS buffer (20 mM Tris–HCl, 150 mM NaCl, pH 7.5 (*see Note 113*)).
5. HPLC control software, such as Empower (Waters).
6. Analysis software, such as Astra (Wyatt).

5.1.3 Small-Angle X-Ray Scattering (SAXS)

1. X-ray source (either home source, such as SAXSpoint 5.0, or synchrotron) (*see Note 114*).

2. CCPO sample at high concentration (150 µL of at least 5 mg/mL) (*see Notes 115 and 116*).
3. Matching buffer (*see Note 117*).
4. ATSAS software package [11].

5.2 Methods

5.2.1 Circular Dichroism Spectroscopy

Measuring Far UV spectra of CD is an important step toward protein structure determination since high signal at specific wavelengths usually indicates the presence of intramolecular (or intermolecular) coiled-coil interactions that are characteristic of CCPO-based structures. Measuring CD signal while performing temperature denaturation also offers information about the stability of the given protein or complex.

1. Turn on the CD machine components following manufacturer's recommendations. For described setup, the order should be: (i) turn on the supply of pressurized nitrogen (*see Note 118*), (ii) water bath for counter cooling the Peltier thermal control block, (iii) thermal control block, (iv) control computer and finally Xenon lamp (*see Note 119*).
2. Clean the cuvette with MQ and 70% ethanol using a pressure pump (or by hand by rinsing the cuvette with MQ and 70% ethanol) and dry the cuvette completely with nitrogen gas. Do this step every time the contents of the cuvette are changed (*see Note 120*).
3. Dilute the purified protein sample to appropriate concentration (*see Note 121*) with the final volume of 270 to 300 µL (*see Note 122*) using SEC buffer (*see Note 123*) and transfer the sample into the cuvette. Finally insert a clean temperature probe into the opening of the cuvette (*see Note 124*).
4. Set the parameters of the measurement in the software (e.g., Chirascan (Applied Photophysics, Leatherhead, UK)): Wavelength of far-UV in the range of 200 to 280 nm, with a step size of 1 nm, a bandwidth of 1 nm, and an integration time of 1 s [5, 6, 12] and measure background buffer at 20 °C.
5. Measure the far-UV spectra of samples at three different temperatures and in-between perform temperature denaturation to assess the stability of the sample. Perform three technical replicates (*see Note 125*) for each sample at a specific temperature (*see Note 126*). The recommended measurement steps:
 - (i). Measure the far-UV spectra at 20 °C (*see Note 127*).
 - (ii). Perform temperature denaturation experiment by heating the samples from 4 to 95 °C (*see Notes 128 and 129*) at a rate of 1 °C/min while measuring the CD signal at 222 nm in triplicates at 1 °C increment (*see Note 130*). In additional settings of temperature ramping, it is important to

set end-of-ramp behavior so that the end temperature at the end of denaturation is sustained (*see Note 131*).

- (iii). Measure the far-UV spectra at 95 °C immediately after thermal denaturation (*see Note 132*).
- (iv). Rapidly cool the sample by setting the temperature of the thermal block back to 20 °C and measure the far-UV spectra (*see Note 133*). The remaining sample should be taken for further SEC-MALS analysis for monodispersity determination.

6. Analyzing and fitting the data

- (i). Make an average of the technical replicates of far-UV spectra measurements and transform measurements acquired by CD into mean residue ellipticity (MRE) using the following equation.

$$\text{MRE} = \frac{A \times M_w}{\text{Res} \times d \times c \times 10}$$

where A is raw data from CD (Deg), Mw is the molecular mass of the protein (g/mol), Res is the number of peptide bonds in the protein (*see Note 134*), d is path length (cm) through the cuvette, and c is the concentration of the measured protein (g/mL). Express MRE data on the ordinate in unit $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ corresponding to wavelengths (nm) on the abscissa. An example of CD spectra is shown in Fig. 8a.

- (ii). Helical content can be derived from the following equation:

$$\text{Helicalcontent(\%)} = \frac{\text{MRE}_{222\text{nm}}}{\text{MRE}_{222\text{nm}}^H \times \left(1 - \frac{2,57}{n}\right)}$$

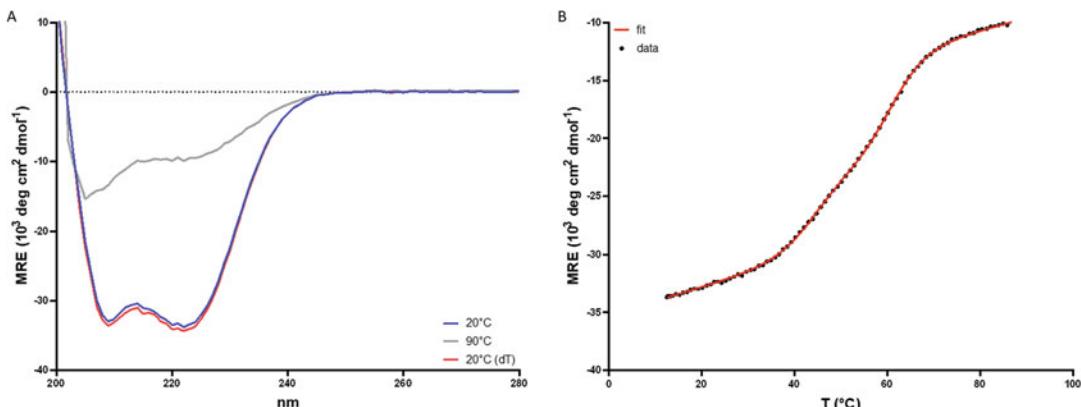


Fig. 8 Example of a CD analysis of CCPo. (a) CD spectra should show a highly helical secondary structure, with MRE between 30 and 35,103 deg cm² dmol⁻¹. (b) Thermal denaturation is usually cooperative, with T_m varying with the T_m of building blocks used

where n is the length of the amino acid sequence, MRE_{222} average mean residue ellipticity at 222 nm, and MRE_{222}^H is the theoretical mean residue ellipticity of an infinitely long helix ($-39,500 \text{ deg cm}^2 \text{ dmol}^{-1}$) [3, 4, 6].

- (iii). Determine the melting temperature of a protein by fitting the temperature denaturation curve to a sigmoidal Boltzmann function. An example of thermal denaturation is shown in Fig. 8b (see Note 135).

5.2.2 Size Exclusion Chromatography Coupled with Multiangle Light Scattering (SEC-MALS)

Size exclusion chromatography, coupled with multi-angle light scattering, UV, and RI detectors (SEC-MALS), is one of the most important routine analyses of CCPs. As SEC-MALS measures both Mw and oligomeric states, it is an essential and quick analysis.

1. Condition the Superdex 200 Increase 10/300 column according to manufacturer's recommendations (see Note 136). In short, wash the 20% EtOH out of the column with 1 CV of MQ, followed by conditioning with SEC-MALS buffer for at least 2 CVs (see Note 137). After the column has been equilibrated, calibrate all detectors, if applicable, and verify that the system operates perfectly by injecting a protein standard.
2. Using 0.1 μm centrifugal filters, filter up to 200 μL of sample, at concentrations of 0.2–0.6 mg/mL. Transfer the samples into the HPLC vial insert, and put the vial with the insert in the autoloader (see Note 138). Input the sample info both in data collection software (e.g., Astra (Wyatt)) and HPLC control software (e.g., Empower (Waters)) (see Note 139) and inject the samples.
3. After the injections, analyze the data following recommendations for the analysis software of choice. Characteristic SEC-MALS profile before and after thermal denaturation is shown in Fig. 9 (see Note 140).
4. Take care to properly clean the column when closing it (see Note 141), potentially using more aggressive cleaning agents when the back-pressure is high or resolution drops.

5.2.3 Small-Angle X-Ray Scattering (SAXS)

As CCPs have higher flexibility compared to most globular protein folds, high-resolution structural characterization still offers some challenges. Therefore, for routine determination of lower resolution structures of CCPs, small-angle X-ray scattering (SAXS) is used. Depending on the nature of the CCP investigated, either batch or SEC-SAXS are used. We will mostly present data collection from the point of a smaller home source, while data analysis is applicable regardless of where data were collected. As home-source SAXS can greatly differ from one another, provided parameters are intended to be used as a more general guideline and a starting point for potential optimization.

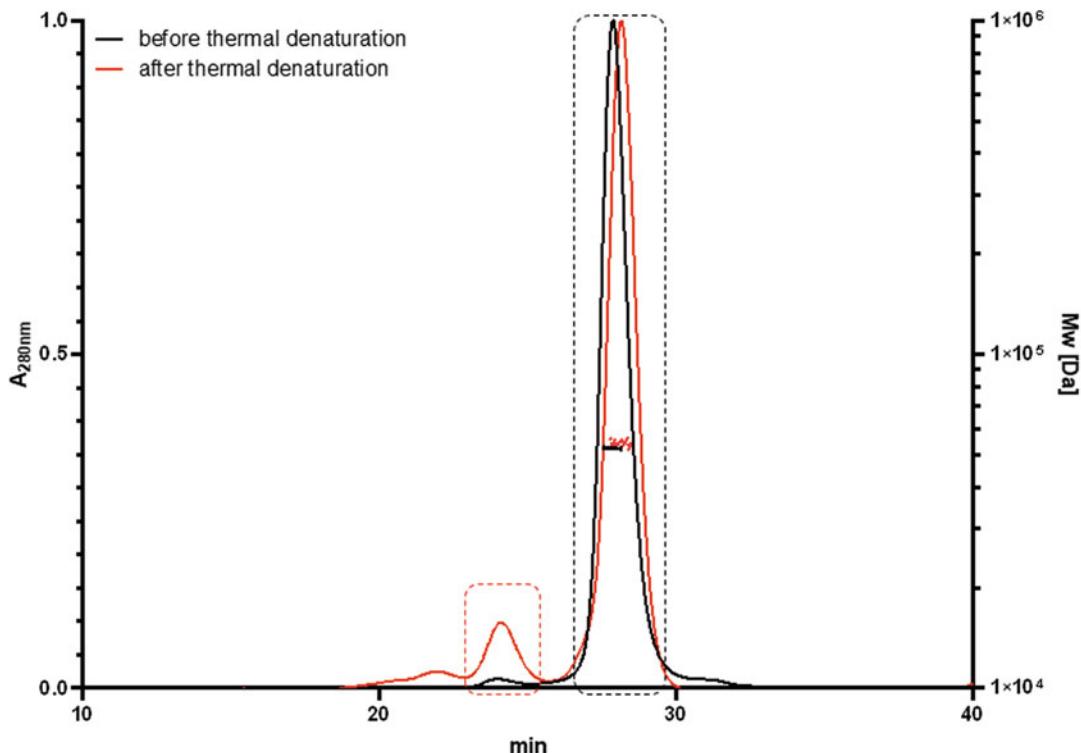


Fig. 9 Characteristic SEC-MALS profile of CCPo from SEC-MALS before and after thermal denaturation, with a monomeric fraction in a black box. Note an increase in trimeric fraction, as a result of thermal refolding (in red box)

1. Data collection:

- Prepare home source for data collection following manufacturer's recommendations and local health and safety rules. SAXS deals with a source of powerful X-rays and strong vacuum, so treat it accordingly.
- If applicable, cool down the sample holding chamber in the autosampler and sample holder. If samples will be loaded via autosampler, rinse and extensively dry the loop, delivery tube and capillary with freshly prepared MQ (see Note 142).
- Filter the samples with centrifugal 0.1 μm filters (see Note 143) and prepare desired concentrations. Transfer the samples into the form, that is suitable for measurements—e.g., for described setup, 60 μL of each sample concentration with its matching buffer and loaded onto the conical bottom 96-well microtiter plates (see Note 144). Align the beam and set up the data collection sequence—e.g., for described setup, 20 μL of each sample is loaded with the ASX autoloader (kept at 4 °C) into the

X-ray beam, with sample holder kept at 10 °C. 2D data are collected in 7 frames of 30 min each with SSD of 900 mm, with transmittance measurement enabled (*see Note 145*).

2. Data analysis (*see Note 146*).

- (i). Convert 2D raw data to 1D raw curves using the software, supplied with the home source. Sample sequence of steps for analysis in SAXS analysis are: Zero point by moments 2D, Masking 2D, Standard operations 2D (transmittance mode), Q transformation, Data reduction 2D, and Export ATSAS.
- (ii). Exported 1D SAXS curves should be checked for radiation damage or potential movement of the sample and potentially discarded. The remaining 1D SAXS curves, belonging to the same sample concentration, are averaged in ATSAS, using Tools → Average in Primus. Save averaged 1D SAXS curves as new data files by Right-click → Saving as (*see Note 147*).
- (iii). Subtract matching buffers from sample concentration importing averaged 1D SAXS curves into Primus. Select the curve, corresponding to one sample concentration and matching buffer (*see Note 148*) and subtract with Tools → Subtract (*see Note 149*).
- (iv). Subtracted 1D SAXS curves can be normalized by correction factor I_{scale} in Primus (*see Note 150*). To normalize, I_{scale} factor can be calculated (*see Note 151*):

$$I_{scalefactor} = \frac{1}{\delta}$$

where δ presents mass concentration in mg/mL

- (v). Alternatively, if M_w calculation from I_0 will not be attempted, crop the subtracted 1D curves with Primus, using From/To columns, avoiding left- and right-most portions of the curve (*see Note 152*). Scale all subtracted 1D SAXS curves, while only central portions are visible (*see Note 153*). To confirm that normalization or scaling was successful, show the remainder of the curves by changing the numbers in From/To to minimum and maximum, respectively. Compare the curves (*see Note 154*).
- (vi). Normalized or scaled SAXS curves should be merged before further analysis. The main idea of merging curves is to make a new composite curve, that is composed of the best regions of different concentrations. With a bit of generalization, SAXS curves, obtained from higher concentrations, normally offer better signal and lower noise at higher angles, but lower angles can be a bit questionable

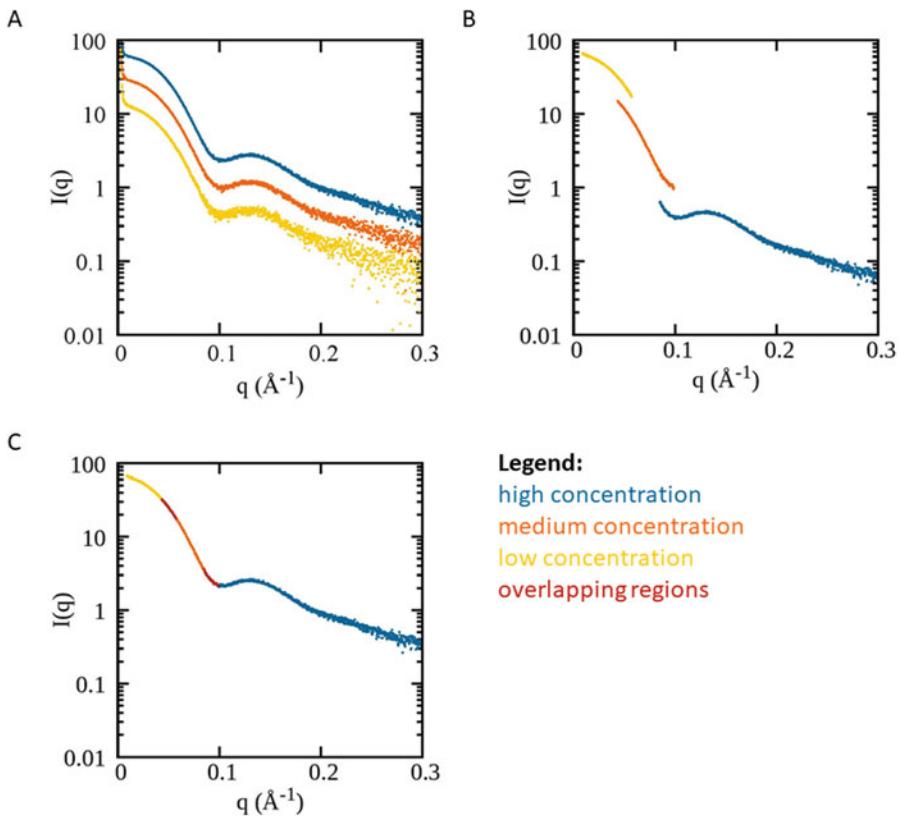


Fig. 10 An example of SAXS curve merging. **(a)** Subtracted SAXS curves are inspected for regions to be used in the final merged SAXS curve, and **(b)** appropriate regions are taken, accounting for sufficient overhangs. Chosen portions of the separate concentrations are then scaled and merged **(c)**

due to oligomerization; while the reverse is true for lower concentrations (Fig. 10a). To merge normalized or scaled SAXS curves, From/To parameters in Primus should be used in such a way that there are at least 50 points of overlap between the curves to be merged (Fig. 10b). Before merging, the curves should be scaled again (in Primus, Tools → Scale) and one should visually confirm that the scaling was correct (Fig. 10c). At this step, merging can simply be achieved by Tools → Merge (*see Note 155*). It is best practice to make several merged SAXS curves, combining more than two normalized or scaled SAXS curves.

- (vii). Merged SAXS curves need to be trimmed—if necessary, remove few points at the lowest angles (*see Note 156*) and determine a threshold for higher angles (*see Note 157*).

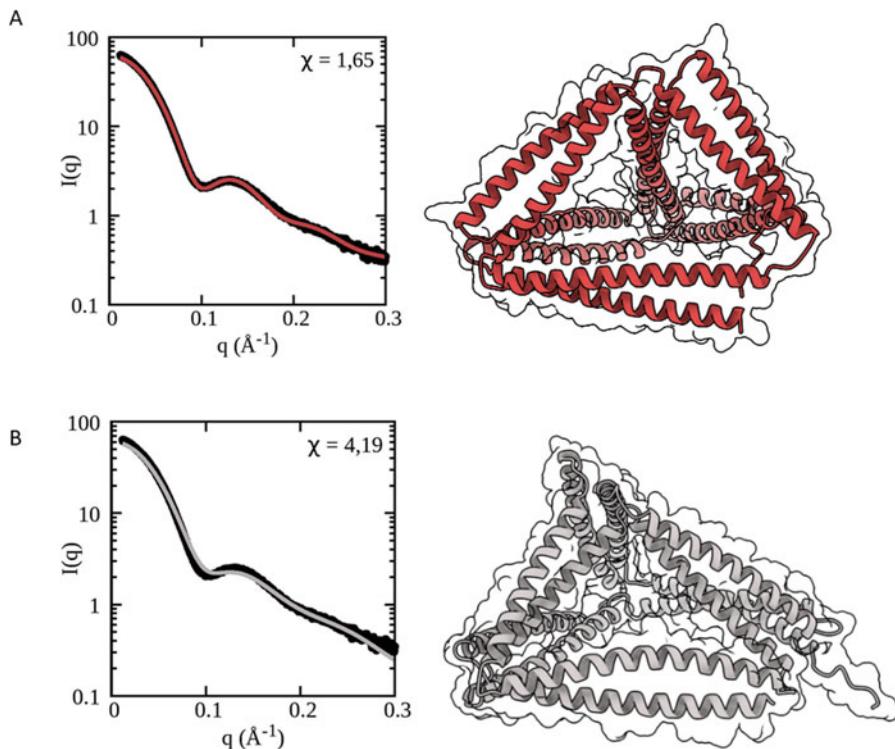


Fig. 11 Example fitting of a good (a) and bad (b) fit. Do note the decrease in the intensity of the distinct minimum and maximum at 0.1 and 0.15 \AA^{-1} , respectively

- (viii). Fit the previously generated models using one of many available SAXS fitting solutions. For CCPD, PEPSI-SAXS repeatedly offers best fits, closely followed by FoXS (*see Note 158*). When checking the models, look for characteristic empty space in the middle of the protein and with straight sides—should these be absent, it might indicate that the measurement or design has not been successful. An example of a good and bad fit is shown in Fig. 11.
- (ix). As a part of a more comprehensive analysis, R_g , D_{\max} , $P(r)$ and Kratky plots should also be obtained. R_g can be estimated using Primus tool Analysis → Radius of Gyration (*see Note 159*) and the same window also lists I_0 value (*see Note 160*). D_{\max} and $P(r)$ function can also be calculated in Primus Analysis → Distance Distribution (*see Note 161*).
- (x). Ab initio reconstruction should also be attempted. ATSAS software package offers a command-line tool—dammin (*see Note 162*). Dammin takes $P(r)$ function as an input (exported from Primus when determining D_{\max}) and is run from the command line using the command:

```
dammin
```

It is advisable to first generate at least 10 reconstructions with default settings. Outputs (files ending with `-1.pdb`) should be copied to a new folder and further processed and averaged with the following list of commands:

```
damsel  
damsup  
damaver  
damfilt
```

To check, if the best fitting model fits in the generated envelope, both should be opened in Chimera and the following command ran:

```
molmap #0 15
```

select mesh in the pop-up window.

- (xi). To fit the model into the generated envelope, open a tool Tools → Volume data → Fit in map. Choose appropriate models and click fit.

6 Notes

1. Detailed instructions on CoCoPOD installation can also be found in provided GitHub repository. There is also a demonstrative video tutorial covering all the steps.
2. On Windows replace export with set.
3. If there are problems with Modeller installation via conda, make sure to use a standalone installer (https://salilab.org/modeller/download_installation.html). Also, ensure to enter the Modeller license key, which is available free of charge for academic non-profit institutions (<https://salilab.org/modeller/registration.html>).
4. Described here is a simple design process of a tetrahedral CCPO. A tutorial is presented in a form of a python Notebook with all the necessary commands, and it also has an accompanying video tutorial on Github repository. Example files can also be found in the CoCoPOD folder (cocopod/examples/TET).
5. Polyhedral shape can also be specified manually by listing polygon faces.

6. Polyhedral topology defines the order of the coiled-coil forming peptide segments in the polypeptide chain. We refer to topology also as double Eurelian trail since the polypeptide chain needs to traverse twice across every edge of a polyhedron, returning to the starting vertex. Selection of starting vertex signifies the circular permutation.
7. To assess different topologies and circular permutations, topological contact order (TCO) and standard deviation of distances between pairs (stdTCO) were introduced. TCO is defined as the average number of segments between each peptide pair and stdTCO calculates the standard deviation of the distances between pairs.
8. More stable peptide pairs should be at the protein termini to stabilize the whole structure. Stable peptide pairs should also be chosen for peptide segments that are far apart in polypeptide sequence since stronger interactions help compensating for entropic penalties bringing these peptide pairs together.
9. Coiled-coil interactions and orthogonality are determined with amino acids at the positions a, d, e, and g. While amino acids at positions b, c, and f do not affect peptide specificity, they can alter coiled-coil stability and solubility [12]. CoCoPOD provides several variations of peptide pairs with different properties, which differ in residues at b, c, and f positions (e.g., P3, P3S, P3SN). S means soluble and SN soluble and negative. Also, coiled coils can be of different lengths (e.g., P3 with 4 heptads, APH with 5 heptads). While this does not present a major problem, polypeptide chains with peptide pairs of the same length give more regular shaped polyhedrons.
10. Stability is not particularly affected by the choice of linkers, for as long as they are long enough. But it can affect flexibility, folding kinetics, and protein solubility.
11. make_config.py file can also be prepared manually, skipping all described steps and going straight to the model building. This is useful when one already has selected topology and only needs to do minor modifications in the protein sequence.
12. Rule of thumb is to create around 50–100 protein models for simpler polyhedral shapes, while more demanding shapes need several thousand models. A number of created models is calculated by multiplying N_fold and N_homology variables.
13. Currently Modeller folds polyhedron peptide pairs in the order based on their stability. Based on recent study [6], peptide pairs with lower TCO form coiled coils first. This will be implemented in the updated CoCoPOD.
14. Use one of the cloning strains, such as Top10 and DH5a.

15. Due to the way BsaI cuts DNA, sequences have to be carefully positioned as not to create incompatible DNA overhangs or frameshifts.
16. The sequence of the vector backbone is not important, as long as it has ampicillin resistance, bacterial ori site, and absence of BsaI recognition site.
17. Alternatively, purified PCR fragments of basic building blocks either from DNA gBlocks or vector (sequence just from left BsaI to right BsaI with a bit of flanking sequence on universal sequences on 5' and 3' for PCR amplification) or just straight gBlocks have also been used with great success.
18. pET-41a with kanamycin resistance was used for this, with ccdb sequence and flanking BsaI restriction sites inserted between NdeI and XhoI restriction sites.
19. When amplifying this DNA vector, use a strain of *E. coli* that contains the ccdb anti-toxin.
20. The following protocol is based on an assumption that GGA building blocks are in a vector with ampicillin resistance, while the accepting vector is modified pET-41a with kanamycin resistance, prepared as described in Note 18.
21. When assembling 12 segments, average starting vector concentrations should be at least 168 ng/ μ L for a final concentration of 14 ng/ μ L of each of the 12 donor vectors to be reachable.
22. Concentration of accepting vector should be from 50 to 100 ng/mL, but as long as enough accepting vector is provided, the concentration does not greatly affect the efficiency of the reaction.
23. Dissolve the white precipitate by incubating and mixing it at 37 °C.
24. If available, NEB reagents should be used, as BsaI-HF v2 (NEB) retains most of its activity in 10 \times T4 ligase buffer (NEB). Changing any or all reagents to different suppliers might greatly decrease the efficiency of the reaction. Moreover, the success of GGA is highly dependent on high enzyme activity and failed GGA attempts normally come down to the low activity of enzymes or technical errors.
25. After transferring the GGA reaction, rinse the PCR tube with 10–20 μ L of sterile MQ to transfer as much DNA as possible, while keeping the cells on the ice the whole time.
26. Use agar plates with the antibiotic, matching the resistance on the accepting vector.
27. Anywhere between 5 and 100 bacterial colonies per plate are normal. The efficiency of GGA can vary greatly, depending on

enzyme activities, the freshness of the buffers, and bacterial transformation efficiency.

28. 6–12 bacterial colonies per GGA should be screened.
29. Many different transfer methods and implements are in use, such as sterile tips, toothpicks, and bacterial loops.
30. Properly label bacterial cultures.
31. Depending on the primers used, observed fragment size might be 100–300 bp bigger than expected, due to the added flanking regions. Special care should be given to fragments close to the expected size, as sometimes just one building block can be missing, or an additional one can be inserted. If available, previously positive CCP DNA segments of the same or similar sizes can be used as a positive control. Moreover, it is advisable to use universal primers (primers that bind to, e.g., T7 promoter and terminator) to distinguish between GGA negative and technical negative colonies.
32. PCR primers, annealing only on coiled-coil regions, should be avoided due to the high similarity of DNA sequence. When designing primers, if possible, position them so they cover two coiled-coils. Moreover, DMSO should be added to all PCR reactions as a precaution, and annealing temperature should be increased above T_m if needed.
33. Gibson assembly and QuickChange can be used, but special care should be taken when choosing overlapping regions. If some additional cloning or modifications have to be done with a certain DNA construct, it is best to put in desired restriction sites or design unique DNA sequences from the beginning.
34. High imidazole concentrations can adversely affect the stability and monodispersity of CCP.
35. Read safety instructions before handling chemicals such as β -mercaptoethanol and SDS.
36. All buffers, MQ, and 20% EtOH should be filtered through 0.22 μm filter and degassed.
37. Other anion exchange resins can be used, but offer varying results.
38. Check that the storage solution is as per the manufacturer's recommendations.
39. Weigh the appropriate mass of the glycerol instead of measuring the volume. Should the CCP contain any Cys, 0.5–1 mM TCEP might be added to the buffer.
40. Glycerol concentration can vary. 10% (V/V) is commonly used for cryoprotection, although as low as 3% has been successfully used. If the sample will not be frozen, glycerol can be omitted altogether.

41. For CCPO of 40–80 kDa, 10 kDa cut-off centricons are recommended, while for >80 kDa, 30 kDa cut-off can be used.
42. pH of TCEP stock should be adjusted to 7.5.
43. Before large-scale production of a new protein for the first time, it is advisable to perform production optimization trials to assess the expression of the protein of interest in different conditions to find the optimal balance between high expression, solubility, and level of degradation. A scheme for these trials involves a parallel variation of different parameters, which commonly are: temperature, the concentration of inducer, duration of induction, media, or additives. In this chapter, we will not be covering optimization of expression and solubility, so please refer to the literature [13].
44. 35–50 mL of overnight culture is usually enough to initiate production of protein in 1 L of culture.
45. In random cases of particular proteins, production has proven to be inefficient when growing bacteria from a stock culture, but was efficient when grown from a single LB plate colony—as a consequence, starting the inoculum from a single colony, growing from a fresh transformation (not older than 1 week) is preferable.
46. The optical density of an overnight culture is usually between 2 and 3.5. If the OD of an overnight culture is $OD < 2$, it is usually due to the leaky expression of a protein. This might be remedied with the addition of 1% glucose into the preculture the night before, to suppress the T7 promoter.
47. The amount of media in which to produce protein in principle correlates to the yield of purified protein. Generally, if the CCPO protein is soluble, one can expect 2–10 mg of purified protein per 1 L of media, which yields enough material from 2 flasks to successfully perform basic protein characterization. However, note that this is a rough estimation and that the rate of protein production can vary enormously even between similar proteins, so the appropriate amount of media in which to produce should be determined empirically for each protein, but usually ranges from 2 to 4 L.
48. To reach the end of log phase of bacterial growth in 1 L of media, it usually takes 1.5–3 h of incubation. If the culture grows extremely slowly (4–6 h to reach stationary phase), adding 1% glucose usually helps the culture to reach $OD_{600\text{nm}}$ of 0.6–0.9 faster.
49. To confirm that there was no leaking of protein production before the induction, 1 mL of culture might be taken for later analysis (Fig. 6).
50. Increasing the final concentration of IPTG above 0.5 mM does not have any effects on the yield of the isolation.

51. If the protein is poorly soluble, overnight incubation at either 24 or 16 °C could be considered. This usually results in a lower rate of overall production, but it might help achieve a higher yield of protein due to a bigger amount of protein in the soluble fraction.
52. If the particular design is produced for the first time, take 1 mL of bacterial culture before harvesting in a 1.5 mL Eppendorf tube and label it “after induction.” Perform an SDS-PAGE gel comparing samples “before induction” and “after induction” to assess production (Fig. 6). Also, you can take 5 mL and 10 mL of bacterial culture for a “thermal lysis test” and a “solubility test,” respectively (Fig. 6), which you should pellet by centrifugation and freeze at –20 °C. Furthermore, depending on equipment available and to ease the process, use the biggest centrifugation tubes available (usually 0.5–1 L).
53. Bacterial pellets can be stored at –80 °C for a few weeks, but yield the best results when used in a week. If stored for longer, pellets can be transferred to 50 mL falcon tubes. The easiest way to transfer pellets is while they are still frozen. Hold the palm against the centrifugation tube where the pellet is located for 15–30 s. Then use a spatula to separate the still frozen pellet from the surface of the centrifugation tube, transfer it to a 50 mL falcon tube, and immediately store it back in the freezer.
54. One can also perform lysis by sonication for the first time. If the isolated protein can be refolded after thermal denaturation as part of CD measurements, then this protein is a good candidate for thermal lysis for all of the subsequent isolations.
55. Take 20 µL of each sample and add 7 µL of 4× SDS loading buffer, mix, incubate at 95 °C for 10 min, do a quick spin, and load 20 µL onto SDS gel.
56. Normally even a faint band on an SDS-PAGE corresponding to the protein in the soluble fraction indicates that thermal lysis should at least produce comparable yields to sonication protocols. Some or even the majority of the protein, however, is still usually located in the insoluble fraction. If based on SDS-PAGE, no protein is present in the soluble fraction, then thermal lysis is not advised.
57. Do not use 2 mL tube because sonication in a 2 mL causes foaming due to insufficient height of the liquid.
58. It is advisable to keep the sample on the ice at all times. But the ice usually cannot support the tube while sonicating. Temporary heating of the sample at this stage is not detrimental since the sample will only be used to run an SDS-PAGE gel.
59. When taking sample for soluble fraction, take care to avoid the pellet.

60. Use a 5 mL or 10 mL pipette to resuspend. One can also use the vortex mixer but it might cause excessive foaming.
61. Add water to ice in the beaker, compress the ice, and add thawed ice. The level of water should be up to approx. 2/3 of the height of the beaker. Sample should always be in contact with ice-cold water during sonication, while the top 1/3 of dry ice functions as a support.
62. As isolation of CCPO is usually performed from a bacterial pellet from 2 L of LB fermentation, the final volume of lysate is 30–35 mL, which should be sufficient for the probe to reach the correct depth if using 50 mL falcon tube and described sonicator. The probe should be submerged in the lysate to achieve efficient lysis. If the probe is not deep enough, or the lysate is foamy, excessive foaming or spraying can occur, which greatly decreases the lysis efficiency. Be mindful of the sound—if it sounds muted and dull, check for excessive foaming. If foaming occurs, quick incubation in a sonication bath should remove most of the foaming. If problems persist, perform short centrifugation (approx. 2 min, 3000–5000 × g), followed by additional resuspension with pipetting before resuming sonication. If working with a pellet from a smaller fermentation, two general solutions are suggested: (i) increase the volume of lysis buffer to a final volume of 30–35 mL or (ii) use smaller, appropriate volumes of lysis buffer, but remove the upper part of the falcon tube with a knife, saw, or scissors, so the sonication probe reaches the appropriate depth.
63. The number of cycles required differs from protein to protein, and it also depends on the quantity of lysis buffer used per cellular pellet. A sufficient number of sonication cycles is indicated by a change of color of lysate to a lighter shade and a decrease of lysate viscosity. For soluble CCPO, the lysate can become almost translucent which indicates a sufficient number of cycles. The sonication heats the lysate, so prolonged sonication is not suggested. If long sonication is required, pay close attention to the temperature of the lysate and make short pauses between cycles to chill the samples if needed.
64. Always take care that the steam has some place to escape.
65. The size and the shape of the column dictate the method of positioning. Two methods are most common: (i) modified falcon stand, which is suspended over some collection vessel, such as a tray, and (ii) 1 L or 2 L glass flasks, with the column inserted into the mouth.

66. The color of the resin has to change from cyan to blue. If the resin is any other color, something is wrong and do not continue with the isolation until the cause of color change is identified. When dealing with low concentrations of reducing agents (check resin specifications), pale blue color is also acceptable.
67. Take care not to disrupt the packing of resin when loading the soluble fraction or subsequent washing with the buffer.
68. Flow-through fraction can be collected in a tube and flown though the column again. In our experience, this does not increase the yields though.
69. Keep the column filled to the top at all times as that increases the flow. Moreover, the addition of a piece of tube to the bottom of the column furthermore increases the flow.
70. In case of a blocked column, a suction with a syringe, connected to the bottom part of the column, can be used to speed up the process. Some additional cautionary Notes: (i) do not provide too much suction as to cause bubble formation and thus disrupting the packing of the resin, (ii) do not reverse the flow of the syringe as to push the dirty buffer back onto the column, (iii) always have the resin covered by a liquid as not to dry up the resin. When rinsing with the help of a syringe, the amount of used buffer increases, but it speeds up the process.
71. For typical isolation of one protein from 2 L of bacterial culture, approx. 300 mL of NiNTA buffer A and 200 mL of NiNTA buffer B are required.
72. Take care to equally distribute the elution buffer across the whole area to increase the concentration and minimize elution volume.
73. Periodically check with spectrophotometer whether the protein is still present in the eluted fractions. Normally the elution is complete in 1.5–2 CV.
74. TCEP and denaturing conditions should be avoided. Moreover, perform the lysis in Strep buffer).
75. Strep-tag should be preferentially added to C-terminal of CCPo, with a short 3-amino-acid linker added in between CCPo and Strep-tag.
76. Some CCPo are stabilized by the addition of the second tag, and in those cases, one affinity chromatographic step can be omitted and sample should be handled as either His- or Strep-tag protein only. Note that StrepTrap usually yields purer protein.
77. 5 mL column is normally sufficient for 1 L of bacterial fermentation.

78. Preferred option is to load through the superloop or dedicated sample pump to avoid the FPLC inline filter.
79. It is advisable to collect the flow-through for SDS-PAGE analysis. Moreover, if the expression of a protein of interest is high, or performance of StrepTrap columns is lower and there is still protein of interest left in the flow-through, the Strep-tag isolation step could be repeated.
80. Value at which it stabilizes is machine-specific and should be determined based on previous experience. Generally, when the UV trace starts levelling out, StrepTrap columns could be considered washed. Alternatively, part of the elute from the column during this stage could be analyzed by spectrophotometer to determine the nature of washed material. However, if the columns are fully saturated, some leaking occurs naturally and one should take care not to mistake this leaking of protein of interest for washing.
81. When Strep-tag isolation is used as a polishing step after NiNTA, washing should be done in 1–2 CV; but if Strep-tag isolation is used as the first step of isolation, then several CV of washing are expected.
82. Usually merged fractions amount to 6–8 mL.
83. If no protein was eluted, check the health of the columns by injecting 2–5 mL of HABA buffer—if HABA changes the color from yellow to red upon binding, the ability of StrepTrap to bind Strep-tag is still present. If StrepTrap turn red with the HABA test and the expression of the CCPO is confirmed by SDS-PAGE, Strep-tag could be inaccessible. In that case, lengthen the linker between CCPO and Strep-tag by 3–6 AA.
84. For as long as 0.5 M NaOH is present on the column (also during the first CV of second MQ wash).
85. Ion-exchange chromatography has the highest chances of success when contaminants in the sample do not originate from the CCPO. If there are CCPO degradation products present, they usually elute in a very narrow range of salt concentrations, which adversely affects the recovery. However, despite everything, this technique should be considered when NiNTA and double-tagging do not offer adequate results.
86. If the column has been stored at +4 °C, take care to run at half the maximum rated flow for 20% EtOH, as 20% ethanol is very viscous at lower temperatures.
87. In both cases, equilibrate with IEX buffer A until conductivity stabilizes. If conductivity meter is not available, follow manufacturer's recommendations on CVs.
88. Collect flow-through fractions for future analysis and potential re-isolation if the IEX column got saturated.

89. If possible, perform gel filtration or dialysis on the same day, as high salt concentrations can cause precipitation.
90. Contact time and flow with 1 M NaOH should be kept within the limits of the resin.
91. Smaller dimensions of Superdex 200 pg are also acceptable, just the volume of injection should be adjusted accordingly—e.g., for Superdex 200 pg 16/600 up to 4 mL or Superdex 200 Increase 10/300 up to 0.5 mL. Check the manufacturer's recommendations for appropriate injection volumes. If a column packed with Superdex 200 pg is not available, other resins should offer acceptable separation; but when choosing the alternative resins, one should take care to consider that calibration curves are calculated using globular proteins, and due to different shape, CCPOs elute as bigger particles than their estimated elution time based on Mw alone. Moreover, the alternative column should offer good separation of particles in the size range of trimeric and monomeric CCPO, as those are expected species of CCPO at this stage.
92. 50 mL superloop is recommended for volumes of above 10 mL, but if it is not available, smaller superloops can be used with a bit of ingenuity. When smaller superloops are used, load as much sample onto the superloop as possible, inject one third to two-thirds of loaded volume, pause the flow, load the remaining sample into the superloop, and inject the rest.
93. If the eluted volume from the previous isolation step is too high, the sample could be additionally concentrated with centricons, but that should be avoided, as CCPO are not stable when concentrated in high imidazole or salt concentrations. If the amount of sample is high enough, two SEC runs are preferred.
94. Elution profile for CCPO is distinct and can normally be easily identified, regardless of the type of CCPO. Normally there is no dimeric fraction, with a distinct trimeric fraction, clearly separated from a monomeric. Should there be some overlap between monomeric and trimeric peaks, it is preferable to avoid the protein fractions that still contain a trimeric form of a CCPO, for analysis, that require a monodisperse sample.
95. CCPO travel differently compared to other proteins; therefore, SDS-PAGE should be used only to estimate the Mw.
96. Due to its distinct shape CCPO elutes as a protein of higher Mw.
97. Take care that the level of MQ, buffer, and sample does not fall below the upper-most part of the membrane as not to dry them up. If the membranes dry up, there is a risk of them being damaged and leaking.

98. Empirically the region of 0.8–1.5 mg/mL seems to be the most likely concentrations when visible aggregation usually occurs—once past 1.5 mg/mL, the CCPO are highly likely to be stable and not form visible aggregates at a concentration up to 10 mg/mL.
99. Use 200 μ L pipette to gently mix the solution in lower part of the concentrator with the solution at the top of the concentrator, as the highest concentration is present near the membranes. Mix until there is no difference in the viscosity, which indicates that the sample and concentration are uniform. If there is some observable aggregation while the protein is at low concentration, the following steps should be followed: (i) remove the aggregates by filtering through 0.22 μ m syringe filter and use another centricon, freshly rinsed with MQ, to continue concentrating; (ii) if the aggregation occurs again, measure the concentration of the (filtered) sample before and after spinning—if the concentration after one round of centrifugation is higher, consider concentrating further on.
100. Minor protein leakage when preparing samples at high concentrations is acceptable.
101. CCPO tend to form a small trimeric fraction upon freezing, but usually the amount is barely detectable and does not cause problems for further analysis and/or use. Moreover, CCPO are stable for years when properly frozen and stored.
102. TEV protease is produced in house with the above-described protocols. Briefly, TEV protease is produced in *E. coli* for 4 h at 30 °C, cell lysis is performed with ultrasonication, and purification is done with NiNTA and SEC. Protein is then concentrated to approx. 20 μ M concentration, aliquoted in low bind tubes, shock frozen in liquid nitrogen, and stored on –80 °C.
103. For removal of cleaved products with reverse NiNTA, usage of TCEP is preferred, as DTT is aggressive and can introduce browning of the resin by reducing nickel ions.
104. Cleavage efficiency can be improved by a 2 h incubation at RT before the overnight incubation.
105. If the efficiency of cleavage is low, adding longer linkers between protein domain and the TEV cleavage site can help. If this is not possible or desirable, use a higher concentration of protease.
106. Reverse NiNTA is used to separate his-tags, uncleaved products, and TEV protease from the cleaved protein after TEV cleavage reaction.

107. Resin volume should be adjusted to the expected amount of protein to be bound.
108. Column needs to be well equilibrated with NiNTA A buffer, especially when using DTT as reducing agent.
109. Some proteins that contain multiple histidine residues in the sequence can retain low binding affinity toward nickel and the protein will not elute with buffer A. In this case, it is best to add 5–10 mM imidazole to the reaction solution before transferring it to the column. Also, after washing with buffer A, continue to wash with buffers with increasing imidazole concentration (20 mM, 30 mM, 40 mM) and analyze fractions on SDS-PAGE. After a certain threshold of imidazole, concentration elution of unwanted his-tagged products should start—discard these fractions.
110. If a sample is prepared in a different buffer, avoid high chloride concentrations since they tend to absorb light in the measured spectra.
111. Take care to minimize the dead volumes between components to increase the resolution. Moreover, should the HPLC system lack the inline degasser, all used buffers should be extensively degassed. Also, just one source of concentration is needed, so either UV-MALS or RI-MALS combinations are acceptable.
112. Avoid silica-based columns.
113. 0.5–1 mM TCEP might be added if CCPo contains Cys.
114. All detailed sample handling and conditions are optimized for the following setup: Anton Paar SAXSpoint 5.0 equipped with Primux 100 micro Cu X-ray source, 2D 1 M EIGER2 R series detector, heated/cooled sample cell holder, and ASX autosampler. Parameters such as sample volume, needed sample volume, exposure time, and frames should be only used as a general guideline when using a different setup—the more different that the setup is, the more these parameters could deviate from the stated ones.
115. Samples at the highest concentration of approx. 7 mg/mL usually yield best results.
116. More complex CCPo (especially when they possess unpaired coiled-coils) are sometimes more prone to the formation of oligomers at high concentrations. In those cases, time-consuming and not readily available SEC-SAXS can still be avoided by careful sample preparation. Instead of diluting a highly concentrated sample to lower concentrations, an adequate volume of the sample should be removed at desired concentrations while the sample is being concentrated

(recommended sample concentrations: 1, 2–2.5, 3–3.5, 4.5–5 > 7 mg/mL). This greatly reduces the occurrence of undesired oligomers.

117. Only applicable for SAXS measurements in batch mode. If possible, buffer from the first round of concentrating should not be used for SAXS as it might contain some impurities either from the centricon membrane or from MQ, used for centricon membrane washing.
118. To decrease the amount of generated ozone upon Xenon lamp warm-up phase, purge the lamp chamber with nitrogen for 5–10 min before turning on the lamp.
119. It is advisable to leave the xenon light on to warm up for 15 min before measurements.
120. The sample in the cuvette can be recovered with long-tip pipette tips, commonly used for SDS sample loading.
121. The sample concentration is determined by the path length of the cuvette used, which is in our case 1 mm. The suggested concentration of protein is 0.15–0.3 mg/mL. Raw CD signal at 222 nm for helical structures at this concentration usually results in the range of 30 to 60 mDeg. Be careful not to use too high of a concentration since this can produce oversaturation of photomultiplier especially at lower wavelengths (180–200 nm) and can damage the detector.
122. 1 mm cuvette has the capacity of 300 μ L, so prepare an adequate amount of sample.
123. α helix-rich proteins are known to have high CD signal; therefore, the presence of high absorbing salt ions does not cause significant concern. This allows us to measure CD spectra of protein in buffers that are used for standard gel filtration purification procedures without the need for lengthy dialysis against low chloride concentration buffers such as sodium or potassium phosphate buffer. However, if the signal of measurements is low or distorted due to oversaturation of the photomultiplier, a different buffer should be considered (e.g., phosphate buffer).
124. Tip of the cuvette where the temperature probe is inserted can also be wrapped in parafilm to prevent evaporation.
125. At given settings, each repeat should take 80 s.
126. Optionally set auto-subtraction of background buffer. Background can also be subtracted manually afterward when analyzing the data.
127. At 20 °C one should expect the CCPO protein sample to be folded. Because CCPO structures consist predominantly of α helices, the CD spectra is very distinct and resembles α helix signal with a double minimum at 208 nm and 222 nm.

128. Temperature is regulated with the thermal control block heating, and there is a difference between set temperature of the heating block and temperature of the actual sample that needs to be accounted for during data analysis. At the end, actual sample temperatures measured directly in the cuvette using a temperature probe are used to report the denaturation curve. As a result, the end temperature is always lower (e.g., 85–90 °C) than the actual set temperature (95 °C).
129. Some CCPo structures are so stable that they do not unfold completely even at a temperature as high as 90 °C. In this case, add a low amount of GdnHCl (0.5–2 M) to the sample which should destabilize the protein and make it susceptible to thermal unfolding. Be careful when measuring far-UV spectra of samples with added GdnHCl since excess Cl⁻ ions can affect the quality of measurements.
130. For a described setup, the measurement lasts roughly 1.5 h per thermal denaturation.
131. When denaturation reaches the end temperature, immediately proceed with the next step of measuring spectra since prolonged periods of a sample at high-temperature cause evaporation of water from the sample which changes concentration which in turn influences the measurement.
132. At 95 °C, it is expected for the protein sample to be fully or partially unfolded. The signal curve at this point should resemble the profile of a random coil. However, some CCPo structures are so stable that retain some folded structure even at 95 °C. For those, look up **Note 129**.
133. Most CCPo-based proteins are capable of spontaneous refolding after thermal denaturation, so the signal should be comparable to the one measured before thermal denaturation. At this stage, it is expected for the signal to be marginally stronger, due to minimal increase in the concentration due to water evaporation.
134. Res is a number of amino acid residues minus 1.
135. Some proteins might have a more complex denaturation curve, e.g., two- or three-state denaturation that require different equations and a thermodynamic model for fitting. In this case, consult the literature [6, 14].
136. Other resins and sizes might also be used, but Superdex 200 Increase 10/300 offers best results for CCPo. Silica-based resins are not advised as CCPOs tend to stick to them, which will increase the back-pressure, and more worryingly, might produce false results when dealing with CCPo components that might cross-react.
137. To protect the longevity of the column, do not wash at maximum flows.

138. The exact concentrations and volumes depend on the specifications and the sensitivity of the system. Using Waters e2695 HPLC system, coupled with a 2489 UV detector (Waters), a Dawn8+ MALS detector (Wyatt) and RI detector RI500 (Shodex), concentrations of above 0.3 mg/mL yield good Mw estimates, providing that the sample is monodisperse. Using higher concentrations decreases the noise and increases the confidence in the measured Mw. For described setup, concentrations of approx. 0.5 mg/mL and injection volumes of 100 μ L are routinely used.
139. Take care that all systems are connected and ready for injection before starting the injection. Be especially mindful that the times of data collection and time of runs match to avoid potential data loss.
140. CCPO elute at lower elution volumes compared to elution times from calibration curves. Monomeric fraction usually amounts to more than 98% of the total mass of the sample, with calculated Mw within 2–3 kDa of theoretical Mw.
141. The column should be washed with water and 20% EtOH at the end of each week or after 30–40 CCPO injections. Take special care if the recovery of the sample is poor—some CCPOs (especially if some of the coiled-coils are not paired by design) are prone to sticking to the resin, which is made even worse when using silica-based columns.
142. Small amounts of Hellmanex or equivalent cleaning solutions might be used, if it is compatible with material surfaces.
143. Alternatively, centrifugation at $20000 \times g$ for 15 min at 4 °C also removes bigger particles.
144. If diluting a sample from high concentration, a twofold serial dilution is customary, although (sample volume permitting), custom higher concentrations could also be made (e.g., instead of only diluting from starting concentration of 7 to 3.5 mg/mL, make additional one at approx. 5 mg/mL).
145. When using the autoinjector, take care that the capillary is adequately washed with MQ and dried between samples and before the first sample. Moreover, the remaining sample after the measurement can be collected and repurposed for other analysis.
146. In case the samples are measured at a synchrotron or some other arrangement, data might be already reanalyzed and thus some steps should be omitted.
147. If Linux is used as an OS, .dat file extension has to be added manually.

148. **Averaged 1D SAXS curves, corresponding to samples** should always be positioned above curves, corresponding to matching buffers—order can be changed by Right click → Move Up/Move Down accordingly.
149. Should the measured buffer not match the sample buffer completely, a buffer mismatch will occur. Buffer mismatch can be identified when lower concentrations are prepared by diluting a stock at high concentration, as with the concentration decreasing, the buffer mismatch is less noticeable due to higher proportion of an additional buffer addition. So far, there have been no observed interactions between CCPD and SEC buffer components which would lead to asymmetrical retention of buffer components when concentrating samples. Therefore, should buffer mismatch occur, it is most likely due to an error during sample preparation.
150. Normalization is only important when attempting to calculate Mw from I_0 from a known sample.
151. After normalization, all normalized curves should overlay.
152. Crop in such a way that only a portion of the curve from $0.3\text{--}0.8 \text{ nm}^{-1}$ is visible.
153. Scaled and cropped 1D SAXS curves should overlay for the whole region. If there is a visible shift in one or more of the curves, shorten the portion of the curve to scale.
154. Loss of resolution with lower concentration at higher angles is expected, but pay special attention to the first 50 points at lower angles. If the normalized or scaled 1D SAXS curves curve up with higher or lower concentration, then there is a concentration-dependent oligomerization effect. For simpler CCPD, there is rarely an increase in oligomerization state, and consequently in the left-most portion of the SAXS curve with rising concentration, while again modified CCPD and CCPD with unpaired coiled-coils have been shown to potentially have some concentration effects.
155. I_0 from a merged SAXS curve should not be used for Mw estimation from a known standard.
156. They are noise due to the beam stopper. Normally, no more than 10–20 points should be removed.
157. There are several ways of determining cutoff at higher angles. First way is to cut x axis at the point where intensity (I) drops for more than 2 orders of magnitude. Other ways of determining where to cut are approximating the point where the SAXS curve levels off, or when the noise becomes too great. Although different traditions rely on different cutoff parameters, the final result is that usually the threshold falls in the region of $2.5\text{--}3 \text{ nm}^{-1}$.

158. Chi of bellow 2 is considered acceptable, but should be improved upon, while chi of bellow 1.4 is considered satisfactory. More importantly, always check few best-fitting models if all of them are of the same shape and not deformed in some way.
159. Take care that Guinier region is appropriate—normally Primus' suggestion offers the appropriate solution.
160. Unless the normalized SAXS curve has been saved and re-opened, I_0 listed is not normalized.
161. If there is oligomerization present in the sample, it will be most clearly presented here. If upon changing the D_{max} value (listed here as R_{max}), $P(r)$ function start rising again at higher r values, this is indicative of oligomers in the sample. A quick way to estimate D_{max} is to set D_{max} as a high value (e.g., 50) and check where the $P(r)$ function crosses the x-axis for the first time and the D_{max} value is normally a little bit lower. Kratky plot can again be visualized in Primus in Plot → Kratky plot. It is normal for CCPO to exhibit flexibility in the Kratky plot, as CCs are not rigidly bound to one another—a small movement of the CCPO could be described as “breathing” and is characteristic of this type of protein folds.
162. Envelopes, generated with dammif, normally do not have a distinctive hole in the middle of the structure; therefore, the usage of dammif is discouraged.

References

1. Gradišar H, Jerala R (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J Pept Sci* 17:100–106. <https://doi.org/10.1002/psc.1331>
2. Gradišar H, Božič Abram S, Doles T et al (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat Chem Biol* 9:362–366. <https://doi.org/10.1038/nchembio.1248>
3. Ljubetić A, Lapenta F, Gradišar H et al (2017) Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat Biotechnol* 35:1094–1101. <https://doi.org/10.1038/nbt.3994>
4. Lapenta F, Aupič J, Vezzoli M et al (2021) Self-assembly and regulation of protein cages from pre-organised coiled-coil modules. *Nat Commun*. <https://doi.org/10.1038/s41467-021-21184-6>
5. Majerle A, Hadzi S, Aupič J et al (2021) A nanobody toolbox targeting dimeric coiled-coil modules for functionalization of designed protein origami structures. *Proc Natl Acad Sci U S A* 118. <https://doi.org/10.1073/pnas.2021899118>
6. Aupič J, Strmšek Ž, Lapenta F et al (2021) Designed folding pathway of modular coiled-coil-based proteins. *Nat Commun*. <https://doi.org/10.1038/s41467-021-21185-5>
7. Eswar N, Webb B, Marti-Renom MA et al (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 50:2.9.1–2.9.31. <https://doi.org/10.1002/0471140864.ps0209s50>
8. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF chimera – a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>
9. McGibbon RT, Beauchamp KA, Harrigan MP et al (2015) MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys J* 109:1528–1532. <https://doi.org/10.1016/J.BPJ.2015.08.015>
10. Sabina B, Gradišar H, Aupič J et al (2021) Triangular in vivo self-assembling coiled-coil

- protein origami. Cite This ACS Chem Biol 16: 310–315. <https://doi.org/10.1021/acschembio.0c00812>
11. Petoukhov MV, Franke D, Shkumatov AV et al (2012) New developments in the *ATSAS* program package for small-angle scattering data analysis. J Appl Crystallogr 45:342–350. <https://doi.org/10.1107/S0021889812007662>
12. Drobnač I, Gradišar H, Ljubetič A et al (2017) Modulation of coiled-coil dimer stability through surface residues while preserving pairing specificity. J Am Chem Soc 139:8229–8236. <https://doi.org/10.1021/jacs.7b01690>
13. Berrow NS, Büssow K, Coutard B et al (2006) Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. Acta Crystallogr Sect D Biol Crystallogr 62:1218–1226. <https://doi.org/10.1107/S0907444906031337>
14. Drobnač I, Vesnauer G, Lah J (2010) Model-based thermodynamic analysis of reversible unfolding processes. J Phys Chem B 114: 8713–8722. <https://doi.org/10.1021/jp100525m>



Chapter 2

Artificial Protein Cages Assembled via Gold Coordination

**Karolina Majsterkiewicz, Izabela Stupka, Kinga Borzęcka-Solarz,
Artur Biela, Szymon Gaweł, Monika Pasternak, and Jonathan Heddle**

Abstract

Artificial protein cages made from multiple copies of a single protein can be produced such that they only assemble upon addition of a metal ion. Consequently, the ability to remove the metal ion triggers protein-cage disassembly. Controlling assembly and disassembly has many potential uses including cargo loading/unloading and hence drug delivery. TRAP-cage is an example of such a protein cage which assembles due to linear coordination bond formation with Au(I) which acts to bridge constituent proteins. Here we describe the method for production and purification of TRAP-cage.

Key words Protein cage, Au(I), Gold ions, Self-assembly, Protein design, Nano-containers, Artificial capsid

1 Introduction

Protein cages are nanometric hollow protein shells, most typically being convex polyhedral in shape. Naturally occurring protein cages have been engineered with various applications in mind with the most well-known examples perhaps being virus capsids for cell delivery [1–3]. Other naturally occurring cages that have been engineered include intracellular proteins such as encapsulins [4–6] and ferritin [7].

As was noted by Crick and Watson [8, 9], virus capsids have to be made from a large number of identical proteins—the nucleic acid sequence required to encode a single protein would be too large to fit within the produced capsid. Additionally, such a large peptide chain would likely be impossible to fold correctly. Therefore, virus capsids and other protein cages typically exist as protein complexes. In nature, the constituent protein building blocks of multi-component protein complexes are bound to partner proteins via protein-protein interactions. These consist of multiple bonds including salt bridges, hydrogen bonds, and hydrophobic packing

interactions. One consequence of these interactions is that such proteins assemble spontaneously and typically require very harsh conditions to disassemble [10], making them of questionable utility in transport of biomacromolecules. One solution to this problem is to design and produce artificial protein cages where the complex protein-protein interactions are replaced with a simpler, single type of interaction, namely, coordination bonds, wherein metal ions act as the “glue” to hold together the protein subunits. This has been achieved with the production of “TRAP-cage”—a large artificial protein cage held together by linear coordination bonds involving Au(I) which bridges the constituent TRAP proteins [11–13].

TRAP (trp RNA-binding attenuation protein) is a protein commonly found in some bacteria including *Geobacillus* spp. where it is involved in the control of tryptophan synthesis. It is notable for its high stability and ring shape. The protein has been well characterized biochemically and structurally [14–17] as well as being used as a construction material for higher order nanostructures [18–20] and the functional ring typically consists of 11 TRAP monomers. Because of its interesting shape, high stability, and ease of modification, TRAP has also become a focus of bionanoscientists aiming to use it as building block for construction of artificial structures and shapes [11–13, 18–21].

TRAP-cage is an artificial cage made from 24 TRAP rings with each ring arranged as if centered on the vertex of a snub cube [11]. The cage-forming TRAP carries a mutation of the lysine at position 35 to cysteine, which lies on the exposed outer rim of the ring and is the only cysteine in the structure (Fig. 1) with an R64S optionally added for further charge modification of the surface [13]. Ten of the eleven monomers in each ring are connected to monomers in a neighboring ring by linear Au(I) coordination bonds between the thiols of the opposing cysteines. The resulting TRAP-cage is highly stable but can be easily disassembled by disruption of the coordinate bond, e.g., by the addition of relatively low concentrations of reducing agents. Below we detail the comprehensive methods for production, purification, and characterization of TRAP-cages.

2 Materials

Prepare all the buffers using ultrapure water (with resistivity approx. 18 mΩ at 25 °C). Reagents used are analytical or molecular biology grade and can be purchased from standard suppliers.

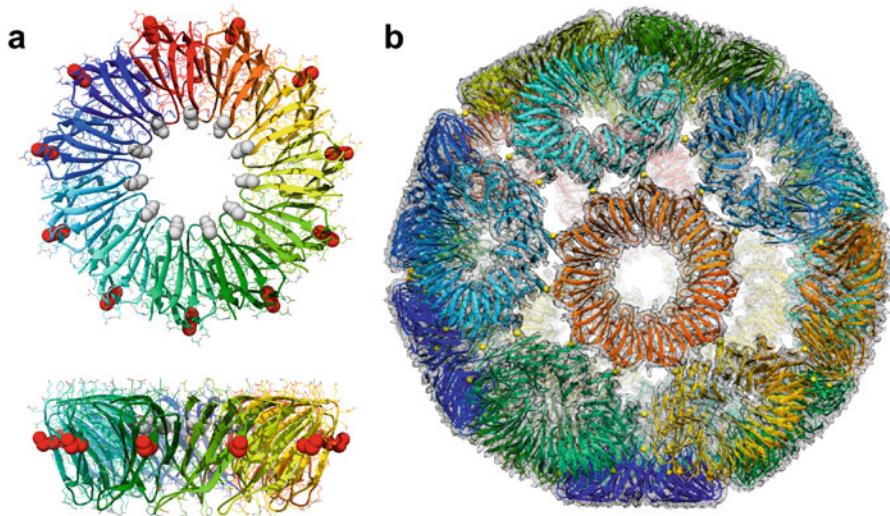


Fig. 1 TRAP and TRAP-cage. (a) Two orthogonal views of the TRAP protein (PDB: 4v4f [22]) with sites of residue change Cys35 and Ser64 shown as spheres on the external and internal rims, respectively. (b) Structure of the TRAP-cage (right-handed version, PDB 6rvw). Cysteine residues are shown in ball and stick representation, while gold(I) residues are shown as spheres between rings. Figure adapted from reference [11]

2.1 TRAP Protein

TRAP (trp RNA-binding attenuation protein) is a ring-shaped protein derived from *Geobacillus stearothermophilus*, a thermophilic, aerobic bacterium [23]. This protein is very interesting in itself due to the large differences in the regulation of the tryptophan biosynthesis genes expression between bacteria with TRAP (e.g., *Geobacillus stearothermophilus* or *Bacillus subtilis*) and bacteria without TRAP (e.g., *Escherichia coli*) [14, 23]. The properly formed TRAP ring consists of 11 monomers with an average mass of 8242.36 Da (original form, PDB: 1QAW [23]) or 8148.22 Da (mutant K35C/R64S, PDB: 6RVW or 6RVV [11]) each, giving the total unliganded mass of approximately 90,666 Da for original, and approximately 89,630 Da for mutant versions. Due to its origin from thermophilic bacteria, TRAP protein is very stable at high temperatures, a feature exploited during the purification process. Due to the introduced K35C/R64S mutations, the theoretical pI value for the mutant is significantly lower (6.39) than the pI value of the natural protein (8.09).

2.2 TRAP K35C/R64S Expression Plasmid

TRAP protein is straightforward to express and purify. The amino acid sequence of the protein is based on PDB entry 1QAW [23], with introduced K35C and R64S mutations. The full sequence can be accessed from the PDB entry for TRAP-cage, 6RVW (right handed version) or 6RVV (left handed) [11]. Reverse translation of the amino acid sequence yields the gene sequence which should be optimized for expression in *E. coli*. Gene synthesis and cloning into pET21b (+) vector plasmid can be performed using a commercial service.

2.3 Protein Expression

1. Chemically competent *E. coli* BL21 (DE3) strain.
2. Incubator for culture plates.
3. Incubator with shaking for 2.5–3.0 L culture flasks.
4. LB Broth medium: 5 g NaCl, 10 g tryptone, 5 g yeast extract per liter.
5. LB agar plates: 15 g agar per liter of LB supplemented with 100 µg/mL ampicillin.
6. Ampicillin stock solution (100 mg/mL) (*see Note 1*).
7. 1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG) stock solution (*see Note 1*).
8. Centrifuge capable of ≥6,000 g relative centrifugal force.
9. –80 °C or –20 °C freezer.

2.4 Protein Purification

1. TRAP preparation Buffer A: 0.05 M Tris–HCl, 0.05 M NaCl, pH 7.9, 0.002 M DTT.
2. TRAP preparation Buffer B: 0.05 M Tris–HCl, 1 M NaCl, pH 7.9, 0.002 M DTT.
3. TRAP SEC buffer: 0.05 M Tris–HCl, 0.15 M NaCl, pH 7.9.
4. Dithiothreitol (DTT) stock solution 1 M (*see Note 1*).
5. Lysozyme, e.g., 20,000 U/mg in pH 6.2 (*see Note 2*).
6. Deoxyribonuclease I (DNase I), e.g., 3000 Kunitz units/mg protein (*see Note 3*).
7. Protease inhibitor tablets (*see Note 4*).
8. Magnetic stirrer.
9. Ultrasonic liquid processor with probe capable of processing 50 mL of liquid sample (*see Note 5*).
10. Centrifuge capable of ≥50,000 g centrifugal force.
11. Water bath with the heating up to 70 °C.
12. 0.22 µm membrane filters (*see Note 6*).
13. 0.22 µm syringe filters and centrifugal concentrators with the 50 kDa cut off (*see Note 7*).
14. Fast protein liquid chromatography (FPLC) system with UV-Vis absorbance detection.
15. SDS-PAGE system, gels, and buffers (for detailed materials—*see subheading 2.5*).
16. UV-Vis absorbance spectrometer.

2.5 Tris-Tricine SDS-PAGE

1. 20% Sodium dodecyl sulfate (SDS) (*see Note 8*).
2. 30% (37.5:1) Acrylamide/bis (*see Note 8*).
3. 10% Ammonium persulfate (APS) (*see Note 1*).

4. N,N,N',N'-tetramethylethylenediamine (TEMED).
5. 3× Stock gel buffer to prepare gels: 3 M Tris–HCl, pH 8.45, 0.3% [w/v] SDS (*see Note 9*).
6. 4% Stacking and 2-layer 10/16% or 1-layer 15% resolving gel (*see Note 10*): 1 M Tris–HCl, pH 8.45, 0.1% [w/v] SDS, Acrylamide/bis (37.5:1), 0.06% APS, 0.1% TEMED.
7. 5× Anode running buffer: 0.5 M Tris–HCl, pH 8.9 (*see Note 9*).
8. 1× Cathode running buffer: 0.1 M Tris–HCl, 0.1 M Tricine, 0.1% [w/v] SDS. Do not adjust pH (*see Note 11*).
9. Stock solution for sample buffer: 1 M Tris–HCl, pH 6.8, 2% [w/v] Bromophenol blue.
10. 4× Reducing sample buffer: 200 mM Tris–HCl, pH 6.8, 0.4% [w/v] Bromophenol blue, 40% [w/v] glycerol, 8% [w/v] SDS, 400 mM β-mercaptoethanol (*see Note 12*).
11. 8 M Urea.
12. Molecular weight marker in range 10–250 kDa.
13. Protein gel stain, e.g., InstantBlue Coomassie Protein Stain.

2.6 Chromatography

Columns

1. Ion exchange column: HiTrap Q Fast Flow, 5 mL (Cytiva).
2. SEC_column_1: HiLoad 26/600 Superdex 200 pg (Cytiva) (*see Note 13*).
3. SEC_column_2: Superose 6 Increase 10/300 GL (Cytiva).

2.7 Source of Metal Ions

TRAP-cages can be assembled using Au(I) or Hg(II) ions. These are provided as chloro[diphenyl(3-sulfonatophenyl)phosphine] gold(I) (Au-TPPMS) and mercury(II) nitrate, respectively (Fig. 2). In our subsequent methods, we concentrate on cage assembly using Au(I).

2.8 TRAP-Cage Self-assembly Reaction and Purification

1. Aqueous Au-TPPMS stock solution (10 mM) (*see Note 12*).
2. Benchtop centrifuge capable of ≥12,000 g centrifugal force.

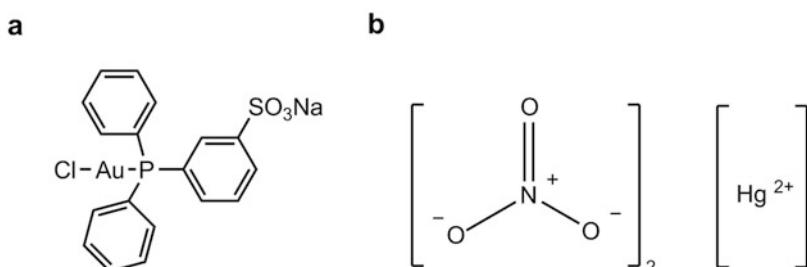


Fig. 2 The chemical structures of Au-TPPMS (a) and mercury(II) nitrate (b)

3. 0.22 µm centrifugal filter (e.g., 0.5 mL) (*see Note 7*).
4. Centrifugal concentrators with the 100 kDa cut off (*see Note 7*).

**2.9 TRAP-Cage
Validation and
Characterization**

1. Native PAGE system, gels, and buffers (for detailed materials—*see* subheading 2.10).
2. Formvar/carbon-coated copper grids, 400 mesh.
3. 3% Phosphotungstic acid stain, pH = 7.4.
4. Filter paper, quantitative grade.
5. Benchtop centrifuge capable of ≥16,800 g centrifugal force.
6. Glow discharger/coater.
7. Transmission electron microscope.
8. Dynamic light scattering system.
9. Quartz cuvette.
10. Thermocycler.
11. Destabilizing agents: DTT, Tris(2-carboxyethyl)phosphine (TCEP), L-cysteine (Cys), glutathione (GHS—reduced and GSSG—oxidized form), SDS, guanidine-HCl (Gdn-HCl), urea.
12. 50 mM Glycine-HCl at pH 2 or 3.
13. 50 mM Sodium acetate at pH 4.
14. 50 mM Potassium phosphate at pH 6 or 12.
15. 50 mM Tris-HCl at pH 8.
16. 50 mM Glycine-NaOH at pH 10.
17. 50 mM KCl-NaOH at pH 13.

2.10 Native PAGE

1. 3–12% Bis-Tris Gels, e.g., Native PAGE Bis-Tris Gels (Novex, Life Technologies) (*see Note 14*).
2. 20× Cathode running buffer additive: 0.4% [w/v] Coomassie Brilliant Blue G-250.
3. 20× Running buffer: 1 M Bis-Tris-HCl, pH 6.8, 1 M Tricine.
4. Stock solution for sample buffer: 1 M Bis-Tris-HCl, pH 7.2, 2% [w/v] Bromophenol blue.
5. 4× Native sample buffer: 200 mM Bis-Tris-HCl, pH 7.2, 0.4% [w/v] Bromophenol blue, 40% [w/v] glycerol.
6. Native protein standard in range 20–1,200 kDa.
7. Fix solution: 40% methanol, 10% acetic acid.
8. Destain solution: 8% acetic acid.
9. Microwave (min. 700 W).

2.11 TRAP-Cage Cellular Toxicity Tests

1. 0.4% Trypan blue (*see Note 15*).
2. Phosphate-buffered saline (PBS).
3. Alamar Blue (AB)—available commercially as a sterile 10× solution (*see Note 16*).
4. Cell culture medium (DMEM).
5. Hemacytometer or automated cell counter.
6. Plate reader, e.g., by TECAN.

2.12 Transformed Cells for Protein Expression

1. Transform chemically competent *E. coli* BL21 (DE3) strain with plasmid pET21b carrying TRAP K35C/R64S gene and plate the cells on LB agar-plates with 100 µg/mL ampicillin. Incubate overnight in 37 °C.

3 Methods

3.1 TRAP Protein Expression and Purification

1. Pick one colony and use it to inoculate 100 mL LB medium containing 100 µg/mL ampicillin (*see Note 17*). Incubate overnight in 37 °C with shaking.
2. Inoculate 1 L LB medium containing 100 µg/mL ampicillin with 10 mL overnight culture (*see Note 17*).
3. Shake the cells in 37 °C until OD₆₀₀ = 0.5–0.7. Then induce with 0.5 mL 1 M IPTG (final concentration 0.5 mM) and shake in 37 °C for 4–5 h (*see Note 18*).
4. Harvest cells by centrifugation at 6,000 g in 4 °C for 20 min.
5. Store the pellet at –80 °C until purification (*see Note 19*).
6. Prepare buffers for TRAP purification (*see subheading 2* for preparation details):
 - (a) TRAP preparation Buffer A: Weigh 7.88 g of Tris–HCl and 2.92 g of NaCl and transfer to a glass beaker. Add water to a volume of 900 mL. Mix and adjust pH with NaOH (*see Note 20*). Make up to 1000 mL with water. Filter the prepared buffer through 0.22 µm filter (*see Note 6*) and then autoclave. Add 2 mL of 1 M DTT (*see Note 21*). Store at room temperature and degas prior to use.
 - (b) TRAP preparation Buffer B: Weigh 7.88 g of Tris–HCl and 58.44 g of NaCl and prepare 1 L solution as in previous step. Store at room temperature and degas prior to use.
 - (c) TRAP SEC buffer: Weigh 7.88 g of Tris–HCl and 8.77 g of NaCl and transfer to a glass beaker. Add water to a volume of 900 mL. Mix and adjust pH with NaOH (*see Note 20*). Make up to 1000 mL with water. Filter the

prepared buffer through 0.22 µm filter (*see Note 6*) and then autoclave. Store at room temperature and degas prior to use.

7. Resuspend half of the cell pellet (0.5 L culture) in 50 mL Buffer A containing 1 protease inhibitor tablet, DNase I, and lysozyme (*see Note 22*). Incubate in room temperature with stirring for 30 min.
8. Lyse the cells by sonication using 50% amplitude, 2 sec ON cycle, and 3 sec OFF cycle for total 10 min. of ON time. Keep sample on ice during the process (*see Note 23*).
9. Centrifuge at 50,000 g for 20 min. to remove the insoluble fraction (*see Note 24*).
10. Incubate the supernatant in 70 °C for 10 min. to remove proteins degradable in high temperatures.
11. Centrifuge at 50,000 g for 20 min. (*see Note 24*).
12. Filter the supernatant through 0.22 µm syringe filter.
13. Assemble four ion exchange chromatography columns together and equilibrate them with 2 column volumes (CV) of Buffer A, 2 CV Buffer B, 2 CV Buffer A on the FPLC system.
14. Load the supernatant using FPLC system and elute with 0–50% NaCl gradient (from 0% to 50% of Buffer B) over 20 CV at 5 mL/min. Flow rate at room temperature monitoring the absorbance at 280 nm (*see Note 25*).
15. Pool the fractions corresponding to TRAP K35C/R64S (*see Note 26*, Fig. 3a) and concentrate them to 5 mL using centrifugal concentrators with the 50 kDa cut off (*see Note 27*).
16. Filter the sample through 0.22 µm syringe filter to remove any aggregates.
17. Equilibrate SEC_column_1 in SEC buffer (*see Note 28*) and load the sample using FPLC system.
18. Elute the protein with SEC buffer at 2.5 mL/min flow rate monitoring the absorbance at 280 nm.
19. Verify the purity of the protein in fractions by Tris-Tricine SDS-PAGE (*see subheading 3.2*). Pool the fractions corresponding to TRAP K35C/R64S (*see Note 29*, Fig. 3b) and concentrate them using centrifugal concentrators with a 50 kDa cut off.
20. Determine the concentration by measuring the absorbance at 280 nm ($\epsilon_{280} = 8,250 \text{ M}^{-1} \text{ cm}^{-1}$) (*see Note 30*).
21. Store the protein in 4 °C (*see Note 31*).

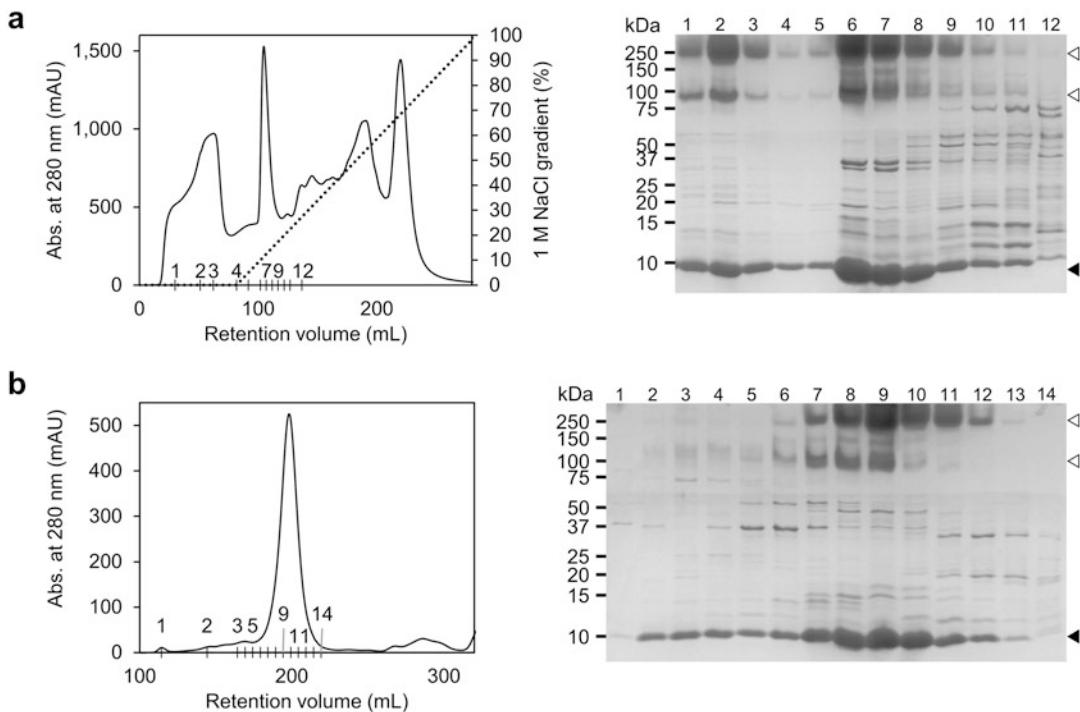


Fig. 3 Example of TRAP purification steps in presence of DTT. **(a)** Ion-exchange chromatogram (left) and corresponding denaturing gel (right). **(b)** Size exclusion chromatogram (left) and corresponding denaturing gel (right). Solid arrowheads indicate the position of TRAP monomers, while hollow arrowheads—the position of TRAP rings and dimers of rings

3.2 Tris-Tricine SDS-PAGE

Tris-Tricine SDS-PAGE is used for optimal separation of proteins with masses below 30 kDa [24]. It is crucial for observation of the 8.2 kDa monomers of TRAP K35C/R64S. TRAP rings are extremely stable; thus, when preparing samples for SDS-PAGE, add 4 M (final concentration) fresh urea and boil samples for 20 min. at 95 °C to remove hydrophobic interactions between monomers. Omitting addition of urea and boiling samples for shorter time, e.g., 10 min., enables observation of a band on the gel corresponding to intact rings (Fig. 3). This can be useful for verification that purified TRAP is forming stable rings.

3.3 TRAP-Cage Self-assembly Reaction

1. Mix TRAP K35C/R64S protein with Au(I)-TPPMS in a 1:1 molar ratio (TRAP monomer:Au(I)) (*see Note 28* and *32*, Fig. 4).
2. Incubate overnight at room temperature (*see Note 33*).
3. Centrifuge at 12,000 g for 5 min. or pass through 0.22 µm centrifugal filter (*see Note 7*) to remove any precipitate (*see Note 34*).
4. Equilibrate SEC_column_2 with SEC buffer and load the sample using an FPLC system.

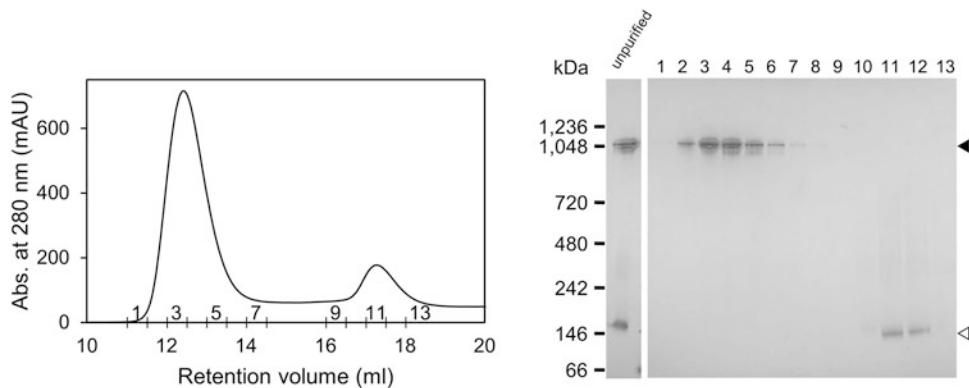


Fig. 4 Example of TRAP-cage purification (0.1 mM scale). Size exclusion chromatogram (left) and corresponding native gel (right). Solid arrowheads indicate the position of TRAP-cage formed with Au(I), while hollow arrowheads—the position of TRAP rings

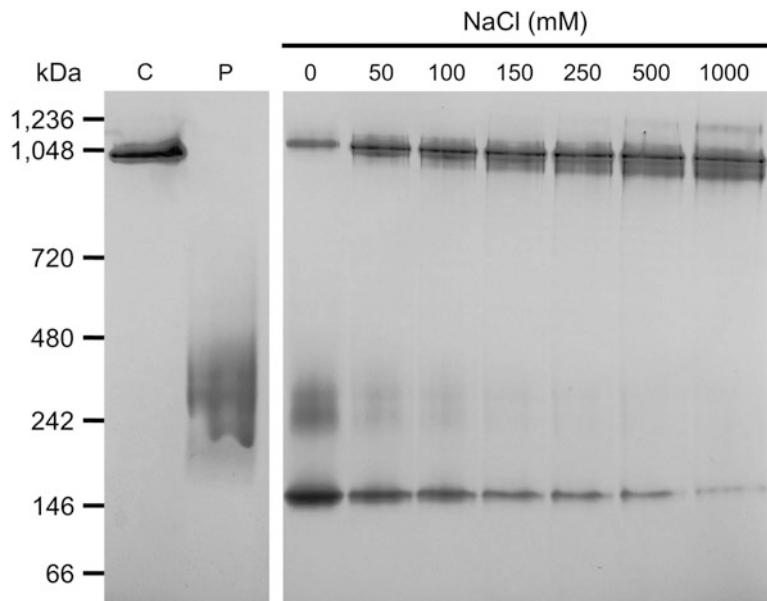


Fig. 5 Influence of ionic strength on TRAP-cage formation. Native gel illustrates cage formation after 60 min. Incubation in presence of increasing concentration of NaCl. C—purified cage; P—unreacted TRAP

5. Elute the protein with SEC buffer at 0.5 mL/min flow rate, monitoring the absorbance at 280 nm.
6. Pool the fractions corresponding to TRAP-cage and concentrate them using centrifugal concentrators with a 100 kDa cut off (see Note 35 and Fig. 5).
7. Determine the concentration by measuring the absorbance at 280 nm ($\epsilon_{280} = 8,250 \text{ M}^{-1} \text{ cm}^{-1}$).

3.4 TRAP-Cage Validation and Characterization

3.4.1 TRAP-Cage Stability Tests

TRAP-cage stability can be easily assessed by native PAGE and TEM. For typical concentrations of destabilizing agents and temperatures in tests, refer to [11].

1. Mix 1 µg of purified TRAP-cage with destabilizing agent (*see Note 36*) or buffer with verified pH and make up to 12 µL total volume with SEC buffer. For temperature stability test, simply dilute 1 µg of purified TRAP-cage with SEC buffer to 12 µL total volume.
2. Leave at room temperature overnight. For temperature stability test, heat samples in the thermocycler.
3. Briefly centrifuge the samples.
4. Take 10 µL of the solution and mix with the native sample buffer.
5. Run native PAGE with not treated sample as a control.
6. Dilute the rest of the sample for TEM analysis.

3.4.2 Dark Blue Native PAGE

In blue native PAGE, Coomassie Brilliant Blue G-250 plays the role of a charge-shift molecule, but does not unfold the protein [25]. Dye molecules bind to protein and give it negative charge, enabling even basic proteins to migrate toward the anode in the neutral pH of native PAGE.

1. Prepare running buffer by diluting stock solution to a working 1× concentration (*see Note 37*). To prepare dark blue cathode running buffer, mix equal volumes of 20× running buffer and 20× additive and dilute with water. Anode running buffer does not require additives.
2. Wash 3–12% gel with water to remove any preservative solution, place the gel in the tank, wash wells with cathode running buffer (*see Note 38*).
3. Mix 10 µL of your sample containing ~1 µg protein with 3.4 µL 4× sample buffer and load on gel.
4. Fill the cathode chamber with dark blue running buffer, and anode chamber with running buffer without additive.
5. Run electrophoresis at 150 V for 1.5 h.
6. Afterward remove the gel from the case, place in clear plastic container with a lid, rinse with water, and immerse in 50 mL of fix solution.
7. Heat in microwave for 45 s with maximum power and shake for 15 min. in RT. Discard fix solution and immerse the gel in 100 mL of destain solution.
8. Heat in microwave for 45 s with maximum power and shake in RT until desired background is reached.

3.4.3 Transmission Electron Microscopy (TEM)

1. Prepare a microscope slide and cover it with parafilm (*see Note 39*).
2. Place the grids on the parafilm using tweezers (*see Note 40*).
3. Glow discharge grids at 80 mA for 70 sec in grid mode (*see Note 41*).
4. Adjust the protein concentration to 0.05 mg/mL.
5. Centrifuge the sample at high speed (16,800 g) for 5 min.
6. Carefully place 4 µL of the protein solution on the grid.
7. Incubate for 1 min. and remove the excess of solution with filter paper (*see Note 42*).
8. Add 4 µL of stain solution.
9. Incubate for 1 min. and remove the excess of solution with the filter paper.
10. Visualize the sample by TEM.

3.4.4 Dynamic Light Scattering (DLS) Measurement

TRAP-cage monodispersity and size can be assessed by DLS. In this method, a laser illuminates the sample and fluctuations of scattered light are detected which gives information about the motion of the particles leading to the calculation of their size (hydrodynamic diameter) and polydispersity index (PDI).

1. Adjust the protein concentration to 0.1 mg/mL.
2. Centrifuge the sample (12,000 g) for 5 min. to remove any precipitate.
3. Transfer 50 µL of the sample to the quartz cuvette (*see Note 43*).
4. Measure the sample using standard settings:
 - (a) Material: protein (RI 1.450).
 - (b) Dispersant: 50 mM Tris–HCl buffer containing 150 mM NaCl, pH 7.9 (temp. 25 °C, viscosity 0.9066 cP, RI 1.332).
 - (c) Temperature 25 °C, 120 s equilibration time.
 - (d) Measurements angle: 173° backscatter, measurement duration: automatic, number of measurements: 3, no delay between measurements.
5. Analyze the data using a “protein analysis” model (Fig. 6).

3.5 TRAP-Cage Cellular Toxicity Tests

Trypan Blue Stain Assay

The dye exclusion test is used to determine the number of viable cells present in a cell suspension. It is based on the principle that live cells possess intact cell membranes that exclude trypan blue dye whereas dead cells do not. In this test, a cell suspension is mixed with dye and then visually examined to determine whether cells take up or exclude dye. A viable cell will have a clear cytoplasm whereas a nonviable cell will have a blue cytoplasm [26].

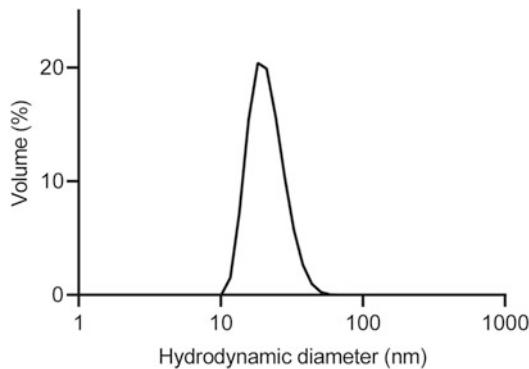


Fig. 6 Dynamic light scattering measurement of Au(I) induced TRAP-cages. Standard measurement of the Au(I)-induced TRAP-cage shows the size of approx. 22 nm (by volume) and PDI = 0.04

1. Prepare the cell suspension—resuspend the cell pellet in PBS or serum-free medium (*see Notes 44 and 45*).
2. Mix the cell suspension at 1:1 with 0.4% trypan blue and incubate ~3 min. at room temperature (*see Note 46*).
3. Load a hemacytometer or counting chamber slide and examine immediately under a microscope or automated cell counter, respectively.
4. Count the unstained (viable) and stained (nonviable) cells separately in the hemacytometer. To calculate the total number of cells per mL of aliquot, add up the number of viable and nonviable cells. Include the dilution factor for trypan blue and multiply number of cells by 2.
5. The cell counter automatically detects the presence of trypan blue dye and automatically reports an accurate cell count of viable and nonviable cell population.
6. Calculate the percentage of viable cells using the following equation:

%Viable cells

$$= \frac{\text{Total number of viable cells per mL of aliquot}}{\text{Total number of cells per mL of aliquot}} \cdot 100$$

3.5.1 Resazurin (Alamar Blue) Assay

This test is based on the natural ability of viable cells to convert resazurin—a blue and nonfluorescent compound, into resorufin—a red and fluorescent molecule by mitochondrial and other reducing enzymes [27].

1. Plate 4×10^4 (suggested cell density) cells per well in a 96-well plate. The optimum cell density may vary between cell types. For each set of conditions, perform the assay in triplicate. Include a control well with untreated cells and a control well with no cells.
2. Remove “old” medium after cell stimulation and add 100 μl per well Alamar Blue reagent diluted 10 times in fresh cell culture medium.
3. Incubate for 1–4 h at 37 °C, 5% CO₂ in an incubator (*see Note 47*).
4. Measure absorbance at a wavelength of 570 nm and 600 nm reading as reference. Or, measure fluorescence with excitation wavelength at 530–560 nm and emission wavelength at 590 nm.
5. The number of viable cells correlates with the magnitude of dye reduction and is expressed as percentage of AB reduction [28]. Calculate the percentage of viable cells using the following equation:

$$\%AB_{\text{reduction}} = \frac{(\varepsilon_{\text{ox}}\lambda_2)(A\lambda_1) - (\varepsilon_{\text{ox}}\lambda_1)(A\lambda_2)}{(\varepsilon_{\text{red}}\lambda_1)(C\lambda_2) - (\varepsilon_{\text{red}}\lambda_2)(C\lambda_1)} \cdot 100$$

where:

$\varepsilon\lambda_1$ and $\varepsilon\lambda_2$ —molar extinction coefficients of Alamar Blue at 570 nm and 600 nm, respectively, in the oxidized (ε_{ox}) and reduced (ε_{red}) forms (*see Note 48*).

$A\lambda_1$ and $A\lambda_2$ —absorbance at 570 nm and 630 nm, respectively.

$C\lambda_1$ and $C\lambda_2$ —absorbance of negative control at 570 nm and 630 nm, respectively.

3.6 Further Characterization

The high-resolution structure of TRAP-cage can be determined using cryo-EM. The detailed methods of data collection and analysis are beyond the scope of this chapter, but here we give a brief overview of how samples are prepared for cryo-EM analysis.

Cryo-EM images are typically obtained using cryo-transmission electron microscopes operating in cryogenic conditions at 200 keV or 300 keV. Field emission guns (FEG) are the preferred electron source, with cold-FEG as the best to date. A very useful set of hardware “additives” in cryo-EM data collection are: an energy filter with a zero-loss function and a phase plate (for small specimens).

Due to cryo-EM sample preparation and data collection complexity, it should always be carried out under supervision or with assistance of an expert in the field. A typical cryo-EM workflow is as follows:

1. Cryo-EM grid preparation—the desired type of grid (*see Note 49*) is plasma-cleaned using a plasma cleaner, e.g., Gatan Solarus, Leica etc., in order to remove any debris deposited on the carbon film during manufacturing and storage. Plasma cleaning conditions can vary depending on the grid type, cleaner type, and sample, but using 70 mA current and about 30 s time is a good starting point.
2. Place 3–4 μL of freshly prepared sample solution onto previously cleaned grid. Avoid cryoprotectants (such as glycerol) in your sample, as they will decrease quality of vitreous ice or even prevent its formation.
3. Plunge-freeze your sample into liquid ethane using, e.g., Leica EM GP2 system or ThermoFisher Vitrobot, or other. Use 100% humidity in order to prevent sample drying during the blotting step. Other parameters need to be adjusted during the grids screening.
4. Transfer frozen grids into the microscope with caution, to avoid any accidental ice-from-air contamination.
5. Collect your data with one of the available cryoEM microscopes; e.g., Titan Krios equipped with specially designed cameras like Gatan K3, or Falcon IV.

4 Notes

1. Store in 1 mL aliquots in -20°C .
2. One unit is defined as the amount of enzyme which decreases the absorbance at 450 nm by 0.001/min. at 25°C and pH 7.0 (*Micrococcus luteus*, ATCC 4698, as substrate) (based on information provided by Sigma-Aldrich).
3. One unit is defined as the amount of enzyme which causes a change in A_{260} of 0.001/min./mL at 25°C and pH 5.0, using DNA as substrate (based on information provided by Sigma-Aldrich).
4. We were using EDTA-free tablets containing the following inhibitors (target enzymes in brackets): Aprotinin (Serine proteases), Bestatin (Aminopeptidases), E-64 (Cysteine proteases), Leupeptin (Serine and cysteine proteases), Sodium fluoride (Serine and threonine phosphatases), Sodium orthovanadate (Tyrosine phosphatases), Sodium pyrophosphate (Serine and threonine phosphatases), β -glycerophosphate (Serine and threonine phosphatases).
5. For comparison, we used a processor having a power supply with 500 W net power output and 20 kHz frequency, standard probe with threaded end, and replaceable tip (diameter 19 mm) with processing capability 50–500 mL.

6. In this study, PVDF (Polyvinylidene fluoride) membranes were used. Other types of membranes to filter buffers can be used as well.
7. We recommend to use filters with PVDF membrane and concentrators with PES (Polyethersulfone) membrane. Other types of membranes can be used as well (e.g., nylon), but avoid cellulose esters as TRAP tends to stick to this type of membrane.
8. Ready-made solution is commercially available. Substituting reagent in powder form by ready-made solution eliminates exposure to powder inhalation.
9. Filter (0.22 µm) and store at RT.
10. Gels with 4% stacking and 15% resolving layer are sufficient when using protocol with sample buffer containing 4 M urea and boiling sample for 20 min.
11. Filter (0.22 µm) before addition of SDS and store in RT. Discard after one use.
12. Store in 100 µL aliquots at -20 °C.
13. Other sizes of Superdex 200 columns can also be used for small-scale purifications, e.g., Superdex 200 Increase 10/300. Remember to set the appropriate flow and pressure for the given columns to avoid compressing the resin.
14. Tris-Glycine native gels with a few layers, e.g., 3.5, 9, 12%, are also sufficient, but the resolution is inferior to gradient gels.
15. Store in dark bottle and filter after prolonged storage.
16. Resazurin dye is light sensitive, protect from prolonged exposure to light.
17. For smaller scale purification, inoculate 4 mL LB medium containing 100 µg/mL ampicillin with one colony. 2 mL of overnight culture can be used for inoculation of 200 mL LB medium.
18. Cells can be also induced with lower IPTG concentration (final 0.2 mM) and shaken for 20 h at 25 °C.
19. Cell pellets can be also stored in -20 °C for up to a several months.
20. When adjusting pH, use concentrated NaOH first (e.g., 10 M), then switch to NaOH with lower concentration (e.g., 5 M) to avoid exceeding the target pH.
21. DTT should be added to buffer after autoclaving, just before use. High temperature and long storage in room temperature causes rapid DTT degradation.

22. Approx. concentrations for the enzymes are: lysozyme—0.1 mg/mL and DNase I—1 µg/mL; however, adding small amount of powder at the end of spatula is sufficient.
23. TRAP K35C/R64S is a thermostable protein; hence, the temperature of the suspension while sonicating does not need to be kept strictly at 4 °C. However, do not let the suspension overheat. Check the position of the container with the suspension in regard to the probe and amount of the ice after half of the sonication time. Adjust the container and add ice if needed.
24. Speed of centrifugation can be reduced to 13,000 g especially when purifying from smaller culture.
25. Assess the amount of protein in supernatant by measuring absorbance at 280 nm (1 Abs unit = 1 mg/mL). Loading supernatant containing more than 700 mg of total protein does not increase the amount of TRAP purified in the process, when using four ion-exchange columns assembled together. You can use 0–100% NaCl gradient, but it is not necessary.
26. TRAP K35C/R64S elutes at 15% NaCl gradient corresponding to approx. 150 mM NaCl. Verify TRAP presence in fractions by Tris-Tricine SDS-PAGE.
27. Pooled fractions prior to concentration can be stored at 4 °C until the next day if needed.
28. SEC purification can be carried out in 0.05 M Tris-HCl, 0.6 M NaCl, pH 7.9, since subsequent cage formation proved to be more efficient in buffer with higher ionic strength (Fig. 5).
29. TRAP K35C/R64S elutes at about 200 mL retention volume.
30. Keep the protein concentration not higher than 1 mM for storage to avoid precipitation.
31. TRAP K35C/R64S has a tendency for aggregation upon the storage at –20 °C. For longer term storage in 4 °C, TRAP K35C/R64S can be supplied with 2 mM DTT, which needs to be removed prior to TRAP-cage assembly. Remove DTT by washing protein solution 3–4 times with SEC buffer using a centrifugal concentrator with a 50 kDa cut off. Avoid dialysis as TRAP K35C/R64S tends to aggregate.
32. For cage formation, try to use freshly prepared TRAP K35C/R64S. Yield of cage formation decreases quickly upon the storage of the protein. We usually keep the final protein concentration in the range 0.1–1 mM.
33. Time of incubation can be decreased to 1 h if needed. TRAP-cage was shown to assemble within a few minutes [11].
34. Addition of Au-TPPMS to the freshly purified TRAP K35C/R64S should not result in any visible precipitation. Precipitation becomes visible upon the addition of Au-TPPMS to the

protein which was stored in 4 °C for longer period of time resulting in lower yield of cage formation.

35. TRAP-cage elutes at about 12.5 mL retention volume.
36. We usually test DTT, TCEP, Cys, GSH, GSSG, SDS, Gdn-HCl, and urea for stability assessment. Adjust pH of the prepared solutions to ~7.9 (especially TCEP and glutathione) to avoid low pH effect on cage stability instead of tested agent.
37. TRAP-cage is thermostable, thus native PAGE can be run at RT. Make sure that running buffers have reached RT, if they were stored at 4 °C.
38. Do not fill the cathode chamber with running buffer before loading samples. Depending on the system you use, running buffer with Coomassie prevents clear observation of the wells. For the same reason, observation of the migrating dye front is also hindered.
39. Parafilm prevents grids from slipping off the slide. Avoid trapping air bubbles under the parafilm.
40. Ensure placing the grids with proper side up. Avoid touching the middle of the grid as it can be easily destroyed. In order to move it, the edges of the grid should be held with tweezers.
41. Given parameters may change depending on the device used. In this study, a Leica EM ACE200 was used.
42. To remove the excess solution, touch only the grid's edge with filter paper.
43. Disposable polystyrene cuvettes can be also used.
44. If you work with adherent cells, use trypsin to get them into suspension instead of scraping.
45. Use PBS or serum-free medium. Serum proteins stain with trypan blue and can produce misleading results.
46. Make sure the stock cell suspension is thoroughly mixed with trypan blue and count cells within 3 to 5 min. After mixing with the reagent.
47. The ability of different cell types to reduce resazurin to resorufin depends on the metabolic capacity of the cell line and the length of time treatment with the reagent. This should be estimated experimentally.
48. Molar extinction coefficients [$M^{-1} \text{ cm}^{-1}$] for Alamar Blue at different wavelengths:
570 nm: $\epsilon_{\text{red}} = 155,677$ and $\epsilon_{\text{ox}} = 80,586$
600 nm: $\epsilon_{\text{red}} = 14,652$ and $\epsilon_{\text{ox}} = 117,216$
49. Most common grids are made of copper and come with variety of pore sizes ranging from 1.2 μm to 3 μm .

Acknowledgments

JGH, AB, and KB-S were funded by a National Science Centre (NCN, Poland) grant no. 2019/34/A/NZ1/00196 (Maestro 11).

Conflict of Interest The authors declare the following competing financial interests J.G.H., A.B., S.G., I.S., K.M., and K. B.-S. are named as inventors on a number of patent applications related to TRAP-cage assembly, decoration, and filling. J.G.H. is the founder of and holds equity in nCage Therapeutics LLC, which aims to commercialize protein cages for therapeutic applications.

References

1. Ma Y, Nolte RJM, Cornelissen JJLM (2012) Virus-based nanocarriers for drug delivery. *Adv Drug Deliv Rev* 64:811–825
2. Bruckman MA, Czapar AE, Steinmetz NF (2018) Drug-loaded plant-virus based nanoparticles for cancer drug delivery. In: Methods in molecular biology. Humana Press Inc., pp 425–436
3. Chung YH, Cai H, Steinmetz NF (2020) Viral nanoparticles for drug delivery, imaging, immunotherapy, and theranostic applications. *Adv Drug Deliv Rev* 156:214–235
4. Giessen TW (2016) Encapsulins: microbial nanocompartments with applications in biomedicine, nanobiotechnology and materials science. *Curr Opin Chem Biol* 34:1–10
5. Künzle M, Mangler J, Lach M, Beck T (2018) Peptide-directed encapsulation of inorganic nanoparticles into protein containers. *Nanoscale* 10:22917–22926
6. Moon H, Lee J, Min J, Kang S (2014) Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplateform. *Biomacromolecules* 15:3794–3801
7. Jutz G, van Rijn P, Miranda BS, Böker A (2015) Ferritin: a versatile building block for bionanotechnology. *Chem Rev* 115:1653–1701
8. Crick FHC, Watson JD (1956) Structure of small viruses. *Nature* 177:473–475
9. Crick F, Watson J (1957) Virus structure: general principles. In: Wolstenholme GEW, Millar ECP (eds) CIBA foundation symposium on the nature of viruses. Wiley Online Library, London, pp 5–13
10. Yoshizawa K, Mishima Y, Park SY, Heddle JG, Tame JRH, Iwahori K, Kobayashi M, Yamashita I (2007) Effect of N-terminal residues on the structural stability of recombinant horse L-chain apoferritin in an acidic environment. *J Biochem* 142:707–713
11. Malay AD, Miyazaki N, Biela A, Chakraborti S, Majsterkiewicz K, Stupka I, Kaplan CS, Kowalczyk A, Piette BMAG, Hochberg GKA, Wu D, Wrobel TP, Fineberg A, Kushwah MS, Kelemen M, Vavpetič P, Pelicon P, Kukura P, Benesch JLP, Iwasaki K, Heddle JG (2019) An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* 569:438–442
12. Malay AD, Heddle JG, Tomita S, Iwasaki K, Miyazaki N, Sumitomo K, Yanagi H, Yamashita I, Uraoka Y (2012) Gold nanoparticle-induced formation of artificial protein capsids. *Nano Lett* 12:2056–2059
13. Imamura M, Uchihashi T, Ando T, Leifert A, Simon U, Malay AD, Heddle JG (2015) Probing structural dynamics of an artificial protein cage using high-speed atomic force microscopy. *Nano Lett* 15:1331–1335
14. Antson AA, Dodson EJ, Dodson G, Greaves RB, Chen XP, Gollnick P (1999) Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* 401:235–242
15. Antson AA, Otridge J, Brzozowski AM, Dodson EJ, Dodson GG, Wilson KS, Smith TM, Yang M, Kurecki T, Gollnick P (1995) The structure of trp RNA-binding attenuation protein. *Nature* 374:693–700
16. Malay AD, Watanabe M, Heddle JG, Tame JRH (2011) Crystal structure of unliganded TRAP: implications for dynamic allostery. *Biochem J* 434:427–434
17. Watanabe M, Heddle JG, Kikuchi K, Unzai S, Akashi S, Park S, Tame JRH (2009) The nature of the TRAP – anti-TRAP complex. *PNAS* 106:2176–2181

18. Heddle JG, Fujiwara I, Yamadaki H, Yoshii S, Nishio K, Addy C, Yamashita I, Tame JRH (2007) Using the ring-shaped protein TRAP to capture and confine gold nanodots on a surface. *Small* 3:1950–1956
19. Heddle JG, Yokoyama T, Yamashita I, Park SY, Tame JRH (2006) Rounding up: engineering 12-membered rings from the cyclic 11-mer TRAP. *Structure* 14:925–933
20. Miranda FF, Iwasaki K, Akashi S, Sumitomo K, Kobayashi M, Yamashita I, Tame JRH, Heddle JG (2009) A self-assembled protein nanotube with high aspect ratio. *Small* 5:2077–2084
21. Watanabe M, Mishima Y, Yamashita I, Park S-Y, Tame JRH, Heddle JG (2008) Intersubunit linker length as a modifier of protein stability: crystal structures and thermostability of mutant TRAP. *Protein Sci* 17:518–526
22. Hopcroft NH, Manfredo A, Wendt AL, Brzozowski AM, Gollnick P, Antson AA (2004) The interaction of RNA with TRAP: the role of triplet repeats and separating spacer nucleotides. *J Mol Biol* 338:43–53
23. Chen XP, Antson AA, Yang M, Li P, Baumann C, Dodson EJ, Dodson GG, Gollnick P (1999) Regulatory features of the trp operon and the crystal structure of the trp RNA-binding attenuation protein from *Bacillus stearothermophilus*. *J Mol Biol* 289:1003–1016
24. Schägger H (2006) Tricine-SDS-PAGE. *Nat Protoc* 1:16–22
25. Wittig I, Braun H-P, Schägger H (2006) Blue native PAGE. *Nat Protoc* 11(1):418–428
26. Strober W (2015) Trypan blue exclusion test of cell viability. *Curr Protoc Immunol* 111:A3. B.1–A3.B.3
27. Rampersad SN (2012) Multiple applications of Alamar Blue as an indicator of metabolic function and cellular health in cell viability bioassays. *Sensors* 12:12347–12360
28. Al-Nasiry S, Hanssens M, Luyten C, Pijnenborg R (2007) The use of Alamar Blue assay for quantitative analysis of viability, migration and invasion of choriocarcinoma cells. *Hum Reprod* 22:1304–1309



Chapter 3

Reassembly Design of Ferritin Cages

Yu Liu and Guanghua Zhao

Abstract

The ferritin family is distributed in nearly all organisms and protects them from iron-induced oxidative damage. Besides, its highly symmetrical structure and biochemical features make it an appealing material for biotechnological applications, such as building blocks for multidimensional assembly, templates for nano-reactors, and scaffolds for encapsulation and delivery of nutrients and drugs. Moreover, it is of great significance to construct ferritin variants with different properties, size, and shape to further broaden its application. In this chapter, we present a routine process of the ferritin redesign and the characterization method of the protein structure to provide a feasible scheme.

Key words Ferritin, Protein cages, Computational design, Characterization method, Self-assembly, Geometry regulation

1 Introduction

The ferritin family is one of the most studied protein nanocages, owing to its highly symmetrical structure and the specific functions performed in the organisms. To date, ferritin is founded in nearly all forms of life (animal, plant, bacteria) with the notable exception of yeast [1]. Structurally, ferritin consists of 24 subunits that can self-assemble into a hollow cage with octahedral (432) symmetry. Functionally, ferritin is an iron storage and detoxification protein function *in vivo* as an iron storage container, where it has the ability to encapsulate and sequester iron up to ~4500 iron atoms within the inner cavity, and such a form is named holoferitin. The holoferitin can be easily converted into its apo form when iron cores from the cavity were removed by reduction of Fe(III) and subsequent chelation of Fe(II). This apo form is termed “apoferitin” [2]. The tight packing arrangement of apoferitin leads to its high stability, and its shell-like structure remains intact upon heating at 80 °C for 10 min; moreover, apoferitin can also tolerate harsh denaturants such as guanidinium chloride and urea, making it an ideal biotemplate for the synthesis of various nanomaterials and a

nanovehicle for the encapsulation and delivery drugs and nutrients [3].

Ferritin was first discovered by Laufberger in 1937 who isolated and purified it from horse spleen by crystallization with cadmium salts [4]. The crystal structure of ferritin was determined much later in 1991 [5], which made great progress in understanding this molecule at the atomic level. Generally, the ferritin protein family can be subdivided into the maxi-ferritins and mini-ferritins. Maxi-ferritin consists of the classic ferritins and bacterio-ferritins which consist of 24 identical or similar subunits (~20 kDa) with an exterior diameter of about 12 nm and a hollow cavity of 8 nm. Each subunit is cylindrical with a length and width of 5 and 2.5 nm, respectively, and consisting of two antiparallel helix pairs (A, B & C, D), a short helix (E), and loops connecting fine helix [6]. For mammalian ferritins, it is usually composed of two types of subunit H-chains (~21.0 kDa) and L-chains (~19.5 kDa) with an amino acid sequence similarity of 55%. Each H-chain contains a dinuclear ferroxidase center, thus responsible for the rapid oxidation of ferrous ions by oxygen or H_2O_2 . Whereas, the L-chain lacks such a center but contains a putative nucleation site that is responsible for the slow oxidation and mineralization nucleation of Fe(II) [7]. Due to the different functions of the two subunits, the proportion of the two subunits is different in various tissues. For example, L-chains are rich in the tissues like the liver which is involved in the long-term storage of iron, while H-chains are rich in heart and muscle with more active iron metabolism. For the bacterio-ferritins, the quaternary structure is almost identical to the mammalian or other classic ferritins, with only one major difference that each bacterio-ferritin contains 12 heme groups in the C_2 interface [8]. The mini-ferritin is often referred to as Dps (DNA-binding protein from starved cells) which is first isolated from *E. coli* in 1992 [9] and its structure was determined later in 1998 [10]. The Dps subunit is a four-helix bundle without an E helix but contains a short helix on the BC loop. However, unlike Maxi-ferritin, 12 identical subunits of Dps assemble in a shell-like structure with tetrahedral point group symmetry.

Compared to other protein cages, ferritin nanocages have one major advantage: high selectivity for cancer cells overexpressing two receptors, namely, transferrin receptor 1 (TfR1) for H-chains ferritin and scavenger receptor class A member 5 (SCARA5) for L-chain ferritin [11, 12]. Such interesting property endows ferritins, especially recombinant human H-chain ferritin (rHuHF), with great potential as an ideal vehicle for encapsulation and delivery of drugs and bioactive nutrients. However, the lack of a clear drug-loading strategy is still a huge problem in the clinical application of ferritin nano-cages and all-natural ferritin molecules are limited to a single size and shape, which further limits their application.

Here, in this chapter, we describe several strategies that we have used to reassemble the design of the ferritin cages. For better drug loading, deletion of several amino acids involved in the AB-loop or C_4 interface yields ferritin disassembly and reassembly under benign experimental conditions [13, 14]. In addition, introducing a His₆ motif at the C_4 interface creates a dual switch (metal switch & pH switch) to control the ferritin assembly with a higher loading efficiency compared to the traditional method [15]. To regulate the geometry of ferritin cages, re-engineering the C_3 - C_4 interface by deleting six amino acid residues or inserting seven amino acid residues can convert the natural 24-mer ferritin into its 8-mer and 16-mer analogue [16, 17]. Meanwhile, the 8-mer ferritin can be further transformed to 16-mer, 24-mer, or 48-mer cages through disulfide bonds [18]. Finally, complete elimination of the ferritin C_3 - C_4 interface by deleting the D and E helix of each subunit facilitates the conversion of native 24-mer protein nanocage into 8-mer nanorings [19]. Since all the fabrication strategies focus on the re-engineering of a small number of amino acid residues, they are conceptually simple, and thus should, in principle, be applicable to design other proteins with high symmetry, leading to the generation of new protein materials with unexplored properties.

2 Materials

Prepare all the solutions with ultrapure water purified by a Millipore water purification system and analytical grade reagents at room temperature (unless indicated otherwise). Store the prepared solutions at 4 °C for up to 2 months.

2.1 Protein Interface Redesign

1. Software: UCSF Chimera. It can be downloaded from the official home page: <http://www.cgl.ucsf.edu/chimera/> (see Note 1).
2. Other commercially available programs such as PyMOL, VMD, and UCSF ChimeraX can perform the same function.

2.2 Protein Expression and Purification

1. Plasmid construction: PET-3a vector and fast site-directed mutagenesis kit.
2. Bacterial culture: BL21 (DE3) *E. coli* for the common protein expression and Rosetta (DE3) *E. coli* is suitable for the expression of eukaryotic genes in the prokaryotic system.
3. Luria-Bertani culture: Tryptone (10 g/L), Yeast extract (5 g/L), and NaCl (10 g/L).
4. Antibiotics and inducers: 250 mg/mL ampicillin stock solution and 1 M isopropyl-D-thiogalactopyranoside (IPTG) stock solution. The working concentrations are 100 µg/mL ampicillin and 400 µM IPTG.

5. Protein preparation: Buffer A 25 mM Tris-HCl (pH 8.0); buffer B 25 mM Tris-HCl and 150 mM NaCl (pH 8.0). 1 mg/mL RNase A & 0.5 mg/mL DNase I to avoid nucleic acid contamination and ammonium sulfate to precipitate the target protein.
6. ÄKTA purifier system to perform the protein purification with a Mono Q 5/50 GL and Superdex 200 10/300 GL gel filtration column.

2.3 Protein Characterization

1. SDS-PAGE Marker: (a) Rabbit Phosphorylase b (97.4 kDa); (b) Rabbit Actin (43.0 kDa); (c) Bovine Carbonic Anhydrase (31.0 kDa); (d) Trypsin Inhibitor (20.1 kDa); (e) Hen Egg White Lysozyme (14.4 kDa).
2. Native-PAGE Marker: (a) Thyroglobulin (669 kDa); (b) Horse Spleen Ferritin (440 kDa); (c) Catalase (232 kDa); (d) lactic dehydrogenase (140 kDa); (e) Bovine Serum Albumin (66 kDa).
3. The initial screening for crystallization conditions was carried out in 96-well hanging drop using commercial kits.
4. The X-ray experiments are performed at beamline BL17U, BL18U, and BL19U of Shanghai Synchrotron Radiation Facility (SSRF).
5. Related crystallography software for solving the crystal structures includes HKL3000, CCP4, PHENIX, and COOT.

3 Methods

3.1 Protein Interface Redesign

Maxi-ferritin such as rHuHF has four inter-subunit interfaces, termed C_2 , C_3 , C_4 , and $C_3\text{-}C_4$ (Fig. 1). According to our previous work, some amino acids on these interfaces play a key role in either changing the ferritin disassembly and reassembly behavior or regulating the geometry of the protein cages [16, 17]. Therefore, it is necessary to carefully analyze the inter-subunit interfaces and find those amino acid residues, which are located at the interfaces, before any protein design (*see Notes 2 and 3*).

1. Open UCSF Chimera and use the Favorites menu to show the command line. Then use the command: **open 2FHA** to fetch the structure of rHuHF from Protein Data Bank (PDB).
2. The rHuHF crystal has an fcc lattice arrangement (F432 space group) and only one subunit is shown. Use the command: **sym** to display the 24-mer molecular structure of rHuHF.
3. The **Find Clashes/Contacts** tool helps you to analyze the contacts quickly. Select the residue or inter-subunit interfaces

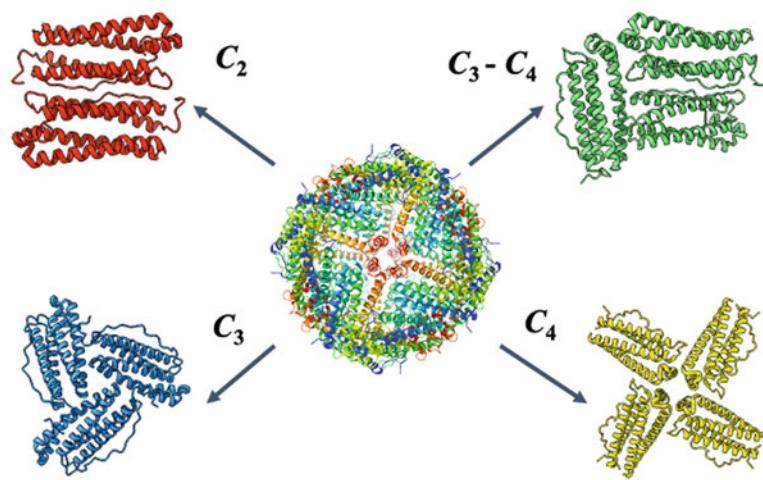


Fig. 1 Each ferritin nanocage consists of four inter-subunit interfaces responsible for its self-assembly

that you want to analyze and use the default contact criteria; the contacts between these atoms will display as pseudo-bonds.

4. The simulation of site-directed mutagenesis can be done by the **Rotamers** tool. It allows amino acid residues to be mutated into a different type and the sidechains rotamers can be viewed, evaluated, and incorporated into structures.
5. Other commercially available programs such as Rosetta can help you design the proteins more rationally (*see Note 4*).

3.2 Protein Preparation

1. Clone the cDNA encoding the full-length amino acid sequence of rHuHF into the pET-3a and verify the DNA sequencing. Then use the fast site-directed mutagenesis kit to perform the variants vector constructions (*see Note 5*).
2. After verifying by the DNA sequencing, transform the new plasmids harboring the desired genes into the *E. coli* BL21 (DE3) bacteria and grow on LB medium supplemented with 50.0 mg/L of ampicillin sodium at 37 °C. Induce the protein expression with 400 µM IPTG when the optical density reached an $\text{Abs}_{600\text{ nm}}$ of 0.6–0.8 (*see Note 6*).
3. Harvest the cells after 5 h of induction and suspend in buffer A. Next, sonicate the cells for 15 min on ice and then centrifuge at $10000 \times g$ for 20 min. Subsequently, incubate the lysate supernatant for 1 h at 37 °C with 1 mg/mL RNase A and 0.5 mg/mL DNase I (*see Note 7*).
4. Collect the supernatant of the resulting crude extract by centrifugation and fractionated by 60% saturation of ammonium sulfate and then resuspend in buffer A and dialyze against the same buffer (*see Note 8*).

5. Purify the protein by passing it through a Mono Q 5/50 GL column followed by gradient elution with 0–0.5 M NaCl and a Superdex 200 10/300 GL column equilibrated with buffer B (*see Note 9*).
6. Using SDS-PAGE to check the purity of the purified protein and determine the protein concentration according to the Lowry method with bovine serum albumin as standard (*see Note 10*).

3.3 Protein Characterization

Compared to the ferritin wild-type, the designed variants may have different properties, such as the molecular weight, geometry shape, and disassembly and reassembly behavior. Thus, a series of comprehensive characterization methods is necessary.

1. Electrophoresis. The Native-PAGE gel is a 4–20% gradient gel, and the SDS-PAGE uses a gel with a concentration of 20%. Mix the protein samples with the same volume of relative loading buffer and the SDS-PAGE samples need to be boiled in water for 5 min. After electrophoresis, stain the protein with Coomassie brilliant blue (R-250) (*see Notes 11 and 12*).
2. Transmission electron microscopy (TEM). Dilute the samples with appropriate buffers to obtain a concentration of about 0.4 mg/mL before being placed on carbon-coated copper grids for 2 min. Subsequently, remove the excess solution with filter paper and stain the samples using 2% uranyl acetate for 2 min.
3. Dynamic Light Scattering (DLS). Perform the DLS measurements at 25 °C using a Viscotek model 802 dynamic light-scattering instrument. And then calculate the size/hydrodynamic diameter distribution of prepared solutions using the OmniSIZE 2.0 software (*see Note 13*).
4. SEC-MALS analysis. Perform the SEC-MALS experiments performed using a DAWN-HELEOS II detector coupled to a Superdex 200 column in buffer B with a flow rate of 0.5 mL/min. Analysis of the data using ASTRA 6 software to determine the weight-averaged molecular mass (*see Note 14*).
5. Analytical ultracentrifugation sedimentation analyses (AUC). Perform the experiments at 10 °C in an XL-I analytical ultracentrifuge equipped with Rayleigh Interference detection (655 nm). Collect the interference profiles every 6 min and analysis the data with the software Sedfit 11.7, GUSSI, and SEDPHAT (*see Note 14*).

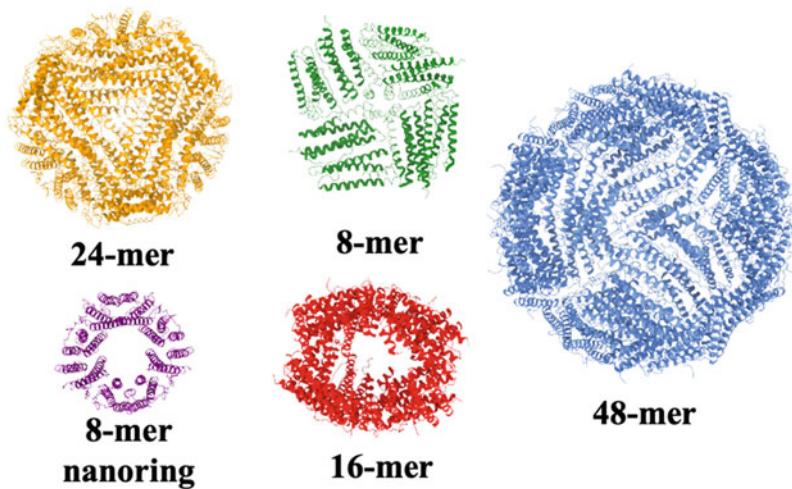


Fig. 2 Crystal structure of ferritin variants with different shapes and sizes

3.4 Protein Crystallization, Data Collection, and Structure Determination

Characterization of the structure of ferritin variants at the atomic level by X-ray crystallography (Fig. 2).

1. Concentrate the protein samples in buffer A to ~10 mg/mL.
2. Using the sitting drop vapor method with the commercial kits by mixing equal volumes of the protein sample and mother liquid to screen for crystallization conditions.
3. A suitable crystallization condition can be further optimized by hanging drop vapor.
4. Collect the X-ray diffraction data at Shanghai Synchrotron Radiation Facility (SSRF) (BL17U, BL18U and BL19U) with merging and scaling by HKL-3000.
5. Determine the structure by molecular replacement using coordinates of the wild-type ferritin as the initial model using the Phenix or CCP4 program package. Carry out following refinement and manual rebuilding by PHENIX and COOT, respectively (*see Note 15*).

4 Notes

1. UCSF Chimera is a powerful program for the interactive visualization and analysis of molecular structures and related data. It is available free of charge for non-commercial use. The detailed tutorials can be found on the home website, which is very convenient for beginners.
2. Redesign of the ferritin nanocage requires the destruction of some inherent driving forces between the subunit-subunit interactions (SSIs) at the key inter-subunit interfaces so that these subunits can rearrange to form new protein nanocage.

3. It is critical to find a balance between destroying the inherent SSIs and keeping subunit folding. Our previous results have shown that deletion of only one or two amino acid residues at the key protein interfaces such as the C_3 - C_4 interface may have a great effect on the whole protein structure.
4. An accurate in silico evaluation of SSIs will simplify the protein design process, preventing time, and effort-consuming experimental protocols. The program such as Rosetta provides multiple applications in protein design, protein docking, de novo modeling, and so on. More detailed information can be found on its website <https://www.rosettacommons.org>.
5. Deletion of a certain number of amino acid residues near the C-terminal can be achieved by introducing an amber mutation at the proper position. But you cannot do this near the N-terminal; instead, you have to re-construct the vector after the mutation.
6. Some variants may not be expressed normally in the previous vector, and you can try different vectors and competent cells. For example, the variant that incorporation of His₆ motifs into the flexible area (DE-loop) near the C_4 interface cannot express in the *E. coli* BL21 (DE3) but express normally in *E. coli* Rosetta (DE3).
7. After sonication, it is convenient and efficient to purify the protein through thermal treatment. The wild-type ferritin has high thermal stability that can heat at 60 °C for 10 min to remove some undesired protein participation. However, some ferritin variants may not have such thermal stability. Compared the heated samples with untreated samples using SDS-PAGE and Native-PAGE to determine whether the variant can be purified by this heating method.
8. The fractional precipitation of ammonium sulfate is also a good method for some ferritin variants which are difficult to purify.
9. A rough molecular weight calculation can be done by the Superdex 200 10/300 GL column with a set of protein standards in the same buffer during the protein purification. For example, the elution volume at ~10 mL in 25 mM Tris-HCl, pH 8.0, and 150 mM NaCl buffer corresponds to 24-mer ferritin with a molecular weight of ~500 kDa. Any change of pH or ion strength may affect the elution volume.
10. Some affinity chromatography methods, such as His-tag or GST-tag, can also be used to purify native ferritin and its variants.
11. After inserting or deleting some amino acids, the subunit molecular weight of ferritin can be preliminarily characterized by SDS-PAGE, and the more accurate molecular weight needs to be determined by MALDI-TOF-MS.

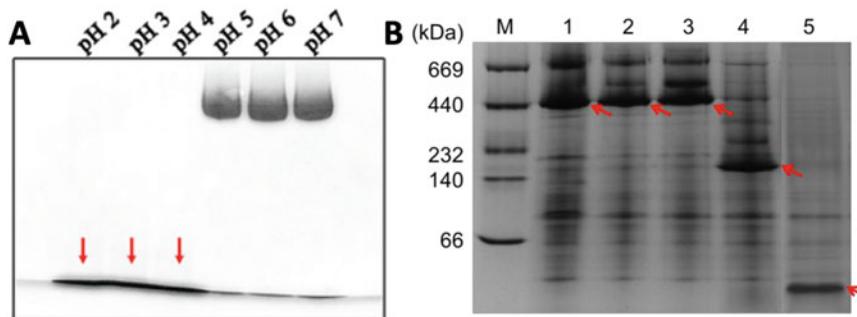


Fig. 3 Native-PAGE analysis of (a) ferritins under various pH conditions and (b) with different molecular weights. (Reprinted from Ref. 13 with permission from The Royal Society of Chemistry, copyright 2019 and Ref. 16 from American Chemical Society, copyright 2016)

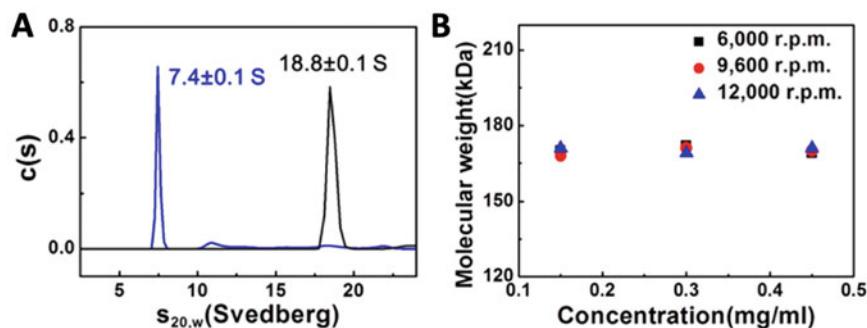


Fig. 4 Characterization of the molecular weight of ferritin variant with (a) analytical sedimentation analyses and (b) equilibrium sedimentation analyses. (Reprinted from ref. 16 with permission from American Chemical Society, copyright 2016)

12. In the case of studying the disassembly conditions of ferritin and its variants, Native-PAGE is a fast and intuitive method. Meanwhile, when the ferritin variants assemble into different assemble states, Native-PAGE can also be used for preliminary characterization (Fig. 3).
13. Dynamic light scattering is more accurate in measuring the hydrodynamic radius of shell-like molecules. Therefore, this method is only suitable for measuring the wild-type ferritin or the 48-mer ferritin variant that we constructed, not for the 8-mer bowl-like or ring-like ferritin variants.
14. Accurate characterization of the molecular weight of a new ferritin variant that assembles into a different assemble state requires a combination of various methods such as SEC-MALS analysis, analytical and equilibrium sedimentation analyses to confirm each other and make the results convincing (Fig. 4).
15. The crystal structure of most variants can be determined by molecular replacement using coordinates of the wild-type ferritin. However, some variants may need SAD, MAD, or other methods to solve the phase.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 31972018).

References

- Theil EC (2004) Iron, ferritin, and nutrition. *Annu Rev Nutr* 24:327–343
- Harrison PM, Arosio P (1996) The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim Biophys Acta Bioenerg* 1275:161–203
- Jutz G, Van RP, Santos Miranda B et al (2015) Ferritin: a versatile building block for bionanotechnology. *Chem Rev* 115:1653–1701
- Laufberger V (1937) Sur la cristallisation de la ferritine. *Bull Soc Chim Biol* 19:1575–1582
- Luzzago A, Levi S, Arosio P et al (1991) Solving the structure of human h-ferritin by genetically engineering intermolecular crystal contacts. *Nature* 349:541–5446
- Crichton RR, Declercq JP (2010) X-ray structures of ferritins and related proteins. *Biochim Biophys Acta – Gen Subj* 1800:706–718
- Crichton RR, Herbas A, Chavez-Alba O et al (1996) Identification of catalytic residues involved in iron uptake by L-chain ferritins. *J Biol Inorg Chem* 1:567–574
- Le BNE, Crow A, Murphy MEP et al (2010) Iron core mineralisation in prokaryotic ferritins. *Biochim Biophys Acta – Gen Subj* 1800:732–744
- Almiron M, Link A, Furlong D et al (1992) A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. *Genes Dev* 6:2646–2654
- Grant R, Filman D, Hogle J et al (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* 5: 294–30311
- Li L, Fang CJ, Ryan JC et al (2010) Binding and uptake of H-ferritin are mediated by human transferrin receptor-1. *Proc Natl Acad Sci U S A* 107:3505–3510
- Montemiglio LC, Testi C, Ceci P et al (2019) Cryo-EM structure of the human ferritin–transferrin receptor 1 complex. *Nat Commun* 10:1–8
- Wang W, Wang L, Li G et al (2019) AB loop engineered ferritin nanocages for drug loading under benign experimental conditions. *Chem Commun* 55:12344–12347
- Chen H, Zhang S, Xu C et al (2016) Engineering protein interfaces yields ferritin disassembly and reassembly under benign experimental conditions. *Chem Commun* 52:7402–7405
- Gu C, Zhang T, Lv C et al (2020) His-mediated reversible self-assembly of ferritin nanocages through two different switches for encapsulation of cargo molecules. *ACS Nano* 14(12):17080–17090
- Zhang S, Zang J, Zhang X et al (2016) “Silent” amino acid residues at key subunit interfaces regulate the geometry of protein nanocages. *ACS Nano* 10:10382–10388
- Zhang S, Zang J, Wang W et al (2016) Conversion of the native 24-mer ferritin nanocage into its non-native 16-mer analogue by insertion of extra amino acid residues. *Angew Chemie Int Ed* 55:16064–16070
- Zang J, Chen H, Zhang X et al (2019) Disulfide-mediated conversion of 8-mer bowl-like protein architecture into three different nanocages. *Nat Commun* 10:1–11
- Wang W, Wang L, Chen H et al (2018) Selective elimination of the key subunit interfaces facilitates conversion of Native 24-mer protein nanocage into 8-mer nanorings. *J Am Chem Soc* 140:14078–14081



Chapter 4

Protein Cages and Nanostructures Constructed from Protein Nanobuilding Blocks

Naoya Kobayashi and Ryoichi Arai

Abstract

Protein cages and nanostructures are promising biocompatible medical materials, such as vaccines and drug carriers. Recent advances in designed protein nanocages and nanostructures have opened up cutting-edge applications in the fields of synthetic biology and biopharmaceuticals. A simple approach for constructing self-assembling protein nanocages and nanostructures is the design of a fusion protein composed of two different proteins forming symmetric oligomers. In this chapter, we describe the design and methods of protein nanobuilding blocks (PN-Blocks) using a dimeric de novo protein WA20 to construct self-assembling protein cages and nanostructures. A protein nanobuilding block (PN-Block), WA20-foldon, was developed by fusing an intermolecularly folded dimeric de novo protein WA20 and a trimeric foldon domain from bacteriophage T4 fibritin. The WA20-foldon self-assembled into several oligomeric nanoarchitectures in multiples of 6-mer. De novo extender protein nanobuilding blocks (ePN-Blocks) were also developed by fusing tandemly two WA20 with various linkers, to construct self-assembling cyclized and extended chain-like nanostructures. These PN-Blocks would be useful for the construction of self-assembling protein cages and nanostructures and their potential applications in the future.

Key words Artificial proteins, Nanostructure, Protein cage, Protein complex, Protein nanobuilding block, Self-assembly

1 Introduction

Protein cages and nanoparticles are promising biocompatible medical materials, such as vaccines and drug carriers [1–6]. Advances in the design of self-assembling supramolecular protein complexes have been reported to form nanocages/nanoparticles and nanostructures/nanoarchitectures [7–12]. Recent advances in designed protein nanocages and nanostructures have opened up cutting-edge applications in the fields of synthetic biology and biopharmaceuticals [13–19]. A simple approach for constructing protein nanocages and nanostructures is the design and construction of a fusion protein composed of two different proteins forming symmetric oligomers. Padilla et al. described a general strategy for

using symmetry to construct protein nanomaterials “nanohedra” [20]. This is a simple and versatile approach to construct protein cage nanoparticles. Three different forms, tetrahedron (12-mer), triangular prism (18-mer), and cube (24-mer), were produced simultaneously [21]. Recently, in reference to the nanohedra strategy, a fusion protein between a pentameric Sm-like protein and a dimeric domain of anti-parallel coiled-coil protein from MyoX (MyoX-coil) was designed [22], and the fusion proteins uniformly self-assembled into the icosahedral 60-meric porous nanoparticle called TIP60 (truncated icosahedral protein composed of 60-mer fusion proteins) [23].

Moreover, in our previous study, utilizing a binary-patterned de novo protein WA20 with a characteristic bisecting-U dimeric structure of a four-helix bundle (PDB ID: 3VJF) [24] (*see Note 1*), protein nanobuilding blocks (PN-Blocks) were developed to construct self-assembling protein cages and nanostructures (Fig. 1). A PN-Block, WA20-foldon, was designed and constructed by fusing the dimeric de novo protein WA20 to a trimeric foldon domain of the T4 phage fibrin (Fig. 1a and b) [25]. The WA20-foldon formed several distinctive types of self-assembling cage-like polyhedral nanoarchitectures in multiples of 6-mer, including a barrel-like hexamer and a tetrahedrally shaped dodecamer (Fig. 1c). In addition, extender protein nanobuilding blocks (ePN-Blocks) were constructed by tandemly joining two copies of WA20 with various linkers (Fig. 1d) [26]. The ePN-Blocks assembled into cyclized and extended chain-type nanostructures (Fig. 1e and f). In this chapter, we focus on the design and construction of PN-Blocks using the intermolecularly folded dimeric de novo protein WA20.

Furthermore, the PN-Block strategy has been proposed as a systematic way for hierarchical design of supramolecular nanostructures as shown in Fig. 2. (1) the PN-Block strategy involves selection of appropriate PN-Block components, including dimeric de novo proteins, symmetrically oligomeric proteins and domains, and various linkers. (2) Subsequent combination and fusion of PN-Block components lead to the development of various PN-Blocks such as vertex-edge PN-Blocks (vPN-Blocks), extender PN-Blocks (ePN-Block), and stopper PN-Blocks (sPN-Blocks). (3) Combination and self-assembly of PN-Blocks produce homo-oligomeric and heterooligomeric PN-Block complexes. (4) Coordination and further assembly of PN-Block complexes produce higher-order supramolecular PN-Block nanoarchitectures in reference to the chemical design of 3D protein crystals [27]. In addition, functional PN-Blocks (lectin nano-blocks) were recently developed by fusing WA20 to a dimeric lectin [28]. The general and systematic strategy for hierarchical design with compatible modularity would expand the possibilities of PN-Blocks as artificial building-block molecules for a wide range of fields including nanotechnology, supramolecular chemistry, and synthetic biology.

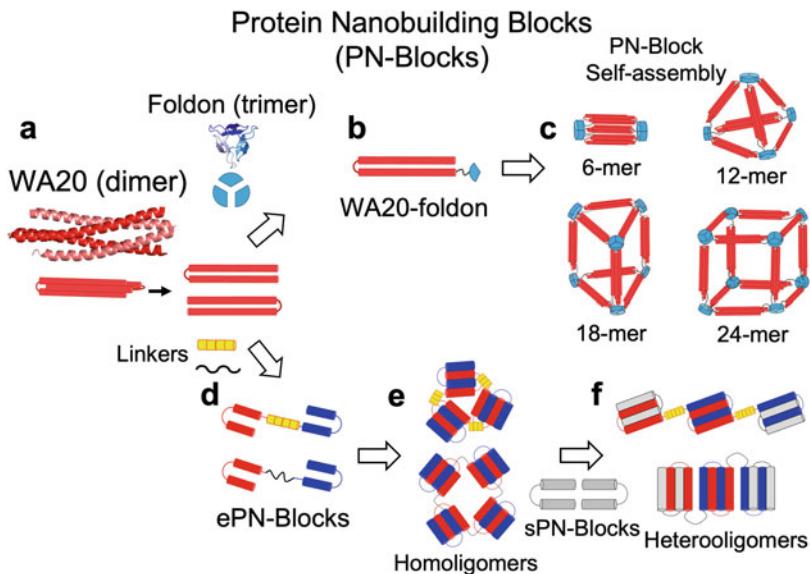


Fig. 1 Schematics of design of protein nanobuilding blocks (PN-Blocks) using a dimeric de novo protein WA20. (a) Ribbon representation and schematic of the intermolecularly folded dimeric de novo protein WA20 (PDB code 3VJF) [24] are shown in red. Schematic of trimeric foldon domains of T4 phage fibrin is shown in cyan. (b) Construction of the WA20-foldon fusion protein as a vertex-edge PN-Block [25]. (c) Schematics of nanoarchitectures designed by expected self-assembling complexes. WA20-foldon is expected to form highly symmetric oligomers in multiples of 6-mer because of the combination of the WA20 dimer and foldon trimer. (d) Construction of extender protein nanobuilding blocks (ePN-Blocks) [26]. The ePN-Blocks were constructed by tandemly fusing two de novo WA20 proteins using various linkers [32–34]. Helical and flexible linkers are shown as a yellow rod and a black curved line, respectively. The first and the second WA20 domains with the same amino acid sequence are shown in red and blue, respectively. (e) Homooligomeric self-assemblies of ePN-Blocks. (f) Reconstruction of heterooligomeric complexes from multicomponent extender PN-Blocks (ePN-Blocks) and stopper PN-Blocks (sPN-Blocks) by denaturation and refolding. (Adapted with permission from [25, 26]. Copyright 2015, 2018, American Chemical Society)

2 Materials

2.1 DNA Construction

1. High fidelity DNA polymerase for polymerase chain reaction (PCR): KOD-Plus-Neo DNA polymerase (Toyobo, Osaka, Japan).
2. Molecular biology reagents: restriction enzymes, DNA ligase, etc.
3. Protein expression plasmid vector: pET-3a, pET-32b(+).

2.2 Protein Expression and Purification

1. *Escherichia coli* for protein expression: BL21 Star (DE3).
2. Culture medium: LB broth.
3. Antibiotic: ampicillin sodium salt.

4. Expression inducer reagent: isopropyl β -D-1-thiogalactopyranoside (IPTG).
5. Reagents for buffer: L-arginine hydrochloride (ArgHCl), 2-[4-(2-Hydroxyethyl)-1-piperazinyl]ethanesulfonic acid (HEPES), glycerol, guanidine hydrochloride (GdnHCl), imidazole, KH_2PO_4 , NaCl , Na_2HPO_4 , NaOH.
6. Affinity resin for immobilized metal ion affinity chromatography (IMAC) purification: HisTrap HP column (Cytiva, Little Chalfont, Buckinghamshire, UK), TALON metal affinity resin (Clontech, Takara Bio, Mountain View, CA, USA).
7. Centrifugal filters for protein concentration: Amicon Ultra centrifugal filters (Merck Millipore, Darmstadt, Germany).
8. Column for size-exclusion chromatography: Superdex 200 Increase 10/300 GL column (Cytiva), HiLoad 16/600 Superdex 200 pg (Cytiva).

2.3 Equipment and Facility

1. Equipment for size exclusion chromatography (SEC)–multi-angle light scattering (SEC-MALS) experiments: 1260 Infinity HPLC system equipped with a Superdex 200 Increase 10/300 GL column (Cytiva), which was connected in line with a mini-DAWN TREOS multi-angle static light scattering detector (Wyatt Technology, Santa Barbara, CA, USA). The data were analyzed using ASTRA 6 software (Wyatt Technology).
2. Facility for Small-Angle X-ray Scattering (SAXS) experiments: synchrotron facility such as Photon Factory beamline (BL-10C or BL-6A) [29, 30] (KEK, Tsukuba, Japan) or laboratory SAXS systems (e.g., Anton Paar, Graz, Austria; RIGAKU, Tokyo, Japan).

3 Methods

3.1 Construction of Polyhedral Protein Nanobuilding Block (PN-Block)

3.1.1 Design of Polyhedral PN-Block (WA20-Foldon)

1. Image shapes and think building blocks to construct polyhedral cages and nanostructures. Polyhedra form with edges and vertexes. As shown in Fig. 2, to construct self-assembling cage-like polyhedral nanoarchitectures, the design of vertex-edge PN-Blocks (vPN-Blocks) is notably a versatile and powerful approach as a geometrically based building block to construct several symmetric polyhedra with three edges from one vertex, such as a tetrahedron, hexahedron, and dodecahedron.
2. Search appropriate proteins structures of symmetric oligomers in protein data bank (PDB). In this case, as shown in Fig. 1a, a dimeric de novo protein WA20 (PDB code 3VJF) [24] and a trimeric foldon domain (PDB code 1RFO) [31] (*see Note 2*) of bacteriophage T4 fibrin were selected for an edge part and a vertex part of polyhedra, respectively.

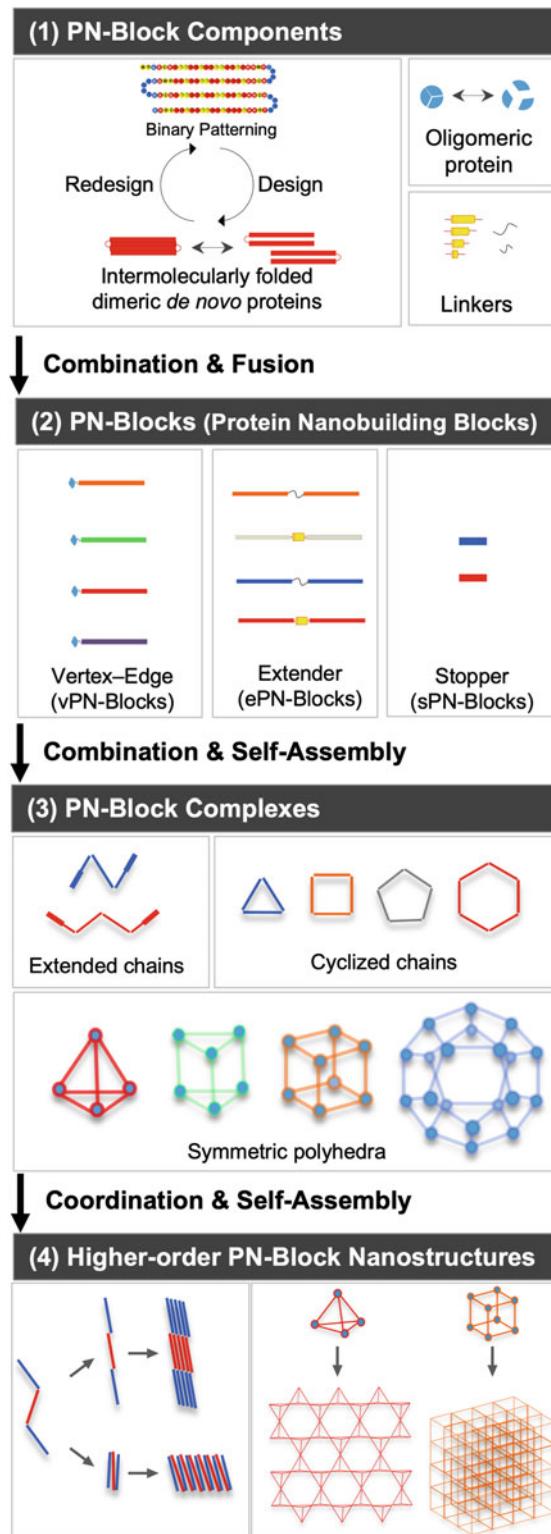


Fig. 2 Schematics of hierarchical design of supramolecular nanostructures using the PN-Block strategy. (1) The PN-Block strategy involves selection of various

3. Design a PN-Block fusion protein (WA20-foldon) by fusing the dimeric de novo protein WA20 and the trimeric foldon domain with a linker such as a short peptide linker (Fig. 1b). In this case, alanine-rich short linker (KLAAA) was used. In the stable self-assembling complexes of WA20-foldon, it was expected to form several oligomers in multiples of 6-mer because of the combination of the WA20 dimer and foldon trimer (Fig. 1c) [25].

3.1.2 Construction of Expression Plasmid of WA20-Foldon

1. Prepare DNA fragments encoding the target protein domains by PCR from an appropriate plasmid or gene library using high-fidelity DNA polymerase. In this case, the DNA fragments encoding the de novo protein WA20 was prepared from a plasmid pET-3a_WA20 [24] by PCR using KOD-Plus-Neo DNA polymerase. The DNA fragment encoding the foldon domain (residues 458–483 in T4 phage fibrin) was prepared by annealing and extension reactions with synthesized oligonucleotides (see Note 3).
2. Insert the DNA fragment into a protein expression plasmid pET vector by digestion with appropriate restriction enzymes and ligation with DNA ligase to construct a protein expression plasmid pET_WA20-foldon (Fig. 3a) [25] (see Notes 3 and 4).

3.1.3 Protein Expression of WA20-Foldon

1. Culture *E. coli* BL21 Star (DE3) harboring the plasmid pET_WA20-foldon in LB broth (Lennox) with 50 µg/mL ampicillin sodium salt at 37 °C.
2. Induce the expression of the WA20-foldon protein with 0.2 mM IPTG at OD₆₀₀ (optical density at 600 nm) = ~0.8, and cells were further cultured for 3–4 h at 37 °C.
3. Harvest *E. coli* cells from the culture by centrifugation.

3.1.4 Purification of WA20-Foldon

1. Extract proteins from the harvested cells by sonication in a lysis buffer (50 mM sodium phosphate buffer (pH 7.0) containing 300 mM NaCl, 10% glycerol).
2. Separate soluble lysate by centrifugation.

←

Fig. 2 (continued) PN-Block components, including intermolecularly folded dimeric de novo proteins by binary pattern design, symmetrically oligomeric proteins and domains, and various linkers. (2) Combination and fusion of PN-Block components produce various PN-Blocks such as vertex-edge PN-Blocks (vPN-Blocks), extender PN-Blocks (ePN-Blocks), and stopper PN-Blocks (sPN-Blocks). (3) Combination and self-assembly of PN-Blocks produce various PN-Block complexes. (4) Coordination and further self-assembly of PN-Block complexes produce higher-order PN-Block supramolecular nanostructures

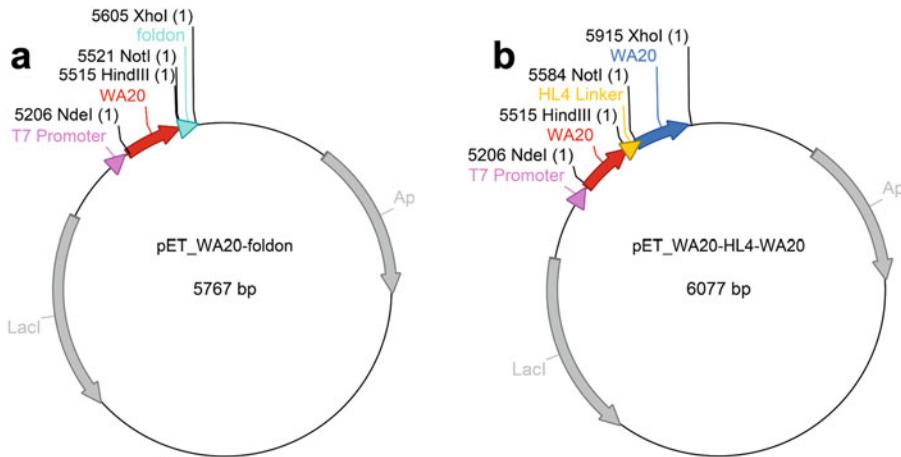


Fig. 3 Plasmid maps of (a) pET_WA20-foldon [25] and (b) pET_WA20-HL4-WA20 [26] for expression of the PN-Block fusion proteins. The graphics were created using A plasmid Editor (ApE) [57]

3. Purify the WA20-foldon protein by immobilized metal ion affinity chromatography (IMAC) with a HisTrap HP column and eluted using a linear gradient of imidazole (equilibration buffer: 20 mM sodium phosphate buffer (pH 7.4) containing 500 mM NaCl, 200 mM ArgHCl, 10% glycerol, and 20 mM imidazole; elution buffer: 20 mM sodium phosphate buffer (pH 7.4) containing 500 mM NaCl, 200 mM ArgHCl, 10% glycerol, and 500 mM imidazole).
4. Concentrate the protein samples with Amicon Ultra centrifugal filters.
5. Purify repeatedly each form of the WA20-foldon protein by size exclusion chromatography (SEC) (20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 200 mM L-ArgHCl, and 10% glycerol) with HiLoad 16/600 Superdex 200 pg and Superdex 200 Increase 10/300 GL columns (Cytiva).

3.2 Construction of Extender Protein Nanobuilding Blocks (ePN-Blocks)

3.2.1 Design of ePN-Blocks

1. Image shapes and think building blocks to construct target structures such as extended and cyclized chain-like nanostructures.
2. Search appropriate oligomeric protein structures in protein data bank (PDB). In this case, the dimeric de novo protein WA20 (PDB code 3VJF) [24] was selected for a PN-Block component as shown in Fig. 1a.
3. Design ePN-Blocks by tandemly fusing two de novo WA20 proteins with linkers of various type and length to construct chain-like oligomers in this case (Fig. 1d) [26]. The two WA20 domains were fused with helical linkers (HL: (EAAAK)_n, n = 2–5) or flexible linkers (FL: (GGGGS)_n, n = 3, 4) [32–

[34]. Because WA20 forms a stable intermolecularly folded dimeric structure [24], ePN-Blocks were expected to self-assemble into cyclized chain-like homooligomers (Fig. 1e).

3.2.2 Construction of Protein Expression

Plasmids of ePN-Blocks

1. Prepare DNA fragments encoding the target protein domains by polymerase chain reaction (PCR) from an appropriate plasmid or gene library using high-fidelity DNA polymerase. In this case, the DNA fragments encoding the de novo protein WA20 was prepared from plasmid pET-3a_WA20 [24] by PCR using KOD-Plus-Neo DNA polymerase. The DNA fragments encoding the various linker genes (HL2, HL3, HL4, FL3, and FL4) [32, 34] were prepared by digestion of the plasmid pET32_EBFP-linker-EGFP [32] with appropriate restriction enzymes (*Hind*III and *Not*I).
2. Insert the DNA fragments into a protein expression plasmid vector such as pET vector by digestion with appropriate restriction enzymes and ligation with DNA ligase to construct protein expression plasmids pET_WA20-linker-WA20 (Fig. 3b) (see Note 4).

3.2.3 Protein Expression of ePN-Blocks

1. Culture *E. coli* BL21 Star (DE3) harboring the protein expression plasmid pET_WA20-linker-WA20 in LB broth (Lennox) with 50 µg/mL ampicillin sodium salt at 37 °C.
2. Induce the expression of the ePN-Block (WA20-linker-WA20) protein with 0.2 mM IPTG at OD₆₀₀ = ~0.8, and cells were further cultured for 3–4 h at 37 °C.
3. Harvest *E. coli* cells from the culture by centrifugation.

3.2.4 Purification of ePN-Blocks

1. Extract proteins from the harvested cells by sonication in a lysis buffer (50 mM sodium phosphate buffer (pH 7.0) containing 300 mM NaCl, 10% glycerol).
2. Separate soluble lysate by centrifugation.
3. Purify the ePN-Block protein by IMAC with TALON metal affinity resin. The equilibration/wash buffer contains 50 mM sodium phosphate buffer (pH 7.0) and 300 mM NaCl, and the elution buffer contains 50 mM sodium phosphate buffer (pH 7.0), 300 mM NaCl, 10% glycerol, and 250 mM imidazole (see Note 5).

3.3 Reconstruction of Heterooligomeric Complexes of ePN-Block and sPN-Block

3.3.1 Denaturation and Refolding of ePN-Block and sPN-Block

To expand the possibilities of the PN-Block strategy, multicomponent PN-Block complexes (esPN-Blocks) were reconstructed from extender PN-Block (ePN-Block) and stopper PN-Block (sPN-Block, i.e., WA20 protein) by denaturation and refolding [26].

1. Mix the ePN-Block protein with the sPN-Block protein (WA20) and denature them in 6 M GdnHCl for 3 h at 25 °C

in 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl and 10% glycerol.

2. Dialyze the denatured proteins three times for refolding for ~4 h against 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 10% glycerol, and 200 mM ArgHCl.

3.3.2 Purification of esPN-Block Complexes

1. Concentrate the refolded protein samples with Amicon Ultra centrifugal filters.
2. Purify the concentrated samples of refolded ePN-Block and sPN-Block heterocomplexes (esPN-Blocks) in 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 10% glycerol, and 200 mM ArgHCl using size exclusion chromatography (SEC) with a Superdex 200 Increase 10/300 GL column, and collect eluted fractions.

3.4 Analyses of PN-Blocks

3.4.1 Polyacrylamide Gel Electrophoresis (PAGE)

1. Perform sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis according to the standard Laemmli procedure to check purity of protein samples.
2. Prepare the gels and buffer without SDS and perform native PAGE analysis to easily and roughly evaluate formation of complexes.

3.4.2 Size Exclusion Chromatography–Multi-Angle Light Scattering (SEC–MALS) Analysis

1. Perform SEC–MALS experiments to determine the absolute molecular weights using a 1260 Infinity HPLC system (Agilent Technologies) equipped with a Superdex 200 Increase 10/300 GL column, which was connected in line with a miniDAWN TREOS multi-angle static light scattering detector (Wyatt Technology). The data were collected in phosphate buffered saline (PBS, pH 7.4: 1 mM KH₂PO₄, 3 mM Na₂HPO₄, and 155 mM NaCl) at 20 °C.
2. Analyze the obtained data using ASTRA 6 software (Wyatt Technology). The $d\eta/dc$ value (0.185 mL/g) was generally used for proteins, and the extinction coefficient for the PN-Block proteins was calculated from the amino acid sequence [35].

3.4.3 Small-Angle X-Ray Scattering (SAXS) Analysis

A typical procedure for SAXS measurement and analysis is as follows:

1. Perform SAXS experiments on protein complexes in 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 200 mM ArgHCl, and 10% glycerol at 20 °C using synchrotron radiation ($\lambda = 0.15$ nm) with a PILATUS3 detector (Dectris, Baden, Switzerland) at a synchrotron facility beamline (Photon Factory BL-10C or BL-6A, KEK, Tsukuba, Japan) [29, 30] to estimate molecular size and shape.

2. Integrate two-dimensional scattering images into one-dimensional scattering intensities ($I(q)$) as a function of the magnitude of the scattering vector $q = (4\pi/\lambda) \sin(\theta/2)$ using SAngler [36], where θ is the total scattering angle.
3. Calculate $p(r)$ for protein particles using the indirect Fourier transformation (IFT) technique with a virtually model-free routine [37–39] (see Note 6).
4. Estimate forward scattering intensity $I(q \rightarrow 0)$ and the radius of gyration R_g using the Guinier approximation [38].
5. Estimate molecular weights of protein complexes with a molecular weight reference standard sample (e.g., lysozyme or ovalbumin) because forward scattering intensity normalized by protein concentration (mg/mL), $I(q \rightarrow 0)/c$, is proportional to weight-average molecular mass (M_w) assuming that proteins have practically identical scattering length density and specific volume and that the structure factor $S(q) \approx 1$ for dilute samples.

3.4.4 Low-Resolution Shape Modeling of PN-Block Complexes Based on SAXS Analysis

The low-resolution dummy atom models can be constructed from the SAXS data as following using ab initio shape modeling programs in the ATSAS program suite [40] for small-angle scattering data analysis from biological macromolecules (Fig. 4).

1. Perform calculations of rapid ab initio shape determination ten times using DAMMIF [41].
2. Align and average the generated models using DAMAVER [42].
3. Modified the DAMAVER model with fixed core by DAMSTART.
4. Perform further refinement of the model using DAMMIN [43].

3.4.5 Rigid-Body Modeling of PN-Block Complexes Based on SAXS Analysis

The rigid-body models of protein complexes can be constructed from the SAXS data based on high-resolution structures of the protein domains in PDB (Fig. 5).

1. Construct the rigid-body models. In the case of WA20-foldon, construct the oligomeric structures using the program COOT [44] based on the crystal structure of WA20 (PDB code 3VJF) [24] and the solution structure of the foldon domain (PDB code 1RFO) [31] with a consideration of their N- and C-terminal directions and two- and three-fold symmetries.
2. Refine the rigid-body models manually and iteratively using the program COOT to minimize differences in the $p(r)$ and $I(q)$ calculated from the models and those obtained from SAXS experiments. The program CRYSTAL [45] in the ATSAS program suite [40] can be used for evaluating the solution scattering from the models and fitting it to experimental scattering curves with the χ^2 value [46].

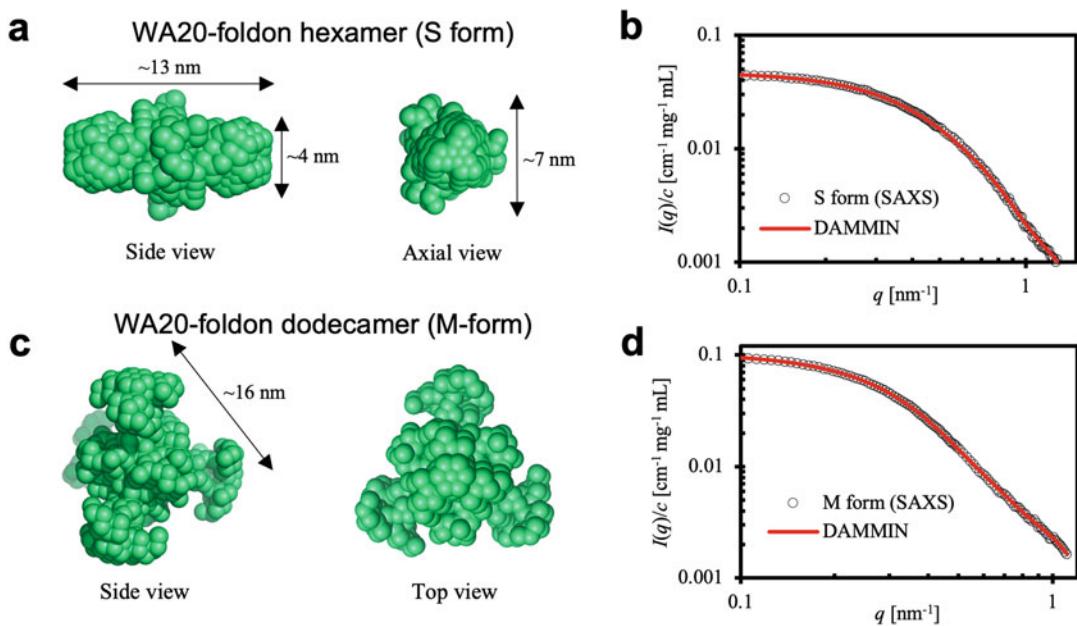


Fig. 4 Low-resolution shape modeling of WA20-foldon complexes based on SAXS analysis [25]. (a) The dummy atom model shape of the S form hexamer of WA20-foldon reconstructed from the SAXS data using the ab initio modeling program DAMMIN [43]. (b) The concentration-normalized SAXS intensity $I(q)/c$ of the S form hexamer of WA20-foldon (black open circle) and that optimized by the DAMMIN procedure (red line). (c) The dummy atom model shape of the M form dodecamer of WA20-foldon reconstructed from the SAXS data using the program DAMMIN. (d) The concentration-normalized SAXS intensity $I(q)/c$ of the M form dodecamer of the WA20-foldon (black open circle) and that optimized by the DAMMIN procedure (red line). (Adapted with permission from [25]. Copyright 2015, American Chemical Society)

3. Perform further refinement of the rigid-body models using some programs such as FoXS, MultiFoXS [47], and CORAL [48] in ATSAS [40] as appropriate.

4 Notes

1. The de novo protein WA20 is one of de novo proteins obtained from a library of binary patterned four-helix bundles [49]. The binary code strategy was developed to construct libraries of novel polypeptides (de novo proteins) that would fold into predetermined structures as a semirational approach for artificial protein design [50]. Using secondary structure motifs with binary patterns of polar and nonpolar residues, de novo proteins with α -helices or β -sheets have been successfully created without reference to natural protein sequences [50]. In addition, recently, hyperstable WA20 mutants were successfully designed [51, 52] to develop thermostable PN-Blocks.

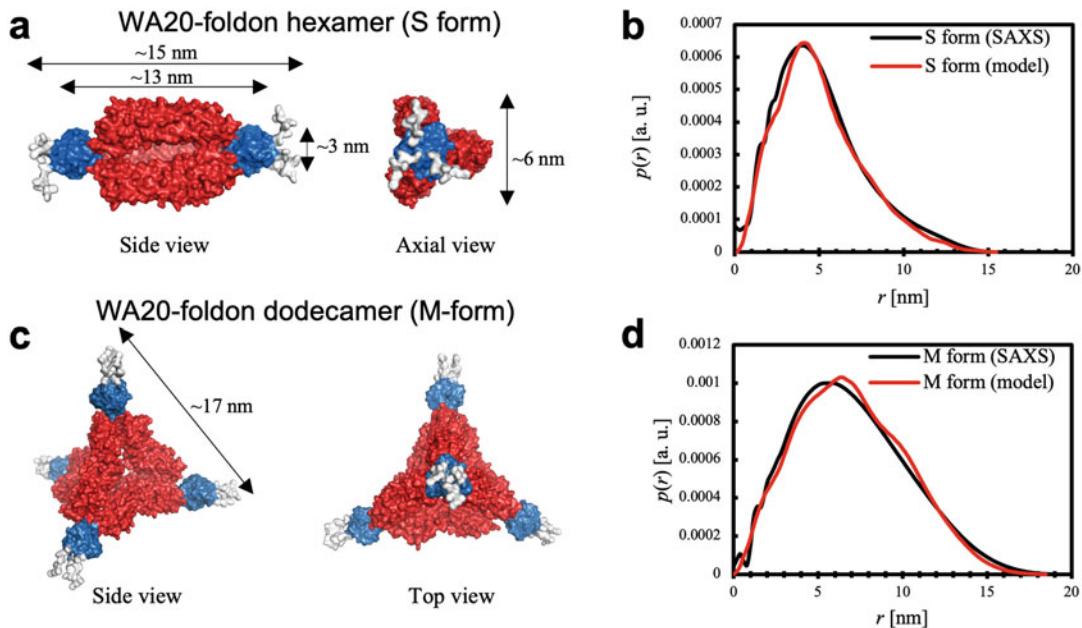


Fig. 5 Rigid-body modeling of WA20-foldon complexes based on SAXS analysis [25]. (a) The rigid-body model structure of the S form hexamer of the WA20-foldon. The domains of the WA20 (PDB code 3VJF) [24], foldon (PDB code 1RFO) [31], and His₆ tag are shown in red, blue, and light gray, respectively. (b) The pair-distance distribution function $p(r)$ of the S form hexamer of the WA20-foldon as obtained by the SAXS experiment (black line) and that simulated from the rigid-body model structure (red line). (c) The rigid-body model structure of the M form dodecamer of WA20-foldon. The domains are shown in the same colors as in (a). (d) The pair-distance distribution function $p(r)$ of the M form dodecamer of the WA20-foldon as obtained by the SAXS experiment (black line) and that simulated from the rigid-body model structure (red line). (Adapted with permission from [25]. Copyright 2015, American Chemical Society)

2. The foldon domain of bacteriophage T4 fibrin [53] consists of only 26 residues. The foldon domain is suitable for a trimeric vertex part because its stable trimerization by fast folding [31] and its application to the construction of engineered bio-nanotubes [54] were reported.
3. The amplified fragment was digested by *Nde*I and *Hind*III and cloned into pET32b(+) between the *Nde*I and *Hind*III sites to construct the plasmid pET_WA20 (a Trx tag was removed and replaced with WA20). The DNA fragment encoding the foldon domain (amino acid residues 458–483 in T4 phage fibrin) was prepared by annealing and extension reactions with the two synthesized oligonucleotides (Foldon_HindIII-NotI_FW: G GGGCAAAGCTTGCAGGCCGTATATTCTGAAGCTC CAAGAGATGGGCAAGCGTACGTTCGTAAAGATG ; Foldon_XhoI_RV: GGGGCCCTCGAGAAAGGTAGAAAG TAATACCCATTGCCATCTTACGAACGTACGCT) using KOD-Plus-Neo DNA polymerase. The DNA fragment encoding the foldon domain digested *Hind*III and *Xho*I and

cloned into pET_WA20 between the *Hind*III and *Xba*I sites to give the expression plasmid pET_WA20-foldon [25]. The amino acid sequence of the WA20-foldon fusion protein with a His₆ tag at the C terminal.

4. Alternatively, restriction-free cloning methods [55, 56] can be used for plasmid construction instead of the conventional method by restriction enzymes and DNA ligation. Also, gene synthesis services can be utilized for not only preparing DNA fragments but also constructing a cloned protein expression plasmid.
5. Because many histidine residues are exposed on the surface of the WA20 structure [24], WA20 and ePN-Block proteins can bind to TALON metal affinity resin even without a His-tag.
6. The scattering intensity for a colloidal dispersion is generally given by the product of the form factor $P(q)$ and structure factor $S(q)$. Hence, $I(q) = n P(q) S(q)$, where n is the number density of the particle. In the present experiments, the structure factor was almost at unity ($I(q) \approx n P(q)$), because interparticle interactions such as the excluded volume effect and electrostatic interactions can be neglected at low protein and high salt concentrations. Thus, the form factor is given by the Fourier transformation of the pair-distance distribution function $p(r)$, which expresses the size and shape of the particle as follows:

$$P(q) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin qr}{qr} dr,$$

where D_{\max} is the maximum intraparticle distance. The indirect Fourier transformation (IFT) technique was used to calculate $p(r)$ for particles with a virtually model-free routine [37–39]. Forward scattering intensity $I(q \rightarrow 0)$ was extrapolated from SAXS data, and the radius of gyration R_g was estimated using the Guinier approximation [38].

Acknowledgments

We thank Prof. Michael Hecht at Princeton University for the kind gift of the expression plasmid of WA20. We thank Prof. Takaaki Sato, Mr. Kouichi Inano, and Dr. Keiichi Yanase at Shinshu University for help in SAXS experiments and analysis. We thank Prof. Nobutaka Shimizu and Photon Factory (PF) staff for help in synchrotron SAXS experiments which were performed at PF, KEK under the approval of PF program advisory committee (Proposal No. 2014G111, 2016G153, and 2016G606). We thank Prof. Nobuyasu Koga, Dr. Rie Koga, and Dr. Takahiro Kosugi at the

Institute for Molecular Science (IMS) for help in SEC-MALS experiments. This work was supported by Joint Research of IMS (IMS program No. 603, 206, 221). This work was supported by JSPS Research Fellowships (DC2) and JSPS KAKENHI Grant Numbers JP14J10185 and JP16H06837 to N.K., and JSPS KAKENHI Grant Numbers JP24113707, JP24780097, JP16K05841, JP16H00761, JP17KK0104, and JP19H02522 to R.A.

References

- Herrera Estrada LP, Champion JA (2015) Protein nanoparticles for therapeutic protein delivery. *Biomater Sci* 3:787–799
- Zhang Y, Ardejani MS, Orner BP (2016) Design and applications of protein-cage-based nanomaterials. *Chem Asian J* 11:2814–2828
- Bhaskar S, Lim S (2017) Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Mater* 9:e371
- Heddle JG, Chakraborti S, Iwasaki K (2017) Natural and artificial protein cages: design, structure and therapeutic applications. *Curr Opin Struct Biol* 43:148–155
- Diaz D, Care A, Sunna A (2018) Bioengineering strategies for protein-based nanoparticles. *Genes (Basel)* 9:370
- Neek M, Kim TI, Wang SW (2019) Protein-based nanoparticles in cancer vaccine development. *Nanomedicine: NBM* 15:164–174
- Yeates TO, Liu Y, Laniado J (2016) The design of symmetric protein nanomaterials comes of age in theory and practice. *Curr Opin Struct Biol* 39:134–143
- Kobayashi N, Arai R (2017) Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Curr Opin Biotech* 46:57–65
- Arai R (2018) Hierarchical design of artificial proteins and complexes toward synthetic structural biology. *Biophys Rev* 10:391–410
- Miyamoto T, Hayashi Y, Yoshida K, Watanabe H, Uchihashi T, Yonezawa K, Shimizu N, Kamikubo H, Hirota S (2019) Construction of a quadrangular tetramer and a cage-like hexamer from three-helix bundle-linked fusion proteins. *ACS Synth Biol* 8: 1112–1120
- Stupka I, Heddle JG (2020) Artificial protein cages – inspiration, construction, and observation. *Curr Opin Struct Biol* 64:66–73
- Laniado J, Cannon KA, Miller JE, Sawaya MR, McNamara DE, Yeates TO (2021) Geometric lessons and design strategies for nanoscale protein cages. *ACS Nano* 15:4277–4286
- Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A, Ravichandran R, Pun SH, King NP, Baker D (2017) Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* 552:415–420
- Terasaka N, Azuma Y, Hilvert D (2018) Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc Natl Acad Sci U S A* 115:5432–5437
- Marcandalli J, Fiala B, Ols S, Perotti M, de van der Schueren W, Snijder J, Hodge E, Benhaim M, Ravichandran R, Carter L, Sheffler W, Brunner L, Lawrenz M, Dubois P, Lanzavecchia A, Sallusto F, Lee KK, Veesler D, Correnti CE, Stewart LJ, Baker D, Lore K, Perez L, King NP (2019) Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus. *Cell* 176:1420–1431
- Edwardson TGW, Tetter S, Hilvert D (2020) Two-tier supramolecular encapsulation of small molecules in a protein cage. *Nat Commun* 11:5410
- Walls AC, Fiala B, Schafer A, Wrenn S, Pham MN, Murphy M, Tse LV, Shehata L, O'Connor MA, Chen C, Navarro MJ, Miranda MC, Pettie D, Ravichandran R, Kraft JC, Ogohara C, Palser A, Chalk S, Lee EC, Guerriero K, Kepf E, Chow CM, Sydeman C, Hodge EA, Brown B, Fuller JT, Dinnon KH 3rd, Gralinski LE, Leist SR, Gully KL, Lewis TB, Guttman M, Chu HY, Lee KK, Fuller DH, Baric RS, Kellam P, Carter L, Pepper M, Sheahan TP, Veesler D, King NP (2020) Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* 183:1367–1382
- Divine R, Dang HV, Ueda G, Fallas JA, Vulovic I, Sheffler W, Saini S, Zhao YT, Raj IX, Morawski PA, Jennewein MF, Homad LJ, Wan YH, Tooley MR, Seeger F, Etemadi A,

- Fahning ML, Lazarovits J, Roederer A, Walls AC, Stewart L, Mazloomi M, King NP, Campbell DJ, McGuire AT, Stamatatos L, Ruohola-Baker H, Mathieu J, Veesler D, Baker D (2021) Designed proteins assemble antibodies into modular nanocages. *Science*:372
19. Ben-Sasson AJ, Watson JL, Sheffler W, Johnson MC, Bittleston A, Somasundaram L, Decarreau J, Jiao F, Chen J, Mela I, Drabek AA, Jarrett SM, Blacklow SC, Kaminski CF, Hura GL, De Yoreo JJ, Kollman JM, Ruohola-Baker H, Derivery E, Baker D (2021) Design of biologically active binary protein 2D materials. *Nature* 589:468–473
20. Padilla JE, Colovos C, Yeates TO (2001) Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci U S A* 98:2217–2221
21. Lai YT, Reading E, Hura GL, Tsai KL, Laganowsky A, Asturias FJ, Tainer JA, Robinson CV, Yeates TO (2014) Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat Chem* 6:1065–1071
22. Kawakami N, Kondo H, Matsuzawa Y, Hayasaka K, Nasu E, Sasahara K, Arai R, Miyamoto K (2018) Design of hollow protein nanoparticles with modifiable interior and exterior surfaces. *Angew Chem Int Ed* 57:12400–12404
23. Obata J, Kawakami N, Tsutsumi A, Nasu E, Miyamoto K, Kikkawa M, Arai R (2021) Icosahedral 60-meric porous structure of designed supramolecular protein nanoparticle TIP60. *Chem Commun* 57:10226–10229
24. Arai R, Kobayashi N, Kimura A, Sato T, Matsuo K, Wang AF, Platt JM, Bradley LH, Hecht MH (2012) Domain-swapped dimeric structure of a stable and functional *de novo* four-helix bundle protein, WA20. *J Phys Chem B* 116:6789–6797
25. Kobayashi N, Yanase K, Sato T, Unzai S, Hecht MH, Arai R (2015) Self-assembling nanoarchitectures created from a protein nano-building block using an intermolecularly folded dimeric *de novo* protein. *J Am Chem Soc* 137: 11285–11293
26. Kobayashi N, Inano K, Sasahara K, Sato T, Miyazawa K, Fukuma T, Hecht MH, Song C, Murata K, Arai R (2018) Self-assembling supramolecular nanostructures constructed from *de novo* extender protein nanobuilding blocks. *ACS Synth Biol* 7:1381–1394
27. Sontz PA, Bailey JB, Ahn S, Tezcan FA (2015) A metal organic framework with spherical protein nodes: rational chemical design of 3D protein crystals. *J Am Chem Soc* 137:11598–11601
28. Irumagawa S, Hiemori K, Saito S, Tateno H, Arai R (2022) Self-assembling lectin nanoblock oligomers enhance binding avidity to glycans. *Int J Mol Sci* 23:676
29. Shimizu N, Mori T, Nagatani Y, Ohta H, Saijo S, Takagi H, Takahashi M, Yatabe K, Kosuge T, Igarashi N (2019) BL-10C, the small-angle x-ray scattering beamline at the photon factory. *AIP Conf Proc* 2054:060041
30. Takagi H, Igarashi N, Mori T, Saijyo S, Ohta H, Nagatani Y, Kosuge T, Shimizu N (2016) Upgrade of small angle x-ray scattering beamline BL-6A at the photon factory. *AIP Conf Proc* 1741:030018
31. Guthe S, Kapinos L, Moglich A, Meier S, Grzesiek S, Kiehaber T (2004) Very fast folding and association of a trimerization domain from bacteriophage T4 fibrin. *J Mol Biol* 337: 905–915
32. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng* 14:529–532
33. Arai R, Wriggers W, Nishikawa Y, Nagamune T, Fujisawa T (2004) Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. *Proteins* 57:829–838
34. Arai R (2021) Design of helical linkers for fusion proteins and protein-based nanostructures. *Methods Enzymol* 647:209–230
35. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption-coefficient of a protein. *Protein Sci* 4:2411–2423
36. Shimizu N, Yatabe K, Nagatani Y, Saijyo S, Kosuge T, Igarashi N (2016) Software development for analysis of small-angle X-ray scattering data. *AIP Conf Proc* 1741:050017
37. Glatter O (1980) Evaluation of small-angle scattering data from lamellar and cylindrical particles by the indirect Fourier transformation method. *J Appl Crystallogr* 13:577–584
38. Glatter O, Kratky O (1982) Small-angle X-ray scattering. Academic Press, New York
39. Brunner-Popela J, Glatter O (1997) Small-angle scattering of interacting particles .1. Basic principles of a global evaluation technique. *J Appl Crystallogr* 30:431–442
40. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, Svergun DI (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* 50:1212–1225

41. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342–346
42. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36:860–864
43. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76:2879–2886
44. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D* 66:486–501
45. Svergun D, Barberato C, Koch MHJ (1995) CRY SOL – a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773
46. Trewella J, Duff AP, Durand D, Gabel F, Guss JM, Hendrickson WA, Hura GL, Jacques DA, Kirby NM, Kwan AH, Perez J, Pollack L, Ryan TM, Sali A, Schneidman-Duhovny D, Schwede T, Svergun DI, Sugiyama M, Tainer JA, Vachette P, Westbrook J, Whitten AE (2017) 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr D* 73:710–728
47. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2016) FoXS, FoXS Dock and Multi-FoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* 44: W424–W429
48. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV, Svergun DI (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45:342–350
49. Patel SC, Bradley LH, Jinadasa SP, Hecht MH (2009) Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins. *Protein Sci* 18:1388–1400
50. Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) De novo proteins from designed combinatorial libraries. *Protein Sci* 13:1711–1723
51. Kimura N, Mochizuki K, Umezawa K, Hecht MH, Arai R (2020) Hyperstable *de novo* protein with a dimeric bisecting topology. *ACS Synth Biol* 9:254–259
52. Irumagawa S, Kobayashi K, Saito Y, Miyata T, Umetsu M, Kameda T, Arai R (2021) Rational thermostabilisation of four-helix bundle dimeric *de novo* proteins. *Sci Rep* 11:7526
53. Tao Y, Strelkov SV, Mesyanzhinov VV, Rossmann MG (1997) Structure of bacteriophage T4 fibrin: a segmented coiled coil and the role of the C-terminal domain. *Structure* 5:789–798
54. Yokoi N, Inaba H, Terauchi M, Stieg AZ, Sanghamitra NJ, Koshiyama T, Yutani K, Kanamaru S, Arisaka F, Hikage T, Suzuki A, Yamane T, Gimzewski JK, Watanabe Y, Kitagawa S, Ueno T (2010) Construction of robust bio-nanotubes using the controlled self-assembly of component proteins of bacteriophage T4. *Small* 6:1873–1879
55. van den Ent F, Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. *J Biochem Biophys Methods* 67:67–74
56. 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* 172:34–44
57. Davis MW, Jorgensen EM (2022) ApE, A plasmid Editor: a freely available DNA manipulation and visualization program. *Front Bioinform* 2:818619



Chapter 5

Preparation of Cage-Like Micellar Assemblies of Engineered Hemoproteins

Koji Oohora and Takashi Hayashi

Abstract

Natural protein assemblies have encouraged scientists to create large supramolecular systems consisting of various protein motifs. In the case of hemoproteins containing heme as a cofactor, several approaches have been reported to form artificial assemblies with various structures such as fibers, sheets, networks, and cages. This chapter describes the design, preparation, and characterization of cage-like micellar assemblies for chemically modified hemoproteins including hydrophilic protein units attached to hydrophobic molecules. Detailed procedures are described for constructing specific systems using cytochrome b_{562} and hexameric tyrosine-coordinated heme protein as hemoprotein units with heme-azobenzene conjugate and poly-*N*-isopropylacrylamide as attached molecules.

Key words Heme, Hemoprotein, Protein modification, Supramolecular assembly

1 Introduction

Artificial protein assemblies have been recognized as an emergent topic over the last decade [1, 2]. Many protein assemblies found in nature have inspired chemists and biologists to create new protein assemblies by artificial design. One general strategy involves the introduction of interprotein interaction sites onto the protein surfaces. For such assemblies, coordination, electrostatic, hydrogen bonding, and hydrophobic and host-guest interactions have been employed. Various protein units have been employed to develop artificial protein assembly systems [3–10], of which hemoproteins with a heme cofactor are among the most common [11]. For example, cytochrome *c* forms assemblies such as ring-shaped oligomers and fibrous polymers via domain-swapping and suitable mutations enhance the efficiency of formation of the assemblies [12–15]. Cytochrome *cb*₅₆₂, an engineered cytochrome b_{562} possessing *c*-type heme covalently linked to two Cys residues in the heme-binding site, and myoglobin generate domain-swapped assemblies

[16]. Cytochrome *c*₅₆₂ is also employed as a building unit to form metal ion-assisted assemblies following incorporation of coordination sites on the protein surface [17–22]. These systems demonstrate a series of 1D, 2D, and 3D assemblies with highly ordered periodic structures. Domain swapping and coordination bond-triggered assemblies both contribute to the formation of cage structures [16, 22]. In hemoproteins which have *b*-type heme bound into a protein matrix by reversible interactions, the heme–heme pocket interaction can be used to form the hemoprotein assemblies [23]. In this system, a synthetic heme with a maleimide or iodoacetamide group at the terminus of the heme-propionate side chain is covalently attached to a Cys residue introduced onto the protein surface by mutagenesis. The first demonstration of formation of a hemoprotein assembly using the heme–heme pocket interaction was reported using an H63C mutant of cytochrome *b*₅₆₂ [24]. Synthetic heme with an iodoacetamide group attached via a flexible linker was reacted with the mutant to construct heme-attached cytochrome *b*₅₆₂. Subsequent removal of heme was found to trigger the formation of a fibrous assembly of cytochrome *b*₅₆₂, which was evaluated by size exclusion chromatography and atomic force microscopy. We have further investigated a series of additional hemoprotein assembly systems using the heme–heme pocket interaction, including rigid helical, star-shaped, and network assemblies using the cytochrome *b*₅₆₂ unit [25–28]. In addition, myoglobin has been found useful as a building block for generating hemoprotein assemblies with unique oxygen-binding and stimuli-responsive materials [29]. Formation of a micellar-type cage-like structure was observed when the H63C mutant of cytochrome *b*₅₆₂ modified by synthetic heme with the iodoacetamide group via an azobenzene or stilbene linker was heated to temperature above 65 °C (see Fig. 1a) [30]. The micellar-type structure consists of the hydrophilic apo-form of the protein units and rigid hydrophobic heme-azobenzene or stilbene moieties, whereas a conventional fibrous assembly by heme–heme pocket interaction is formed at room temperature. Interestingly, this micellar-type cage-like structure is maintained as a meta-stable form even after cooling and slowly converts to the fibrous assembly over a period of 96 h at 25 °C.

Hexameric tyrosine-coordinated heme protein (HTHP) has also been found to be suitable for hemoprotein assemblies. The inherent function of HTHP is unknown but is believed to provide peroxidase, catalase, or heme-storage/carrier activity [31]. HTHP is a homo-hexamer with *C*₆-symmetric structure with each domain including a single heme molecule. In addition to this highly symmetric structure, its remarkable thermal stability ($T_m > 130$ °C) is an advantageous characteristic for a building block of a hemoprotein assembly. Using this protein as a core unit for branching, we demonstrated the formation of a star-shaped assembly from the apo-form of HTHP and cytochrome *b*₅₆₂ via the heme–heme

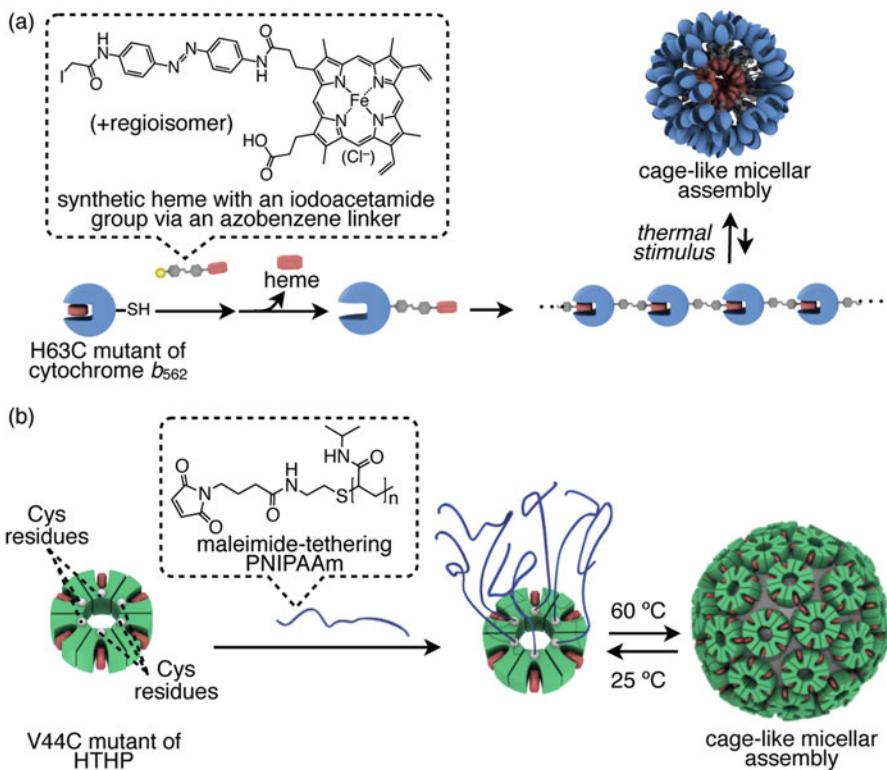


Fig. 1 Schematic representation of cage-like micellar assemblies of (a) covalently heme-attached cytochrome *b*₅₆₂ mutant and (b) poly-*N*-isopropylacrylamide-bound HTHP mutant

pocket interaction [27]. Furthermore, functional molecules with a maleimide group can be introduced at Cys residues mutated on the HTHP surface to form supramolecular assembly systems. For example, pyrene molecule-attached HTHP is assembled to generate a supramolecular dimer which is constructed from multimeric interprotein interactions between pyrene molecules [32]. FGG tripeptide-attached HTHP provides a two-dimensional sheet upon addition of cucurbit[8]uril, where two FGG tripeptide molecules are introduced into the cavity of one cucurbit[8]uril molecule with moderate affinity [33]. In addition, HTHP modified with poly-*N*-isopropylacrylamide (PNIPAAm), a thermal responsive polymer, yields a cage-like micellar assembly at 60 °C, where HTHP and PNIPAAm are hydrophilic and hydrophobic moieties, respectively (see Fig. 1b) [34]. Similar types of assemblies have been reported using reconstituted hemoproteins with artificial heme tethering synthetic polymers such as PNIPAAm and polystyrene [35, 36].

This chapter describes the details of practical procedures for producing two different types of cage-like micellar assemblies of hemoproteins (see Fig. 1). In particular, we demonstrate the design of the protein units and the preparation and characterization of the cage-like micellar structures.

2 Materials

All aqueous solutions should be prepared using deionized water with electrical resistivity above 18 MΩcm. Reagents of the highest available grade should be used as obtained from commercial sources. Typically, further purification is not required unless indicated otherwise. Detailed procedures to synthesize heme with an iodoacetamide group via an azobenzene linker are described in our previous paper [30].

2.1 Expression and Purification of Cytochrome *b*₅₆₂ Mutant

1. LB medium prepared by dissolution of 10 g tryptone, 5 g yeast extract, 10 g NaCl, and 200 μL of 5 M NaOH aqueous solution in 1 L of water. The medium is sterilized in an autoclave.
2. 2×YT medium prepared by dissolving 16 g tryptone, 10 g yeast extract, 5 g NaCl, and 200 μL of 5 M NaOH aqueous solution in 1 L of water. The medium is sterilized in an autoclave.
3. Plasmid coding the target protein. The cytochrome *b*₅₆₂ H63C mutant amino acid sequence is as follows:

ADLEDNMETLNDNLKVIEKADNAAQVKDALKM
RAAALDAQKATPPKLEDKSPDSPEMKDFRCGFIDLVG
QIDDALKLANEGKVKEAQAAEQLKTTRNAYHQKYR

The gene encoding this protein is inserted into the multi-cloning site of a pUC118 vector.

4. 100 mg/mL ampicillin aqueous solution.
5. 10 mM Tris–HCl buffer solution (pH 7.3).
6. CHCl₃.
7. 1 M HCl aqueous solution.
8. CM Sepharose Fast Flow (Cytiva).
9. Solution A: 50 mM KH₂PO₄ (pH 4.5) aqueous solution containing 0.1 mM ethylenediaminetetraacetic acid (EDTA).
10. Solution B: Solution A containing 50 mM KCl.
11. Solution C: Solution A containing 150 mM KCl.
12. Amicon stirred ultrafiltration cell with a 5 kDa molecular weight cut-off membrane.
13. Sephadex G-50 superfine (Sigma Aldrich).
14. 50 mM Tris–HCl buffer solution (pH 8.0) containing 0.1 mM EDTA.

2.2 Expression and Purification of HTHP Mutant

1. LB medium prepared by dissolution of 10 g tryptone, 5 g yeast extract, 10 g NaCl, and 200 μL of 5 M NaOH aqueous solution in 1 L of water. The medium is sterilized in an autoclave.

2. Plasmid coding the target protein. The V44C HTHP mutant amino acid sequence is:

SETWLPTLVTATPQEGFDLAVKLSRIAVKKTQP-
DAQVRDTLRACYEKDANALIAAVSAAVATHFQTIAAN-
DYWKD.

The gene encoding this protein is inserted into the multi-cloning site of a PDEST14 vector.

3. 100 mg/mL ampicillin aqueous solution.
4. Isopropyl- β -D-1-thiogalactopyranoside.
5. 10 mM Tris-HCl buffer solution (pH 8.0).
6. 1 mM hemin stock solution dissolved in an aqueous solution of 0.1 M NaOH.
7. DEAE Sepharose Fast Flow (Cytiva).
8. Solution D: 10 mM potassium phosphate buffer (pH 6.0).
9. Solution E: Solution D containing 250 mM NaCl.
10. Solution F: Solution D containing 375 mM NaCl.
11. Amicon stirred ultrafiltration cell with a 10 kDa molecular weight cut-off membrane.
12. HiPrep 16/60 Sephadryl S-200 HR column (Cytiva).
13. 100 mM potassium phosphate buffer (pH 7.0).

2.3 Modification of

Cytochrome b₅₆₂ Mutant by Synthetic Heme

1. 1 M dithiothreitol (DTT) aqueous solution.
2. HiTrap Desalting column (Cytiva).
3. 100 mM NaHCO₃ buffer (pH 9.0).
4. Tetrahydrofuran.
5. Synthetic heme with an iodoacetamide group via an azobenzene linker.
6. 2-butanone.
7. 50 mM Tris-HCl buffer (pH 7.3) containing 0.1 mM EDTA.
8. Water and acetonitrile containing 0.1% trifluoroacetic acid (TFA).
9. YMC-Pack Pro C18 reverse phase column (YMC).

2.4 Modification of

HTHP with Maleimide-Tethering PNIPAAm

1. 1 M DTT aqueous solution.
2. HiTrap Desalting column (Cytiva).
3. Maleimide-tethering PNIPAAm ($M_n = 5500$).
4. 100 mM potassium phosphate buffer (pH 7.0).
5. DEAE Sepharose Fast Flow (Cytiva).
6. 100 mM potassium phosphate buffer (pH 7.0) containing 0.5 M NaCl.

2.5 Characterization of Cage-like Micellar Structures

1. Aldehyde-modified carbon grid (C-SMART Lysine TEM carbon grid, Alliance Biosystems).
2. Staining solution: NANO-W (Nanoprobes: methylamine tungstate).
3. 0.1% (3-aminopropyl)triethoxysilane aqueous solution.
4. Mica.
5. 0.25% glutaraldehyde aqueous solution.

3 Methods

3.1 Design of Protein Mutants

To obtain the cage-like structure formed by the heme-attached cytochrome b_{562} mutant, the introduced cysteine should be located on the opposite side of the heme-binding site. This prohibits self-binding of synthetic heme to the attached protein, which can outcompete the formation of the micellar structure. His63 in cytochrome b_{562} is clearly located on the opposite side of the heme-binding site (Fig. 2). Thus, the H63C mutant is used as the building block in the reported literature [30].

To obtain the cage-like micellar structure formed by PNIPAAm-attached HTHP, the shape of HTHP should be considered. Due to the cylindrical shape of HTHP, the positioning options for introducing PNIPAAm into the protein including side and base positions. For the micellar structure, the hydrophobic areas should be dense, indicating that it is better to modify the base than the side (*see Note 1*). Thus, the V44C mutant has been found to provide an appropriate building block (Fig. 3) [34]. In

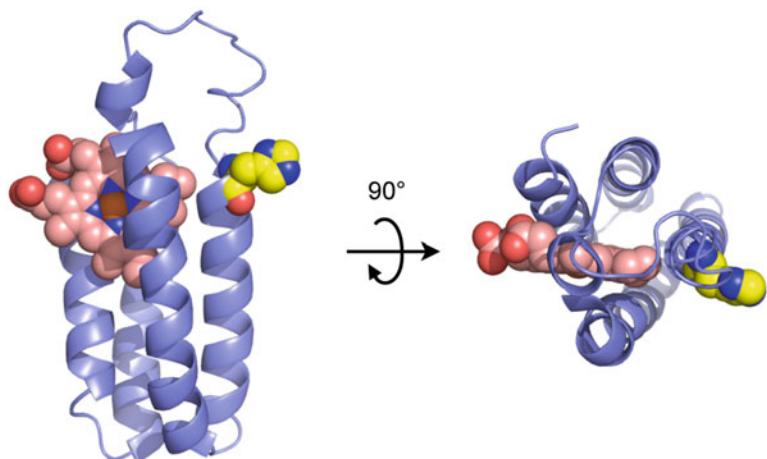


Fig. 2 Solution structure of cytochrome b_{562} (PDB ID: 1QPU). The heme molecule and the Hi63 residue are highlighted in pink and yellow, respectively

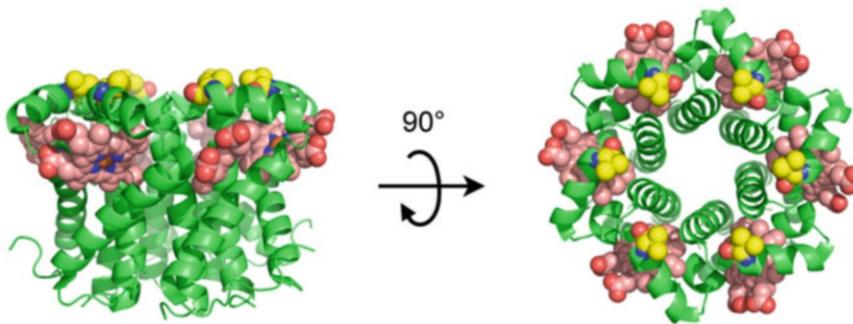


Fig. 3 Crystal structure of HTHP (PDB ID: 2OYY). The heme molecules and the Val44 residues are highlighted in pink and yellow, respectively

contrast, to create other large structures such as sheets, side modifications should be useful.

3.2 Expression and Purification of Cytochrome b_{562} Mutant

1. Transform the *Escherichia coli* strain TG1 with a plasmid encoding the H63C mutant and inoculate the bacteria onto an agar plate of LB medium containing ampicillin.
2. Inoculate one of the colonies into 5 mL of LB medium containing 5 μ L of 100 mg/mL ampicillin aqueous solution and shake it at 37 °C for 12 h.
3. Inoculate the cultured solution into 100 mL of LB medium containing 100 μ L of 100 mg/mL ampicillin aqueous solution and shake it at 37 °C until the OD₆₀₀ value reaches 1.0.
4. Inoculate each 25 mL of the cultured solution into 1 L of 2×YT medium (*see Note 2*) and shake at 37 °C overnight.
5. Collect the cells by centrifugation at 4 °C and suspend in 140 mL of 10 mM Tris–HCl buffer solution (pH 7.3).
6. Add 3 mL of CHCl₃ into the suspended solution and stir at 4 °C for 1 h.
7. Add 280 mL of 10 mM Tris–HCl buffer solution (pH 7.3) into the solution and stir at 4 °C for 1 h.
8. Remove the precipitates by centrifugation.
9. Adjust pH of the supernatant to 4.5 upon addition of 1 M HCl at 4 °C and stir the solution at 4 °C for 1 h.
10. Remove the precipitates by centrifugation.
11. Load the supernatant onto a CM Sepharose Fast Flow cation exchange column (swelled gel volume = 100 mL) equilibrated with solution A.
12. Wash the column with 1 L of solution A, then 300 mL of solution B.
13. Purify the protein by a linear gradient using solutions B and C.

14. Collect fractions with a R_z value (= absorbance at 418 nm/absorbance at 280 nm) greater than 5.
15. Concentrate the solution using an Amicon stirred ultrafiltration cell with a 5 kDa molecular weight cut-off membrane until the total volume is 1 mL.
16. Purify the protein solution using a Sephadex G-50 gel filtration column (swelled gel volume = 180 mL) equilibrated with a 50 mM Tris-HCl buffer solution (pH 8.0) containing 0.1 mM EDTA.
17. Collect the fractions with a R_z value greater than 6 and concentrate the solution.
18. The resulting solution can be stored at -80°C for at least 6 months.

3.3 Expression and Purification of HTHP Mutant

1. Transform the *Escherichia coli* strain BL21(DE3) with a plasmid encoding the V44C mutant and inoculate the bacteria onto an agar plate of LB medium containing ampicillin.
2. Inoculate one of the colonies into 5 mL of LB medium containing 5 μL of 100 mg/mL ampicillin aqueous solution and shake at 37°C for 4 h.
3. Inoculate the cultured solution into 100 mL of LB medium containing 100 μL of 100 mg/mL ampicillin aqueous solution and shake at 37°C until the OD_{600} value reaches 0.5.
4. Inoculate each 25 mL of the cultured solution into 1.5 L of LB medium (*see Note 2*) and shake at 37°C for 2 h. Then add the isopropyl- β -D-1-thiogalactopyranoside (1 mM as a final concentration) before further shaking at 37°C for 8 h.
5. Collect the cells by centrifugation at 4°C and suspend in 50 mL of 10 mM Tris-HCl buffer solution (pH 8.0).
6. Lyse the cells by freeze-thaw cycles with subsequent sonication at 4°C .
7. Remove the precipitates by centrifugation.
8. Add 3 mL of a 1 mM hemin stock solution dissolved in an aqueous solution of 0.1 M NaOH into the supernatant at 4°C and stir at 4°C for 10 min.
9. Incubate the solution at 80°C for 10 min and remove the precipitates by centrifugation.
10. Load the supernatant onto a DEAE Sepharose Fast Flow anion exchange column (swelled gel volume = 100 mL) equilibrated with solution D.
11. Wash the column with 500 mL of solution E.
12. Purify the protein with a linear gradient using solutions D and F.

13. Collect the fractions with a of R_z value (= absorbance at 404 nm/absorbance at 280 nm) greater than 2.2.
14. Concentrate the solution using an Amicon stirred ultrafiltration cell with a 10 kDa molecular weight cut-off membrane until the total volume is 1 mL.
15. Purify the protein solution using a HiPrep 16/60 Sephadryl S-200 HR column (120 mL) equilibrated with 100 mM potassium phosphate buffer (pH 7.0).
16. Collect the fractions with a of R_z value greater than 2.5 and concentrate the solution.
17. The resulting solution can be stored at -80°C for at least 6 months.

3.4 Modification of Cytochrome b_{562} Mutant by Synthetic Heme

1. Add a 40 μL of 1 M DTT as a reductant into a 400 μL of 1 mM H63C mutant solution and incubate the solution at 50°C for 10 min.
2. Purify the protein solution using a HiTrap Desalting column equilibrated with 100 mM NaHCO₃ buffer (pH 9.0).
3. Add 0.66 mL of a tetrahydrofuran solution of 4 mM heme with an iodoacetamide group connected via an azobenzene linker (*see Note 3*) into 5.8 mL of 67 μM protein solution.
4. Stir the solution in the dark under an N₂ atmosphere at room temperature for 8 h.
5. Add 1 M HCl to pH 1.8 to acidify the solution and extract free heme molecules with cooled 2-butanone.
6. Neutralize the solution by dialysis using 50 mM Tris-HCl buffer (pH 7.3) containing 0.1 mM EDTA.
7. Purify the protein by HPLC using a C18 column with a water/acetonitrile eluent containing 0.1% TFA (*see Note 4*).
8. Collect the fractions with visible absorption and lyophilize.
9. The resulting powder can be stored at -80°C for at least 6 months and is dissolved in 50 mM Tris-HCl buffer (pH 7.3) containing 0.1 mM EDTA prior to use.
10. Heat the protein solution at 80°C for 2 min and cool it in another water bath at 25°C to obtain the cage-like micellar assembly (*see Note 5*).

3.5 Modification of HTHP with Maleimide-Tethering PNIPAAm

1. Add a 100 μL of 1 M DTT aqueous solution into 900 μL of 500 μM V44C mutant solution and incubate the solution at 50°C for 1 h.
2. Purify the protein solution using a HiTrap Desalting column equilibrated with 100 mM potassium phosphate buffer (pH 7.0).

3. Dissolve 24.7 mg maleimide-terminated PNIPAAm (10 eq) into 500 μ L of 100 mM potassium phosphate buffer (pH 7.0).
4. Purify the polymer solution using a HiTrap Desalting column equilibrated with 100 mM potassium phosphate buffer (pH 7.0).
5. Add 1.0 mL of maleimide-terminated PNIPAAm solution into 2.0 mL of purified protein solution and incubate the mixture at 4 °C for 2 h.
6. Load the solution onto a DEAE Sepharose Fast Flow anion exchange column (swelled column volume = 2.0 mL) equilibrated with 100 mM potassium phosphate buffer (pH 7.0).
7. Elute the target protein using 100 mM potassium phosphate buffer (pH 7.0) containing 0.5 M NaCl.
8. Dialyze the solution with 100 mM potassium phosphate buffer (pH 7.0).
9. The resulting solution can be stored at –80 °C for at least 6 months.
10. Heat the protein solution at 60 °C to obtain the cage-like micellar assembly (*see Note 6*).

3.6 Characterization of Cage-Like Micellar Structures

The secondary structure of the protein region in cage-like micellar structures is evaluated by circular dichroism (CD) spectroscopy. The UV-vis spectrum in the visible region is also useful for confirming the coordination state of heme to determine if heme is properly bound within the protein matrix. This section describes the procedure for preparing samples for size evaluation.

3.6.1 Sample Preparation for Dynamic Light Scattering

Prepare a protein solution with appropriate concentrations (10 to 20 mg/mL, *see Note 7*). Filter the protein solution using a 0.22 μ m filter and add the protein solution to a cuvette prior to the measurement.

3.6.2 Sample Preparation for Transmission Electron Microscopy

Use an aldehyde-modified carbon grid to immobilize the protein on the grid. Prepare the protein solution with appropriate concentrations (0.05 to 0.5 mg/mL). Place 10 μ L of the protein solution on the grid for 10 min. Transfer the grid onto a 10 μ L staining solution for 5 s. Remove the residual staining solution carefully with a filter paper. Transfer the grid to the apparatus after drying under ambient conditions. A representative transmission electron microscopic image is shown in Fig. 4.

3.6.3 Sample Preparation for Atomic Force Microscopy

Place 3 μ L of 0.1% (3-aminopropyl)triethoxysilane aqueous solution on freshly cleaved mica for 3 min. Wash the substrate with 70 μ L of water. Place 3 μ L of 0.25% glutaraldehyde aqueous solution on the substrate for 3 min. Wash the substrate with 70 μ L of water. Place 3 μ L of the 0.5 mg/mL protein solution onto a modified mica substrate for 3 min. Place the substrate onto the apparatus after

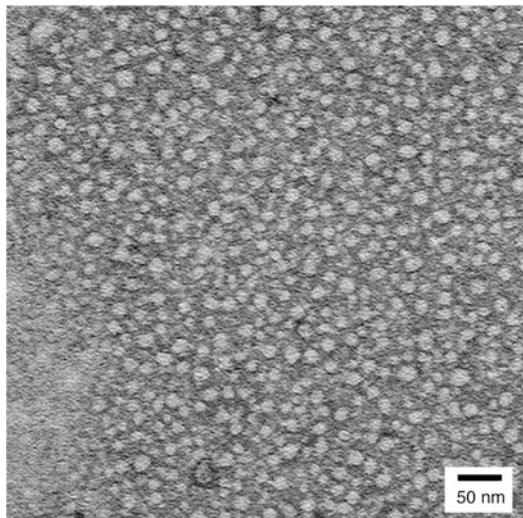


Fig. 4 Transmission electron microscopic image of a cage-like micellar assembly of the H63C mutant of cytochrome b_{562} modified by synthetic heme with an iodoacetamide group via an azobenzene linker

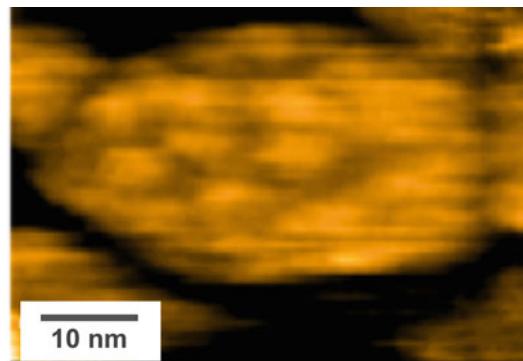


Fig. 5 Atomic force microscopic image of a cage-like micellar assembly of the V44C mutant of HTHP modified with poly-*N*-isopropylacrylamide

washing with 100 mM potassium phosphate buffer (pH 7.0). A representative atomic force microscopic image is shown in Fig. 5.

4 Notes

1. Other mutation positions on the sides of HTHP do not result in micellar assemblies. Instead, larger amorphous assemblies are found as precipitates.
2. For efficient aeration, a flask with baffles is used for culture.

3. A maleimide group is more reactive toward Cys residues than an iodoacetamide group, but synthesis of heme with the maleimide group via an azobenzene linker has not been successful.
4. No purification is required if the modification occurs quantitatively. For this system, the modification yield may be as high as 60% and purification is performed under denaturing conditions.
5. The obtained cage-like micellar assembly is a kinetically trapped meta-stable state which spontaneously converts to the supramolecular fibrous assembly after 96 h.
6. The obtained cage-like micellar assembly is stable at 60 °C. The assembly immediately dissociates upon cooling to 25 °C. If measurements of the cage-like micellar assembly at 25 °C are required, covalent cross-linkage with glutaraldehyde at 60 °C is available.
7. Dynamic light scattering measurements are much more sensitive to contaminants over micrometer diameters compared to samples with diameters of tens of nanometers.

Acknowledgments

We acknowledge supports from JSPS KAKENHI JP15H05804, JP15K21707, JP20H00403, and JP20H02755.

References

1. Luo Q, Hou C, Bai Y, Wang R, Liu J (2016) Protein assembly: versatile approaches to construct highly ordered nanostructures. *Chem Rev* 116:13571–13632
2. Kobayashi N, Arai R (2017) Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Curr Opin Biotech* 46:57–65
3. van Dun S, Ottmann C, Milroy LG, Brunsved L (2017) Supramolecular chemistry targeting proteins. *J Am Chem Soc* 139:13960–13968
4. Fegan A, White B, Carlson JCT, Wagner CR (2010) Chemically controlled protein assembly: techniques and applications. *Chem Rev* 110:3315–3336
5. Lai YT, Cascio D, Yeates TO (2012) Structure of a 16-nm cage designed by using protein oligomers. *Science* 336:1129
6. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang P, Ravichandran R, Yi S, Davis TN, Gonen T, King NP, Baker D (2016) Design of a hyper-stable 60-subunit protein icosahedron. *Nature* 535:136–139
7. Burazerovic S, Gradinaru J, Pierron J, Ward TR (2007) Hierarchical self-assembly of one-dimensional streptavidin bundles as a collagen mimetic for the biomineratization of calcite. *Angew Chem Int Ed* 46:5510–5514
8. Nguyen TK, Negishi H, Abe S, Ueno T (2019) Construction of supramolecular nanotubes from protein crystals. *Chem Sci* 10:1046–1051
9. Malay AD, Miyazaki N, Biela A, Chakraborti S, Majsterkiewicz K, Stupka I, Kaplan CS, Kowalczyk A, Piette BMAG, Hochberg GKA, Wu D, Wrobel TP, Fineberg A, Kushwah MS, Kelemen M, Vavpetič P, Pelicon P, Kukura P, Benesch JLP, Iwasaki K, Heddle JG (2019) An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* 569:438–442
10. Simon AJ, Zhou Y, Ramasubramani V, Glaser J, Pothukuchi A, Gollihar J, Gerberich JC,

- Leggere JC, Morrow BR, Jung C, Glotzer SC, Taylor DW, Ellington AD (2019) Supercharging enables organized assembly of synthetic biomolecules. *Nat Chem* 11:204–212
11. Oohora K, Hayashi T (2014) Hemoprotein-based supramolecular assembling systems. *Curr Opin Chem Biol* 19:154–161
 12. Hirota S (2019) Oligomerization of cytochrome c, myoglobin, and related heme proteins by 3D domain swapping. *J Inorg Biochem* 194:170–179
 13. Hirota S, Hattori Y, Nagao S, Taketa M, Komori H, Kamikubo H, Wang Z, Takahashi I, Negi S, Sugiura Y, Kataoka M, Higuchi Y (2010) Cytochrome c polymerization by successive domain swapping at the C-terminal helix. *Proc Natl Acad Sci* 107:12854–12859
 14. Zhang M, Nakanishi T, Yamanaka M, Nagao S, Yanagisawa S, Shomura Y, Shibata N, Ogura T, Higuchi Y, Hirota S (2017) Rational design of domain-swapping-based c-type cytochrome heterodimers by using chimeric proteins. *Chembiochem* 18:1712–1715
 15. Hayashi Y, Yamanaka M, Nagao S, Komori H, Higuchi Y, Hirota S (2016) Domain swapping oligomerization of thermostable c-type cytochrome in *E. coli* cells. *Sci Rep* 6:19334
 16. Miyamoto T, Kurabayashi M, Nagao S, Shomura Y, Higuchi Y, Hirota S (2015) Domain-swapped cytochrome *c*b₅₆₂ dimer and its nanocage encapsulating a Zn-SO₄ cluster in the internal cavity. *Chem Sci* 6:7336–7342
 17. Churchfield LA, Tezcan FA (2019) Design and construction of functional supramolecular metalloprotein assemblies. *Acc Chem Res* 52: 345–355
 18. Radford RJ, Tezcan FA (2009) A superprotein triangle driven by nickel(II) coordination: exploiting non-natural metal ligands in protein self-assembly. *J Am Chem Soc* 131:9136–9137
 19. Ni TW, Tezcan FA (2010) Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew Chem Int Ed* 49:7014–7018
 20. Song WJ, Tezcan FA (2014) A designed supramolecular protein assembly with *in vivo* enzymatic activity. *Science* 346:1525–1528
 21. Brodin JD, Smith SJ, Carr JR, Tezcan FA (2015) Designed, helical protein nanotubes with variable diameters from a single building block. *J Am Chem Soc* 137:10468–10471
 22. Golub E, Subramanian RH, Esselborn J, Alberstein RG, Bailey JB, Chiong JA, Yan X, Booth T, Baker TS, Tezcan FA (2020) Constructing protein polyhedra via orthogonal chemical interactions. *Nature* 578:172–176
 23. Oohora K, Onoda A, Hayashi T (2012) Supramolecular assembling systems formed by heme–heme pocket interactions in hemoproteins. *Chem Commun* 48:11714–11726
 24. Kitagishi H, Oohora K, Yamaguchi H, Sato H, Matsuo T, Harada A, Hayashi T (2007) Supramolecular hemoprotein linear assembly by successive interprotein heme–heme pocket interactions. *J Am Chem Soc* 129:10326–10327
 25. Oohora K, Fujimaki N, Kajihara R, Watanabe H, Uchihashi T, Hayashi T (2018) Supramolecular hemoprotein assembly with a periodic structure showing heme–heme exciton coupling. *J Am Chem Soc* 140:10145–10148
 26. Oohora K, Kajihara R, Jiromaru M, Kitagishi H, Hayashi T (2019) Arginine residues provide a multivalent effect for cellular uptake of a hemoprotein assembly. *Chem Lett* 48:295–298
 27. Soon JW, Oohora K, Hirayama S, Hayashi T (2021) A supramolecular assembly of hemoproteins formed in a star-shaped structure via heme–heme pocket interactions. *Int J Mol Sci* 22:1012
 28. Kitagishi H, Kakikura Y, Yamaguchi H, Oohora K, Harada A, Hayashi T (2009) Self-assembly of one- and two-dimensional hemoprotein systems by polymerization through heme–heme pocket interactions. *Angew Chem Int Ed* 48:1271–1274
 29. Oohora K, Onoda A, Kitagishi H, Yamaguchi H, Harada A, Hayashi T (2011) A chemically-controlled supramolecular protein-polymer formed by a myoglobin-based self-assembly system. *Chem Sci* 2:1033–1038
 30. Oohora K, Onuma Y, Tanaka Y, Onoda A, Hayashi T (2017) A supramolecular assembly based on an engineered hemoprotein exhibiting a thermal stimulus-driven conversion to a new distinct supramolecular structure. *Chem Commun* 53:6879–6882
 31. Jeoung JH, Pippig DA, Martins BM, Wagener N, Dobbek H (2007) HTHP: a novel class of hexameric, tyrosine-coordinated heme proteins. *J Mol Biol* 368:1122–1131
 32. Oohora K, Hirayama S, Mashima T, Hayashi T (2020) Supramolecular dimerization of a hexameric hemoprotein via multiple pyrene-

- pyrene interactions. *J Porphyrins Phthalocyanines* 24:259–267
33. Oohora K, Hirayama S, Uchihashi T, Hayashi T (2020) Construction of a hexameric hemoprotein sheet and direct observation of dynamic process of its formation. *Chem Lett* 49:186–190
34. Hirayama S, Oohora K, Uchihashi T, Hayashi T (2020) Thermoresponsive micellar assembly constructed from a hexameric hemoprotein modified with poly(*N*-isopropylacrylamide)
- toward an artificial light-harvesting system. *J Am Chem Soc* 142:1822–1831
35. Wan X, Liu S (2010) Fabrication of a thermo-responsive biohybrid double hydrophilic block copolymer by a cofactor reconstitution approach. *Macromol Rapid Commun* 31: 2070–2076
36. Reynhout IC, Cornelissen JJLM, Nolte RJM (2007) Self-assembled architectures from biohybrid triblock copolymers. *J Am Chem Soc* 129:2327–2332

Part II

Functionalization of Protein Cages



Chapter 6

Fabrication of Protein Macromolecular Frameworks (PMFs) and Their Application in Catalytic Materials

Masaki Uchida, Ekaterina Selivanovitch, Kimberly McCoy,
and Trevor Douglas

Abstract

The construction of three-dimensional (3D) array materials from nanoscale building blocks has drawn significant interest because of their potential to exhibit collective properties and functions arising from the interactions between individual building blocks. Protein cages such as virus-like particles (VLPs) have distinct advantages as building blocks for higher-order assemblies because they are extremely homogeneous in size and can be engineered with new functionalities by chemical and/or genetic modification. In this chapter, we describe a protocol for constructing a new class of protein-based superlattices, called protein macromolecular frameworks (PMFs). We also describe an exemplary method to evaluate the catalytic activity of enzyme-enclosed PMFs, which exhibit enhanced catalytic activity due to the preferential partitioning of charged substrates into the PMF.

Key words Virus-like particle (VLP), Protein array, Protein macromolecular framework (PMF), Superlattice, Self-assembly, Enzyme encapsulation, Collective behavior

1 Introduction

Protein cages, including virus-like particles (VLPs), provide unique platforms for encapsulation of functional cargos and have been utilized in various applications including catalysis and vaccine development [1, 2]. For example, we have demonstrated encapsulation of various enzymes inside the VLPs derived from P22 bacteriophage [3–6]. The enzyme cargos maintain their catalytic activities, while the capsid shell serves as a protective layer for the enzymes [4, 7], leading to the development of more robust catalytic materials. A key feature of protein cages, their easy functionalization via chemical or genetic modification, also makes them ideal building blocks to construct higher-order assemblies with designed functionalities [2, 8]. A range of interactions, including electrostatic [9–15], hydrophobic [16], complementary DNA [17–19],

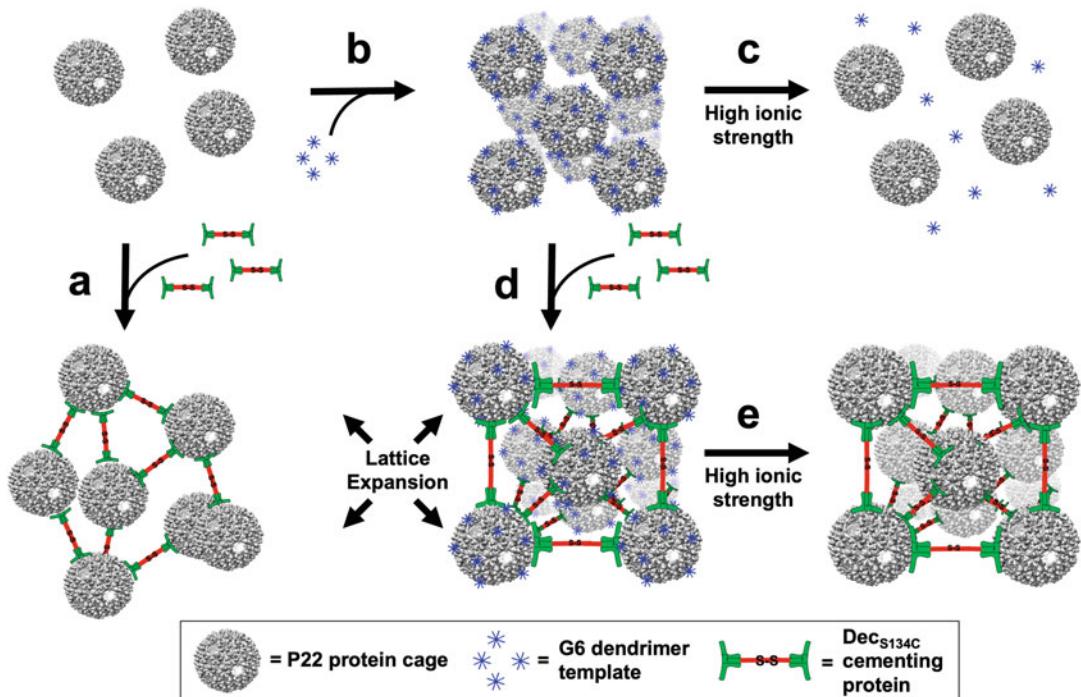


Fig. 1 Schematic cartoon depicting protein macromolecular framework (PMF) formation process. Negatively charged P22-E2 VLPs are mixed with either (a) DecS134C linker protein or (b) positively charged PAMAM generation 6 dendrimers (G6), resulting in formation of amorphous array or an ordered array, respectively. When ionic strength is increased, (c) the array disassembles. (d) DecS134C linker proteins are added to G6 templated array to lock the structure in place while concomitantly increasing the lattice parameter approximately 4 nm. (e) Ionic strength is then increased, removing the G6 from the array but preserving the structure, yielding a PMF. (Adapted from Ref. 21 with permission)

metal-ligand coordination [20], and specific protein-protein [21, 22] interactions have been exploited to mediate higher-order assembly of protein cages into three-dimensional (3D) arrays.

We have recently developed a new class of 3D protein arrays, protein macromolecule framework (PMF), by using P22 VLP and a capsid decoration protein (Dec) as building blocks (Fig. 1). The P22 VLP has three well-characterized morphologies, procapsid (PC) [23], expanded (EX) [24], and wiffleball (WB) [25, 26]. Dec is a homotrimeric protein and binds to the quasi and true threefold symmetry sites (80 binding sites per VLP) on the EX and WB forms of P22 capsid [27, 28]. Point mutation of Dec, in which the C-terminal serine is replaced with a cysteine (DecS134C), results in the formation of a head-to-head dimer of trimers via disulfide bond formation upon oxidation, enabling DecS134C to act as a ditopic linker bridging between P22 VLPs (Fig. 1) [29]. We have demonstrated that DecS134C bridges neighboring VLPs in an ordered array of P22 VLPs and stabilizes the lattice, resulting in the formation of the PMF (Figs. 1 and 2).

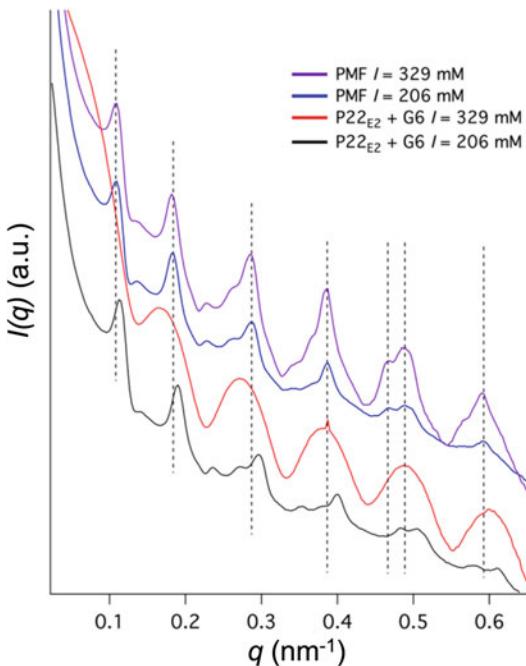


Fig. 2 Small-angle X-ray scattering of the templated assembly of PMFs. $P22-E2$ assembled with G6 results in an ordered array with a face-centered cubic (FCC) structure (black). When this ordered array is transferred into a high ionic strength buffer ($I = 329 \text{ mM}$), structure factor contributions are significantly reduced, indicating disassembly of the array into individual VLPs (red). Addition of DecS134C linker protein to a G6 templated array (blue) shows preservation of long-range order and a shift in peak positions (vertical dashed lines). This peak shift corresponds to an increase in the lattice parameter from 92.6 nm to 96.8 nm . When PMFs is transferred into a high ionic strength buffer (purple), long-range order and peak position are preserved. (Adapted from Ref. 21 with permission)

[21, 22], which is conceptually analogous to a metal-organic framework (MOF) [30] but is constructed entirely of protein building blocks.

We have shown that PMFs can also be fabricated from enzyme-encapsulated $P22$ VLPs [21, 22]. The robustness and versatility of the PMFs allow us to impart desired functionality by encapsulating any enzyme inside individual $P22$ VLPs, while the interstitial space and porosity of the PMF provides room for the diffusion of substrates with size- and charge-based selectivity. Because $P22$ VLP carries a negative exterior surface charge, the interstitial space of the PMF is highly negatively charged. Thus, positively charged substrates can be preferentially taken up into the PMF. This selectivity leads to a unique collective behavior where the catalytic activity of the PMFs is significantly higher than the enzyme-enclosed free VLPs for charged substrates [22]. Here we provide a protocol for

the fabrication of PMFs from P22 VLPs and the Dec linkers. We also provide an exemplary protocol for a catalytic activity assay by using a PMF constructed from alcohol dehydrogenase D (AdhD) encapsulated P22 VLPs.

2 Materials

All solutions were prepared using ultrapure water (resistivity of 18.2 MΩ·cm at 25 °C) unless otherwise noted. Filter all buffer solutions through 0.2 μm filters.

2.1 Preparation of VLP Building Blocks

1. P22 PMF is typically constructed with P22-E2, a mutant of P22 VLP with two repeats of the hepta-peptide sequence (VAALEKE) fused to the C-terminus of the coat protein and displayed on the exterior surface of the VLP [13]. The procapsid form of P22-E2 VLP and an enzyme encapsulated P22-E2 VLP were heterologously expressed in, and purified from, *E. coli* (BL21 (DE3)). The cloning, transformation, expression, and purification procedure was described in previous papers [3, 21, 22, 31].
2. A mutant of the decoration protein (DecS134C) was heterologously expressed in, and purified from *E. coli* (BL21 (DE3)). The cloning, transformation, expression, and purification procedure was described in previous papers [21, 22, 29].
3. Phosphate buffers: Phosphate buffers with two different ionic strengths (I) are prepared. (a) 50 mM sodium phosphate, 100 mM sodium chloride ($I = 206$ mM), pH 7.0. Weigh 2.92 g NaH₂PO₄•H₂O, 7.73 g Na₂HPO₄•7H₂O and 5.84 g NaCl. Dissolve them in 900 mL water. (b) 80 mM sodium phosphate, 160 mM sodium chloride ($I = 329$ mM), pH 7.0. Weigh 4.67 g NaH₂PO₄•H₂O, 12.37 g Na₂HPO₄•H₂O and 9.35 g NaCl. Dissolve them in 900 mL water. Adjust pH of the both solutions to 7.0 with 1 M NaOH. Make up to 1.0 L with water. Filter through 0.22 μm membrane filter. Store at room temperature.
4. 0.2% (w/v) sodium dodecyl sulfate (SDS) in $I = 206$ mM phosphate buffer.
5. Water bath with capability to heat up to 75 °C.
6. Dithiothreitol (DTT).
7. Copper (II) sulfate (CuSO₄).
8. 12–14 kDa molecular weight cutoff dialysis tubing and closures.

2.2 Construction of PMFs

1. Polyamidoamine (PAMAM) generation 6 (G6) dendrimer.
2. 12–14 kDa molecular weight cutoff dialysis tubing and closures.
3. Phosphate buffers: Phosphate buffers with 50 mM sodium phosphate, 100 mM sodium chloride, pH 7.0 ($I = 206$ mM) and 80 mM sodium phosphate, 160 mM sodium chloride, pH 7.0 ($I = 329$ mM) (see subheading 2.1).

2.3 Activity Assay of AdhD-Enclosed PMF with Charged Substrates

1. PMFs constituted with the EX form of AdhD-encapsulated P22-E2.
2. Phosphate buffers: Phosphate buffers with three different ionic strengths (I), i.e., low salt (LS, $I = 41$ mM), intermediate salt (IS, $I = 206$ mM), and high salt (HS, $I = 508$ mM), are used for the activity assay. LS has 10 mM sodium phosphate, 20 mM sodium chloride, pH 7.0. IS has 50 mM sodium phosphate, 100 mM sodium chloride, pH 7.0. HS has 50 mM sodium phosphate, 400 mM sodium chloride, pH 7.0.
3. Nicotinamide adenine dinucleotide (NADH) conjugated with PAMAM G0.5, G1.5, or their variants. They are referred to as NADH_{-x}^(y) in this chapter, where x and y represent the generation and charge of a dendrimer conjugated to NADH, respectively. The synthesis of the NADH variants has been described elsewhere [22, 32].
4. Acetoin.
5. UV-Vis spectrometer or Plate reader.

3 Methods

3.1 Transformation of the P22 VLP

Morphology from PC to EX Form

1. Adjust the concentration of P22-E2 PC to 1 mg/mL in $I = 206$ mM phosphate buffer.
2. Add a freshly prepared 0.2% (w/v) SDS solution in $I = 206$ mM phosphate buffer to an equal volume of 1 mg/mL P22-E2 PC solution.
3. Incubate the mixed solution for 5 min at room temperature.
4. Centrifuge the sample at 17,000 g for 5 min at room temperature to remove aggregated protein.
5. Pellet P22-E2 by ultracentrifugation at approximately 209,000 g (45,000 rpm) (F50L-8×39 rotor, Piramoon Technologies) for 50 min at 4 °C. Discard the supernatant to remove SDS. Resuspend the P22-E2 pellet into $I = 329$ mM phosphate buffer. Repeat this ultracentrifugation and resuspension process twice to ensure the removal of SDS [22, 33].
6. Adjust the concentration of the EX form of P22-E2 to 1 mg/mL (see Note 1).

3.2 Transformation of the P22 VLP

Morphology from PC to WB Form

1. Adjust the concentration of P22-E2 PC to 1 mg/mL in $I = 206$ mM phosphate buffer.
2. Heat the P22 sample in a 75 °C water bath for 25 min, then cool on ice for at least 15 min [21, 26].
3. Centrifuge the sample at 17,000 g for 5 min at room temperature to remove aggregated protein.
4. Pellet P22-E2 by ultracentrifugation at approximately 209,000 g (45,000 rpm) (F50L-8×39 rotor) for 50 min at 4 °C to remove scaffolding protein and coat protein released from the VLPs [25], followed by resuspension into $I = 329$ mM phosphate buffer.
5. Adjust the concentration of the WB sample to 1 mg/mL with the same buffer (*see Note 1*).

3.3 Preparation of Ditopic Protein Linker from DecS134C

1. Adjust the concentration of DecS134C to 1 mg/mL in $I = 206$ mM phosphate buffer.
2. Add DTT to the DecS134C solution to obtain a final concentration of 5 mM DTT.
3. Incubate the DecS134C solution for 3 h at room temperature (*see Note 2*).
4. Transfer the DecS134C solution into a 12–14 kDa molecular weight cutoff dialysis tubing and dialyzed against $I = 206$ mM phosphate buffer at 4 °C overnight. The buffer solution should be replaced once with fresh buffer to ensure the removal of DTT from the DecS134C solution.
5. Add CuSO₄ to the reduced DecS134C solution to obtain a final concentration of 20 μM CuSO₄, which will oxidize DecS134C.
6. Incubate the DecS134C solution at 4 °C overnight, then heat at 60 °C for 20 min to form the head-to-head Dec dimer via disulfide bonds (*see Note 3*).
7. Centrifuge the solution at 17,000 g for 5 min at room temperature to remove aggregated protein.
8. Transfer the supernatant into a 12–14 kDa molecular weight cutoff dialysis tubing and dialyze against $I = 206$ mM phosphate buffer at 4 °C overnight to remove CuSO₄.

3.4 Construction of PMFs from the EX or WB Form of P22 VLP

1. Prepare a fresh solution of PAMAM G6 dendrimer in a 1:4 dilution of $I = 329$ mM phosphate buffer.
2. Slowly add the dendrimer solution to the solution containing P22-E2 VLPs (either EX or WB form) at room temperature in a 1000-fold excess of dendrimers to VLP (*see Note 4*).
3. Transfer the mixed solution into a 12–14 kDa molecular weight cutoff dialysis tubing and dialyze against $I = 206$ mM

phosphate buffer at room temperature for at least 3 h (*see Note 5*).

4. Adjust the concentration of DecS134C to 2 mg/mL with $I = 206$ mM phosphate buffer (*see Note 6*).
5. Add the DecS134C solution to the P22-G6 solution at room temperature in a 160: 1 molar ratio of trimeric DecS134C: P22 VLP (*see Note 7*).
6. Incubate the sample solution at room temperature for at least 30 min to ensure binding of DecS134C linkers to P22 VLPs, which will lead to the formation of the PMFs constituted of DecS134C linked P22 VLPs.
7. Centrifuge the PMF sample solution at 5000 g for 3 min at room temperature and then gently remove the supernatant without disturbing the a small white pellet containing the PMFs.
8. To remove the PAMAM G6 dendrimers from the PMF, resuspend the pellet in $I = 329$ mM phosphate buffer, followed by gentle centrifugation at 5000 g for 3 min at room temperature. Carefully remove the supernatant containing released PAMAM G6 dendrimers (*see Note 8*). Repeat this process twice to ensure that all dendrimers are removed from PMFs. Resuspend PMFs in $I = 206$ mM phosphate buffer and stored at 4 °C until further use.

3.5 Activity Assay of AdhD-Enclosed PMF with Charged Substrates

1. Prepare acetoin in the three different buffers (LS, IS, and HS) with concentrations ranging from 0 to 120 mM (*see Note 9*).
2. Prepare NADH variants, $\text{NADH}_{0.5}^{(-)}$, $\text{NADH}_{0.5}^{(+)}$, $\text{NADH}_{1.5}^{(-)}$, and $\text{NADH}_{1.5}^{(+)}$, in water at a high concentration, typically around 16 mM.
3. Adjust the AdhD-enclosed PMF concentration to 1 mg/mL of P22 VLPs.
4. Centrifuge the PMF at 5000 g for 3 min at room temperature to form a pellet, and then remove the supernatant without disturbing the pellet.
5. Resuspend the pellet in the same volume of LS, IS, or HS buffer containing 0–120 mM acetoin and transfer to reaction vials.
6. Add the NADH variants to the reaction vials in a 1:100 dilution just before the activity assay. The final concentration of NADH is typically around 160 μM.
7. Constantly agitate the sample solutions containing PMFs, acetoin, and NADH variants to prevent the PMFs from settling during the enzyme activity assay at room temperature.

8. Upon reaching each individual time-point, remove the respective aliquots from the reaction vials and centrifuge ($17,000 \times g$, 1 min) (see Note 10). Recover the supernatant.
9. Measure the absorbance at 340 nm (oxidation of NADH to NAD^+) of the supernatant using a UV-Vis spectrophotometer or a plate reader (see Note 11). Repeat this step a total minimum of three times for each condition.
10. Use the Michaelis-Menten kinetics model to extract the Michaelis-Menten constant (K_M) and the turnover rate (k_{cat}) (see Note 12).

4 Notes

1. Morphological transformation of PC to EX and WB forms can be confirmed with native agarose gel electrophoresis as described in previous papers [21, 22, 25, 26, 34].
2. The C-terminal cysteine of DecS134C tends to form disulfide bonds with other thiol containing molecules such as glutathione and cysteine during the expression process. DecS134C needs to be reduced to increase thiol availability for the desired formation of the head-to-head Dec dimer.
3. It is common that some aggregates form during this step.
4. The mixed solution should be clear, indicating no higher-order assembly in the $I = 329$ mM buffer.
5. The mixed solution should become turbid in the $I = 206$ mM buffer, indicating formation of a higher-order assembly (Figs. 1 and 2).
6. DecS134C can be concentrated by using a centrifugal filter device.
7. There are 80 total Dec binding sites per P22 VLP [28]. Addition of 160-fold Dec per P22 VLP results in a twofold excess of Dec per binding site.
8. The PMF of P22 VLP should stay as an assembled array material in the $I = 329$ mM buffer as the DecS134C linker cements the assembly (Figs. 1 and 2). The assembly without addition of DecS134C, on the other hand, disassembles in the same buffer [21].
9. Concentration of acetoin for the activity assay is typically 0, 1, 3, 10, 25, 50, 80, and 120 mM.
10. Typically, the solutions are aliquoted 3, 6, 9, 15, 21, 30, 60, and 120 min after addition of the NADH variants.
11. Wavelength at 600 nm is also monitored to ensure that the PMFs are removed from the solution. The optical density at

600 nm of the supernatant should be nearly zero. The presence of the PMFs results in a higher value of the optical density due to light scattering.

12. The concentration of AdhD was determined by measuring the absorbance at 280 nm and the molecular weight values of AdhD-encapsulated P22 VLP determined by SEC-MALS, as previously described [3, 35].

Acknowledgments

This work was supported by a grant from the Human Frontier Science Program (HFSP) 4124801. M.U. was supported in part by the National Science Foundation grant CMMI-1922883. E.S. was partially supported by the Graduate Training Program in Quantitative and Chemical Biology under Award T32 GM109825 and Indiana University. T.D. was additionally supported by the National Science Foundation through grant 1720625.

References

1. Douglas T, Young M (2006) Viruses: making friends with old foes. *Science* 312:873–875
2. Aumiller WM, Uchida M, Douglas T (2018) Protein cage assembly across multiple length scales. *Chem Soc Rev* 47:3433–3469
3. Patterson DP, Prevelige PE, Douglas T (2012) Nanoreactors by programmed enzyme encapsulation inside the capsid of the bacteriophage P22. *ACS Nano* 6:5000–5009
4. Jordan PC, Patterson DP, Saboda KN, Edwards EJ, Miettinen HM, Basu G, Thielges MC, Douglas T (2016) Self-assembling biomolecular catalysts for hydrogen production. *Nat Chem* 8:179–185
5. Patterson DP, Schwarz B, Waters RS, Gedeon T, Douglas T (2014) Encapsulation of an enzyme cascade within the bacteriophage p22 virus-like particle. *ACS Chem Biol* 9:359–365
6. Wang Y, Uchida M, Waghwanı HK, Douglas T (2020) Synthetic virus-like particles for glutathione biosynthesis. *ACS Synth Biol* 9:3298–3310
7. O’Neil A, Prevelige PE, Douglas T (2013) Stabilizing viral nano-reactors for nerve-agent degradation. *Biomater Sci* 1:881–886
8. Selivanovitch E, Douglas T (2019) Virus capsid assembly across different length scales inspire the development of virus-based biomaterials. *Curr Opin Virol* 36:38–46
9. Kostiainen MA, Kasyutich O, Cornelissen JJLM, Nolte RJM (2010) Self-assembly and optically triggered disassembly of hierarchical dendron-virus complexes. *Nat Chem* 2:394–399
10. Kostiainen MA, Hiekkataipale P, Laiho A, Lemieux V, Seitsonen J, Ruokolainen J, Ceci P (2012) Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nat Nanotechnol* 8:52–56
11. Liljeström V, Ora A, Hassinen J, Rekola HT, Nonappa HM, Hynnen V, Joensuu JJ, Ras RHA, Törmä P, Ikkala O, Kostiainen MA (2017) Cooperative colloidal self-assembly of metal-protein superlattice wires. *Nat Commun* 8:671
12. Künzle M, Eckert T, Beck T (2016) Binary protein crystals for the assembly of inorganic nanoparticle superlattices. *J Am Chem Soc* 138:12731–12734
13. Uchida M, McCoy K, Fukuto M, Yang L, Yoshimura H, Miettinen HM, Lafrance B, Patterson DP, Schwarz B, Karty JA, Prevelige PE Jr, Lee B, Douglas T (2018) Modular self-assembly of protein cage lattices for multistep catalysis. *ACS Nano* 12:942–953
14. Chakraborti S, Korpi A, Kumar M, Stepien P, Kostiainen MA, Heddle JG (2019) Three-dimensional protein cage array capable of active enzyme capture and artificial chaperone activity. *Nano Lett* 19:3918–3924

15. Ramberg KO, Engilberge S, Skorek T, Crowley PB (2021) Facile fabrication of protein-macrocycle frameworks. *J Am Chem Soc* 143: 1896–1907
16. Zheng B, Zhou K, Zhang T, Lv C, Zhao G (2019) Designed two- and three-dimensional protein nanocage networks driven by hydrophobic interactions contributed by amyloidogenic motifs. *Nano Lett* 19:4023–4028
17. Strable E, Johnson JE, Finn MG (2004) Natural nanochemical building blocks: icosahedral virus particles organized by attached oligonucleotides. *Nano Lett* 4:1385–1389
18. McMillan JR, Brodin JD, Millan JA, Lee B, Olvera de la Cruz M, Mirkin CA (2017) Modulating nanoparticle superlattice structure using proteins with tunable bond distributions. *J Am Chem Soc* 139:1754–1757
19. Tian Y, Lhermitte JR, Bai L, Vo T, Xin HL, Li H, Li R, Fukuto M, Yager KG, Kahn JS, Xiong Y, Minevich B, Kumar SK, Gang O (2020) Ordered three-dimensional nanomaterials using DNA-prescribed and valence-controlled material voxels. *Nat Mater* 19:789–796
20. Bailey JB, Zhang L, Chiong JA, Ahn S, Tezcan FA (2017) Synthetic modularity of protein-metal–organic frameworks. *J Am Chem Soc* 139:8160–8166
21. McCoy K, Uchida M, Lee B, Douglas T (2018) Templated assembly of a functional ordered protein macromolecular framework from p22 virus-like particles. *ACS Nano* 12:3541–3550
22. Selivanovitch E, Uchida M, Lee B, Douglas T (2021) Substrate partitioning into protein macromolecular frameworks for enhanced catalytic turnover. *ACS Nano* 15(10): 15687–15699
23. Thuman-Commike PA, Greene B, Jakana J, Prasad BV, King J, Prevelige PE Jr, Chiu W (1996) Three-dimensional structure of scaffolding-containing phage p22 procapsids by electron cryo-microscopy. *J Mol Biol* 260: 85–98
24. Tuma R, Prevelige PE, Thomas GJ (1998) Mechanism of capsid maturation in a double-stranded DNA virus. *Proc Natl Acad Sci U S A* 95:9885–9890
25. Teschke CM, McGough A, Thuman-Commike PA (2003) Penton release from P22 heat-expanded capsids suggests importance of stabilizing penton-hexon interactions during capsid maturation. *Biophys J* 84:2585–2592
26. Kang S, Uchida M, O’Neil A, Li R, Prevelige PE, Douglas T (2010) Implementation of P22 viral capsids as nanoplatforms. *Biomacromolecules* 11:2804–2809
27. Tang L, Gilcrease EB, Casjens SR, Johnson JE (2006) Highly discriminatory binding of capsid-cementing proteins in bacteriophage L. *Structure* 14:837–845
28. Parent KN, Deedas CT, Egelman EH, Casjens SR, Baker TS, Teschke CM (2012) Stepwise molecular display utilizing icosahedral and helical complexes of phage coat and decoration proteins in the development of robust nanoscale display vehicles. *Biomaterials* 33:5628–5637
29. Uchida M, Lafrance B, Broomell CC, Prevelige PE Jr, Douglas T (2015) Higher order assembly of Virus-like Particles (VLPs) mediated by multi-valent protein linkers. *Small* 11:1562–1570
30. Yaghi OM, O’Keeffe M, Ockwig NW, Chae HK, Eddaoudi M, Kim J (2003) Reticular synthesis and the design of new materials. *Nature* 423:705–714
31. McCoy K, Douglas T (2018) In vivo packaging of protein cargo inside of virus-like particle P22. *Methods Mol Biol* 1776:295–302
32. Selivanovitch E, LaFrance B, Douglas T (2021) Molecular exclusion limits for diffusion across a porous capsid. *Nat Commun* 12:2903
33. Selivanovitch E, Koliyatt R, Douglas T (2019) Chemically induced morphogenesis of P22 virus-like particles by the surfactant sodium dodecyl sulfate. *Biomacromolecules* 20:389–400
34. Lucon J, Qazi S, Uchida M, Bedwell GJ, Lafrance B, Prevelige PE, Douglas T (2012) Use of the interior cavity of the P22 capsid for site-specific initiation of atom-transfer radical polymerization with high-density cargo loading. *Nat Chem* 4:781–788
35. Sharma J, Douglas T (2020) Tuning the catalytic properties of P22 nanoreactors through compositional control. *Nanoscale* 12:336–346



Chapter 7

Iron Accumulation in Ferritin

Akankshika Parida and Rabindra K. Behera

Abstract

Understanding the iron accumulation process in ferritin protein nanocages has remained a centerpiece in the field of iron biochemistry/biomineratization, which ultimately has implications in health and diseases. Although mechanistic differences of iron acquisition and mineralization exist in the superfamily of ferritins, we describe the techniques that can be used to investigate the accumulation of iron in all the ferritin proteins by *in vitro* iron mineralization process. In this chapter, we report that the non-denaturing polyacrylamide gel electrophoresis coupled with Prussian blue staining (*in-gel* assay) can be useful to investigate the iron-loading efficiency in ferritin protein nanocage, by estimating the relative amount of iron incorporated inside it. Similarly, the absolute size of the iron mineral core and the amount of total iron accumulated inside its nanocavity can be determined by using transmission electron microscopy and spectrophotometry, respectively.

Key words Iron, Ferritin nanocage, Iron mineralization, Native PAGE, Iron staining, TEM, Iron quantification

1 Introduction

Iron, one of the most essential elements, is required ($\sim 10^{-5} - 10^{-3}$ M) for various important biological processes such as oxygen binding/transport, N₂ fixation, and synthesis of RBCs, DNA, and ATP (during both oxidative/phosphorylation) [1, 2]. However, in spite of its large abundance on the Earth's crust (~1 M), its bioavailability is limited due to the poor solubility of Fe³⁺ ($\sim 10^{-18}$ M under physiological conditions and $\sim 10^{-9}$ M in seawater) [1–3]. Although the solubility of Fe²⁺ ion is sufficiently high (~0.1 M) in aqueous solution, it gets oxidized readily under aerobic environment. Moreover, the excess free Fe²⁺ ion can be highly toxic (generates ROS via Fenton reaction) and can promote infection; thus, iron concentration inside the living organism needs to be tightly regulated and well managed [3–6].

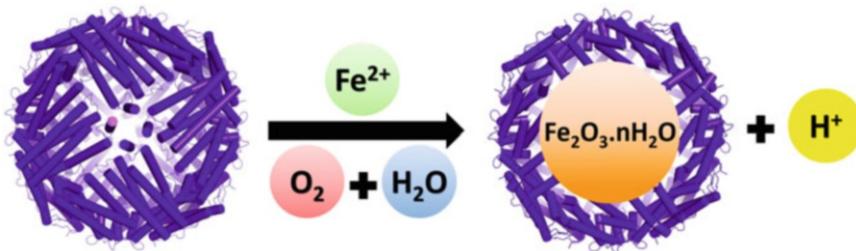


Fig. 1 Iron accumulation in ferritin cage. Illustration of a 24-mer ferritin nanocage, which rapidly scavenges free Fe^{2+} ions and accumulates in the form of a ferric oxyhydroxide mineral after ferroxidase/mineralization reaction

Ferritin, the cellular iron repository in almost all organisms (except yeast), sequesters toxic Fe^{2+} ion (prevents ROS formation and thus acts as an antioxidant) and decreases the iron solubility gap by synthesizing protein-coated hydrated ferric oxide minerals ($\text{Fe}_2\text{O}_3\cdot\text{nH}_2\text{O}$) [1, 4, 7–12]. These proteins have spherical nanocage architecture (external diameter ~12 nm; internal diameter ~8 nm), self-assembled from 24 subunits (four α -helical bundles) in maxi-ferritins [4, 6, 12, 13]. Self-assembly of ferritin subunits generates various pores for different functions (rapid sequestration of Fe^{2+} , entry of phosphate/ O_2 and release of $\text{Fe}^{2+}/\text{H}^+$) by controlling pore electrostatics and forms a central hollow nanocavity (~30% of total cage volume) that can accommodate up to 4500 iron atoms [14–18]. Once acquired, the Fe^{2+} ions are ferried to the enzymatic sites (ferroxidase centers) for rapid oxidation, and subsequently undergo a cascade of complex, hydrolysis reactions to form ferric oxy-hydroxide mineral inside the ferritin nanocavity (Fig. 1) [6, 8, 12, 15, 19]. The protein encapsulation keeps these mineral in a safe, soluble, and bioavailable form for cellular metabolic requirements [1, 4, 19–21].

The mammalian ferritin nanocages are heteropolymeric, i.e., comprising of two types of ferritin subunits, which are both functionally and genetically distinct and their ratio is tissue specific [1, 4, 8, 22]. H-ferritin (heavy chain) contains the di-iron catalytic site and involves in rapid iron uptake/oxidation, whereas the L-ferritin (light chain) lacks catalytic site but stabilizes the iron mineral electrostatically to facilitate its formation [23–25]. These H and L subunits not only act cooperatively/synergistically to facilitate the kinetics of ferroxidase activity but also influence the iron accumulation ability of ferritins to meet cell/tissue specific requirements (heart vs. liver/spleen) [7, 8, 19, 23].

Ferritins isolated from different source may have different iron storage/accumulation capacity and precipitates (forms aggregates) when in vitro iron loading reaches beyond its limit [7, 16, 20, 26]. Similarly, iron-induced precipitation also arises during neuroferritinopathy due to impaired iron storage capacity of ferritins

[27]. Normally, *in vivo*, these soluble ferritin nanocages store ~500–2500 Fe/cage, which varies during iron deficiency/overload and neurodegenerative disorders [27, 28]. In humans, apart from haemochromatosis, a genetic disorder, iron overload also occurs in sickle cell disease and thalassemia cases due to the blood transfusion at regular intervals. Ferritin also accumulates excess amount of iron under these abnormal circumstances [19, 29]. Moreover, the effect of phosphate (another component of iron mineral in native ferritins) on the synthesis, structure, and reactivity of ferritin mineral core were also investigated, which not only influenced the mineralization processes but also decreased the size/crystallinity of the iron mineral core [4, 12, 30, 31].

Understanding the uptake and accumulation of iron in ferritin nanocage has gathered remarkable attention over the decades owing to its exceptional relevance in health and diseases [1, 4, 14, 19, 27, 32]. Numerous studies have been focused on exploring mechanisms involved in the iron acquisition/accumulation process in ferritins from different organisms, and in structurally altered ferritins, to shed light on structure/function relationships and to understand the ferritin self-assembly [4–6, 8, 12, 13, 15, 23, 27]. Recently, the intrinsic nature (i.e., iron scavenging/accumulating and reversible self-assembly) of ferritin have been exploited to synthesize superparamagnetic iron oxide nanoparticles and various important size-constrained nanoparticles (as nanoreactor) for biomedical applications such as magnetic resonance imaging (MRI) [18, 23, 25, 33].

This chapter describes the superior ability of frog M ferritin to store large amounts of iron and investigation of iron mineral core formed inside the nanocages by non-denaturing polyacrylamide gel electrophoresis (native-PAGE) and transmission electron microscopy (TEM). The protocol also describes the spectrophotometric quantification of the total amount of iron accumulated inside the ferritin nanocages via acid/heat-based denaturation trailed by ferrozine assay. This protocol can also be used to study ferritin self-assembly (cage integrity) and their iron loading ability for all the superfamily of ferritins, despite their mechanistic differences of iron acquisition and mineralization reaction [6, 12, 27, 34].

2 Materials

2.1 Preparation of Ferritin and Fe^{2+} Solution for *In Vitro* Iron Mineralization

1. Ferritin solution: Overexpress and purify the frog M ferritin in *E. coli* [BL21(DE3)pLysS] as reported earlier [15, 34, 35] and determine the stock concentration by Bradford assay taking bovine serum albumin as standard (*see Note 1*).
2. Working buffer: 100 mM MOPS with 100 mM NaCl (pH 7.0): Dissolve 11.56 g of MOPS with sodium salt in 500 mL (*see Note 2*).

3. 50 mM FeSO₄ solution: Dissolve 13.9 mg of FeSO₄.7H₂O in 1 mL of 1 mM HCl (*see Note 3*).

2.2 Preparation of Non-denaturing Polyacrylamide Gel for Checking Iron Accumulation in Ferritin Cage

1. Resolving gel buffer: 50 mL of 1.5 M Tris-free base (pH 8.8). Weigh 9.1 g of Tris-free base and transfer it to a measuring cylinder. Add 30 mL of distilled water. Mix and adjust the pH to 8.8 with HCl. Make up the volume to 50 mL and store at 4 °C.
2. Stacking gel buffer: 50 mL of 1 M Tris-free base (pH 6.8). Weigh 6.07 g of Tris-free base and transfer it to a measuring cylinder. Add 30 mL of distilled H₂O. Mix and adjust the pH to 6.8 with HCl. Make up the volume to 50 mL and store at 4 °C.
3. Acrylamide/bis solution: Weigh 10 g and 0.265 g of acrylamide and bis-acrylamide, respectively, and transfer to a 100 mL graduated glass beaker. Add 10 mL of distilled H₂O and slightly heat the samples at ~60–70 °C for ~1–2 min (*see Note 4*). Make up the volume to 25 mL and store in a bottle wrapped with aluminum foil or an amber bottle at 4 °C.
4. 10% Ammonium persulfate: 0.1 g in 1 mL of distilled water (*see Note 5*).
5. N, N, N, N'-Tetramethyl-ethylenediamine (TEMED).
6. 10× Native-PAGE running buffer (1 L): Weigh 144 g of Glycine and 30.2 g of Tris-free base. Add distilled water to make a final volume of 1 L and store at room temperature. The solution is diluted to 1× for electrophoresis, i.e., 100 mL of 10× buffer added to 900 mL of distilled H₂O. The pH of the 1× solution should be ~8.3, without adjustment.
7. For iron staining (Prussian blue formation): (a) 2% K₄FeCN₆; 10 g in 500 mL of distilled water (47.35 mM) and store the solution in dark place. (b) 2% HCl; 10 mL (from ~12 N stock concentration) in 500 mL of distilled water (~ 240 mM).
8. 2× sample loading buffer (10 mL): Weigh 10 mg of bromophenol blue and add 2.5 mL of stacking gel buffer (pH 6.8), 2 mL of 20% glycerol and make up the volume by adding 5.5 mL of distilled water. Store one aliquot at 4 °C and rest of the aliquots at –20 °C.
9. Staining solution (500 mL): Weigh 0.25 g of Coomassie Brilliant Blue R250 and dissolve in 200 mL methanol. Then add 25 mL acetic acid and make up the volume by adding 275 mL of distilled water.
10. Destaining solution (1 L): Mix 100 mL methanol, 100 mL acetic acid, and 800 mL distilled water.

2.3 Preparation of Ferritin-Mount Grid for Transmission Electron Microscopy (TEM)

1. 1 mg/mL of mineralized ferritin sample (~ 500 Fe/cage).
2. 1% (v/v) stock solution of Gadolinium acetate tetrahydrate (uranyl acetate alternative): Add 2 µL in 200 µL of Millipore water.
3. 300-mesh carbon (type B)-coated copper grid.

2.4 Quantification of Accumulated Iron inside Ferritin Nanocage

1. PD-10 column (GE HealthCare).
2. 2 M HCl: Add 1.67 mL of HCl fuming 37% to 8.33 mL of distilled water to make a final volume of 10 mL.
3. 20 mM NH₂OH.HCl: Dissolve 0.0138 g of NH₂OH.HCl in 10 mL of distilled water.
4. 2.5 M CH₃COONa: Dissolve 2.05 g of CH₃COONa in 10 mL of distilled water.
5. 1 M NaOH: Dissolve 0.4 g of NaOH in 10 mL of distilled water.
6. 15 mM ferrozine: Dissolve 0.0738 g of ferrozine in 10 mL of distilled water and store in dark condition.
7. 2 µM mineralized ferritin.
8. 100 mM MOPS, 100 mM NaCl (pH 7.0).

2.5 Equipment and Software

1. AKTA Prime plus protein purifier (GE HealthCare).
2. HERMLE Table-Top centrifuge.
3. SHIMADZU UV-visible spectrophotometer (UV-1900).
4. Mini-PROTEAN® Tetra System Cell and PowerPac™ supplies from BIORAD.
5. FEI, Tecnai G2 TF30-ST transmission electron microscope equipped with a LaB6 electron gun operating at 300 keV.
6. pH meter.
7. Origin 8.5 Pro.
8. ImageJ Software (freeware, National Institute of Health).

2.6 Other Consumables

1. Q-Sepharose fast-flow column (GE HealthCare).
2. 30 kDa ultracentrifugal filter unit (Millipore).
3. PD-10 desalting column (GE HealthCare).
4. Whatman qualitative filter paper (Grade 1).
5. 0.2 µm Millex syringe filter.

3 Methods

3.1 In Vitro Iron Mineralization in Ferritin

- For ~500 Fe/cage: Add 4 μ L of freshly prepared 50 mM FeSO_4 solution (final concentration; 1 mM) to 2 μM of ferritin cage (~1 mg/mL) in 100 mM MOPS, 100 mM NaCl (pH 7.0) to achieve 500 Fe/cage in 200 μL .
- Similarly, add 8–36 μ L of freshly prepared 50 mM FeSO_4 solution (final concentration; 2–9 mM) sequentially to 2 μM of ferritin cage (~1 mg/mL) in 100 mM MOPS, 100 mM NaCl (pH 7.0) to achieve 0–4500 Fe/cage in 200 μL .
- Finally, incubate these solutions for 2 h in room temperature followed by overnight incubation at 4 °C (Fig. 2a). For control reactions, add similar amounts of FeSO_4 to (1) buffer: 100 mM MOPS/100 mM NaCl (pH 7.0) and (2) bovine serum albumin in 100 mM MOPS/100 mM NaCl (pH 7.0) (Fig. 2b).

3.2 Checking Iron Accumulation in Ferritin Cage by Native-PAGE

- Preparation of 5% (w/v) non-denaturing polyacrylamide gel: Mix 1.25 mL of resolving buffer, 0.625 mL of acrylamide/bis solution, and 3.075 mL of distilled water in a 25 mL Buchner flask and degas the solution (Table 1). Add 50 μL of 10% ammonium persulfate solution, 3.5 μL of TEMED and mix

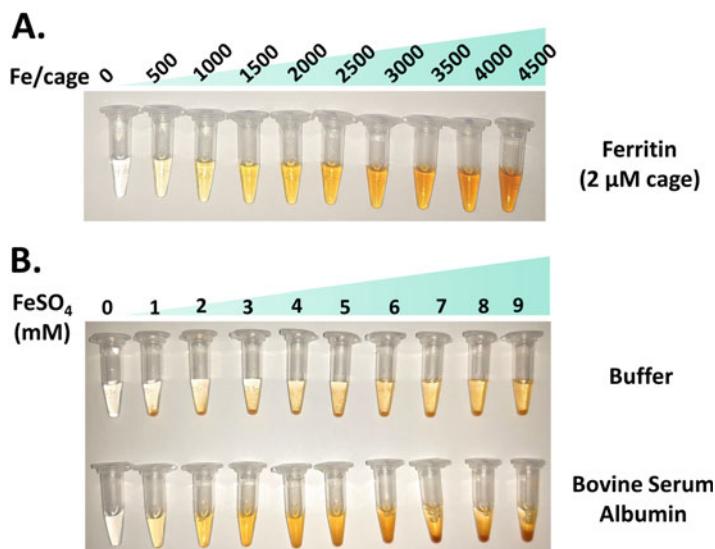


Fig. 2 Comparison of iron mineralization in ferritin protein cage, buffer, and bovine serum albumin. (a) The soluble ferritin protein nanocage rapidly sequesters Fe^{2+} and synthesizes ferric oxyhydroxide mineral in its central nanocavity by ferroxidase activity and thus prevents autoxidation/precipitation in buffer (outside the ferritin cage). (b) Addition of FeSO_4 to either buffer (100 mM MOPS/100 mM NaCl, pH 7.0) or bovine serum albumin (in same buffer) results in precipitation (due to formation of insoluble ferric oxyhydroxide aggregates)

Table 1
Composition of different percent (%) of resolving gel for native-PAGE (see Note 7)

Components	5%	6%	7%
Acrylamide/bis solution	0.625 mL	0.750 mL	0.875 mL
1.5 M resolving gel buffer (pH 8.8)	1.250 mL	1.250 mL	1.250 mL
Distilled water	3.075 mL	2.947 mL	2.822 mL
10% ammonium persulfate	50 µL	50 µL	50 µL
TEMED	3.5 µL	3.5 µL	3.5 µL
Total volume	~5 mL	~5 mL	~5 mL

the solution properly. Cast the gel using spacer plate (1 mm integrated spacer) and short plate on a casting stand. Leave space for stacking gel and slowly overlay with ethanol or water (*see Note 6*). Leave the resolving gel for 15–20 min to polymerize and solidify.

2. Mix 0.5 mL of stacking gel buffer, 0.2 mL of acrylamide/bis solution, and 1.3 mL of distilled water in a 25 mL Buchner flask and degas the solution. Add 20 µL of 10% ammonium persulfate solution and 1.8 µL of TEMED and mix the solution properly. Cast the stacking gel by pouring the above solution after removing the overlaid water/ethanol and immediately put the 10-well gel comb without creating any air bubbles. Leave the stacking gel for 20–30 min to polymerize and solidify.
3. Remove the 10-well comb gently without rupturing the wells. Mix 20 µL (i.e., 20 µg) of iron-loaded ferritin samples with 6 µL of 2× loading dye and load the mixture on to the wells. Electrophorese at 100 V for 1 h in ice (*see Note 8*) and continue till the dye front (from bromophenol blue) reaches the bottom of the gel.
4. After electrophoresis, pry the glass plates with the help of a gel releaser. The gel remains on one of the glass plates. Carefully transfer the gel and proceed for iron staining/protein staining.
5. For iron staining: Following electrophoresis, transfer the gel to a staining box containing 2% HCl and 2% K₄FeCN₆ solution in 1:1 (v/v) ratio. Incubate the gel for ~30 min–1 h for the appearance of Prussian blue band (Fig. 3a). The Fe³⁺ ions accumulated inside the ferritin nanocage as iron mineral reacts with acidified K₄FeCN₆ solution to form Prussian blue precipitate (Eq. 1).



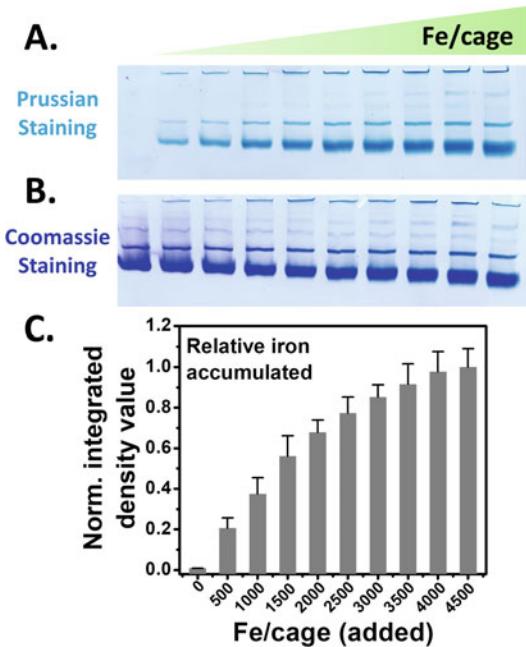


Fig. 3 Native-PAGE-based analysis of iron accumulation in ferritin. Ferritin protein samples loaded with increasing amount of iron (mineralized ferritin) were run in a 5% (w/v) non-denaturing gel at 100 V for 1 h in ice. After electrophoresis, the gel was treated with acidified K_4FeCN_6 solution to visualize the accumulated ferric iron mineral inside the ferritin by formation of Prussian blue precipitate (**a**), followed by visualization of ferritin protein cage by Coomassie staining (**b**). The relative amount of “accumulated iron” within the ferritin protein nanocage was estimated from normalized integrated density values using ImageJ analysis (**c**)

6. For protein staining: Wash the gel with distilled water and stain with Coomassie for ~30 min. Decant the staining solution carefully and add destaining solution to remove the dye from protein-free parts (gel matrix) to visualize the protein bands only (Fig. 3b, see Note 9).
7. Relative quantification of accumulated iron inside the ferritin nanocage: ImageJ software can be used to estimate the relative amount of iron accumulated inside the ferritin protein cage from the Prussian staining profiles (normalized integrated density values). The values can be normalized or plotted as such to obtain the histograms represented in Fig. 3c.

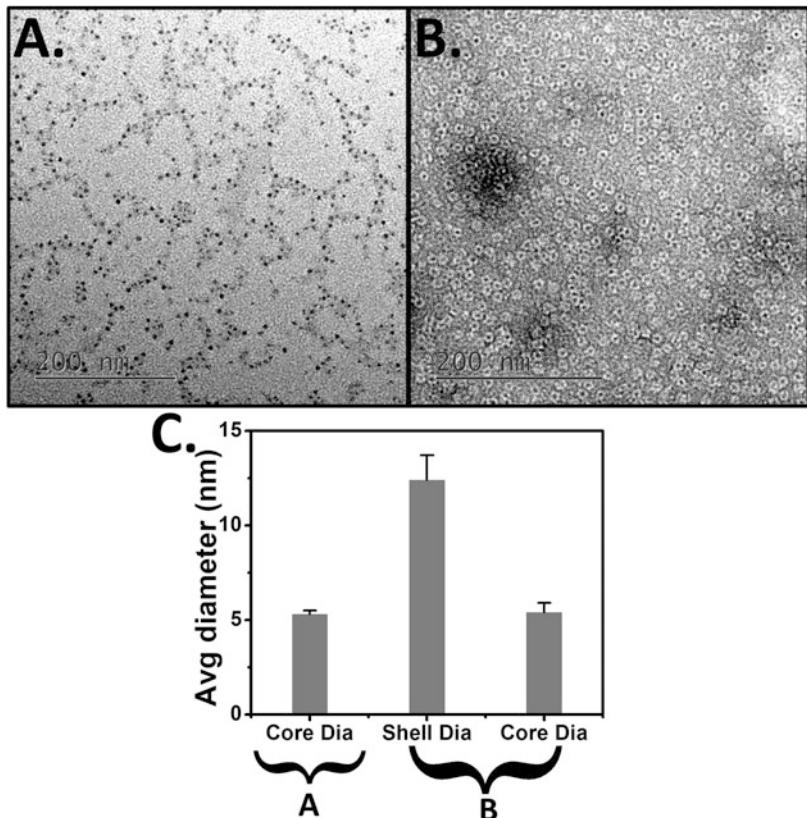


Fig. 4 TEM micrographs of iron oxide nanoparticles accumulated inside the ferritin nanocage. 1 mg/mL of ferritin loaded with 500 Fe/cage without negative staining (**a**) and with 1% (v/v) gadolinium acetate tetrahydrate (uranyl acetate alternative) staining (**b**). Scale bars are 200 nm as indicated at the bottom left-hand side of each panel. (**c**) Average sizes of ferritin protein cages and iron mineral cores were estimated using ImageJ analysis from the TEM images of stained (i.e., **a**) and unstained (i.e., **b**) ferritin samples

3.3 Estimation of Size of Ferritin Iron Mineral Core and its Shell by Transmission Electron Microscopy (TEM)

1. Visualization of iron mineral core (without negative staining): Deposit 5 μ L of 1 mg/mL mineralized ferritin sample (~500 Fe/cage) on to a 300-mesh carbon-coated copper grid and allow it to absorb on the grid for ~1 min. Wash the grid with Millipore water and soak away the excess solution using a Whatman filter paper. Air-dry the grid in a vacuum desiccator and place them into a grid chamber. Observe the grid with a transmission electron microscope equipped with a LaB6 electron gun operating at 300 keV to visualize the bare iron mineral core (Fig. 4a).
2. Visualization of iron mineral core with ferritin protein shell (with negative staining): Follow the same procedure for grid preparation as mentioned above. Stain the water washed ferritin-mount grid twice with 20 μ L of 1% (w/v) gadolinium acetate tetrahydrate (uranyl acetate alternative) solution (for 30 s) (see Note 10). Soak away the excess solution with filter

paper and air-dry the grids in a vacuum desiccator. Record the TEM micrographs to visualize the iron mineral core along with its ferritin protein shell (Fig. 4b).

3. Analysis of size of ferritin iron mineral core and its shell: Use the ImageJ software to estimate the average diameters of both ferritin shell (for negatively stained samples) and its encapsulated iron mineral core along with shell (for both stained and unstained samples). Calibrate the image manually by establishing the image scale using the scale bar at the bottom left-hand side of the image. Adjust the threshold of the image and analyze the particle size. For negatively stained images, draw a circle around the particle of interest (≥ 50 particles from a single micrograph) and calculate the diameter of that circle (Fig. 4c).

3.4 Quantification of Accumulated Iron inside the Ferritin Nanocage

Step 1: Pass the mineralized ferritin sample through a PD-10 column and 30 kDa ultracentrifugal membrane filter to remove the unbound iron (if any) and determine the concentration of ferritin cage by Bradford assay.

Step 2: Ferritin cage disintegration and iron mineral reduction/dissolution: Prepare a solution (total volume 500 μ L) comprising of following; 10 μ L of 2 μ M mineralized ferritin (final concentration; 0.04 μ M cage in 500 μ L), 125 μ L of 2 M HCl (final concentration; 500 mM in 500 μ L), and 125 μ L of 20 mM $\text{NH}_2\text{OH} \cdot \text{HCl}$ (final concentration; 5 mM in 500 μ L). Incubate the solution for 30 min at room temperature and then boil it for ~10 min.

Step 3: Neutralization and Fe^{2+} chelation: Add 250 μ L of 15 mM ferrozine (final concentration; 2.5 mM) and 500 μ L of 2.5 M CH_3COONa followed by ~200 μ L of 1 M NaOH to neutralize the solution. Add ~50 μ L of distilled water to make up the volume to 1500 μ L.

Step 4: Incubate the solution (from step 3) for 30 min. Centrifuge the solution ~10,000 rpm for 10–15 min at room temperature to remove protein aggregates (if any).

Step 5: Collect 1 mL of supernatant and record the UV-visible absorption spectra to quantify the amount of iron accumulated per ferritin cage by using the absorbance value at 562 nm for the formation of $[\text{Fe}(\text{Fz})_3]^{4-}$ complex ($\epsilon = 25.4 \text{ mM}^{-1} \text{ cm}^{-1}$) (Fig. 5, see Notes 11 and 12).

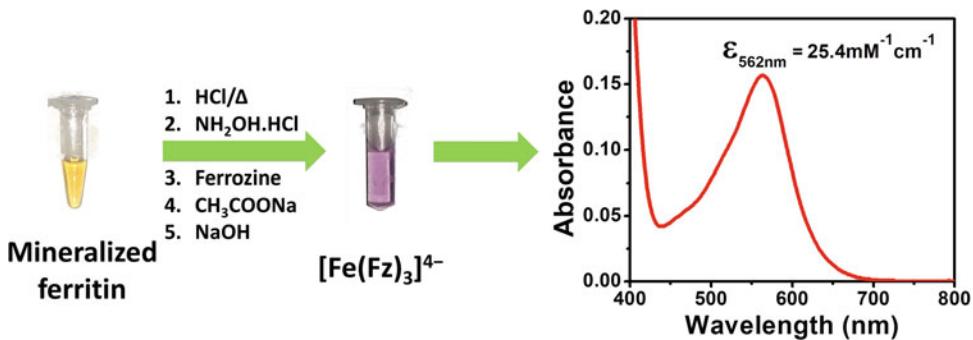


Fig. 5 Quantification of iron accumulated inside the ferritin nanocage. The mineralized ferritin sample (~500 Fe/cage; added) is disassembled by acid and heat treatment followed by reduction by using NH₂OH. HCl. Fe²⁺ chelator, ferrozine, was added to the solution and its pH was neutralized by adding CH₃COONa and NaOH. The purple color developed due to the formation of [Fe(Fz)₃]⁴⁻ complex was analyzed, which corresponds to the accumulated iron (~460 ± 30 Fe/cage) inside the ferritin nanocage

4 Notes

1. Recombinant frog M ferritin was overexpressed in *E. coli* [BL21(DE3)pLysS] and purified by heat treatment and ammonium sulfate precipitation followed by ion-exchange chromatography (Q-Sepharose fast-flow column; GE HealthCare), as reported earlier [34, 35]. Finally, the purified ferritin fractions were concentrated using 30 kDa ultracentrifugal filter units (Millipore) and its concentration was determined using Bradford assay.
2. Usually high concentrations of buffers are used for iron mineralization process to minimize the pH drop due to the generation of H⁺ ions during ferroxidase/mineralization reaction. A variety of buffers (MOPS, MES (pH 6.0–6.5), HEPES, etc.) can be used based on the stability and solubility of ferritins.
3. FeSO₄ solutions are prepared freshly, just before its addition to ferritin protein, in 1 mM HCl to minimize its autoxidation. Ammonium Fe(II) sulfate (Mohr's salt) can also be used instead of ferrous sulfate.
4. The mixture can be left overnight in a stirrer in order to dissolve completely. However, heating the mixture for ~1–2 min at ~60–70 °C can solubilize the acrylamide/bis instantly. Extreme caution must be taken while weighing and handling acrylamide, as the unpolymerized acrylamide is a neurotoxin.
5. For best results, prepare ammonium persulfate (APS) solution freshly each time before use.

6. If the acrylamide/bis solution is lower than 8%, overlay with water and for gels of 10% or greater acrylamide/bis concentration, use 20% ethanol. Atmospheric oxygen can prevent the polymerization of acrylamide and overlaying with water or ethanol inhibits contact with atmospheric oxygen. In addition, this also helps to level the resolving gel.
7. Using 4% non-denaturing gel saves time but is too delicate to handle, which may compromise with the resolution (sieving). In contrast, higher percentage (>7%) polyacrylamide gels may prevent maxi ferritins (24-mer) to enter. However, 10% gels can be used for mini ferritins (12-mer, Dps proteins).
8. Electrophoresis at high voltage not only deforms the gel but also denatures the proteins' structure (may affect ferritin cage assembly). Hence, it is advisable to keep the Mini-PROTEAN® Tetra System Cell in ice during electrophoresis.
9. The Prussian blue staining followed by Coomassie staining can differentiate the formation of non-specific aggregates (due to autoxidation and hydrolysis reactions) from the ferritin iron mineral.
10. Although aqueous solution of uranyl acetate is extensively used as a staining reagent for the TEM of biological samples, its biological toxicity and environmentally hazardous radioactive nature restricts its use under the regulation of nuclear fuel materials. However, gadolinium triacetate (1–10% w/v) is non-radioactive and possesses excellent ability to stain various biological samples.
11. Other Fe^{2+} chelators, such as 2, 2-Bipyridine [36] and 1,10-Phenanthroline, can also be used, but ferrozine is preferred due to its stable complex, $[\text{Fe}(\text{Fz})_3]^{4-}$, over a wide range of pH (4–9) and its high extinction coefficient ($25.4 \text{ mM}^{-1} \text{ cm}^{-1}$).
Iron quantification from Fig. 5: $0.04 \mu\text{M}$ of ferritin cage in $500 \mu\text{L}$ finally diluted to $1500 \mu\text{L}$, i.e., three times, so the final ferritin concentration becomes $0.0133 \mu\text{M}$. As $A_{562\text{nm}} = 0.157$; $[\text{Fe}^{2+}] = 6.2 \mu\text{M}$, i.e., $0.0133 \mu\text{M}$ of ferritin contains $6.2 \mu\text{M}$ of iron. Hence, number of iron per cage = $6.2/0.0133 = 466 \text{ Fe/cage}$.
However, higher extinction coefficients for $[\text{Fe}(\text{Fz})_3]^{4-}$ at 562 nm are also used ($\sim 27.9\text{--}31.0 \text{ mM}^{-1} \text{ cm}^{-1}$) depending upon the experimental conditions such as ferrozine:iron [37], concentration of buffer, aerobic vs. anaerobic conditions, etc.
12. The amount of incorporated iron can also be quantified using techniques such as atomic absorption spectroscopy (AAS) or inductively coupled plasma optical emission spectrometry (ICP-OES)/mass spectrometry (ICP-MS).

Acknowledgments

This work was supported by the Science and Engineering Research Board (SERB), India (CRG/2020/005332) to R.K.B. The authors are thankful to Prof. Elizabeth C. Theil (CHORI and NCSU, USA) for her generous support in providing the ferritin clone. The authors are also thankful to Mr. Rohit Kumar Raut, Mr. Abhinav Mohanty, Ms. Swagatika Mallik, and Ms. Tanaya Subudhi for their experimental assistance.

References

- Theil EC, Toshia T, Behera RK (2016) Solving biology's iron chemistry problem with ferritin protein nanocages. *Acc Chem Res* 49:784–791
- Crichton R (2016) The essential role of iron in biology. In: Iron metabolism: from molecular mechanisms to clinical consequences. Wiley, p 22
- Koppenol WH, Hider RH (2019) Iron and redox cycling. Do's and don'ts. *Free Radic Biol Med* 133:3–10
- Theil EC, Behera RK, Toshia T (2013) Ferritins for chemistry and for life. *Coord Chem Rev* 257:579–586
- Theil EC (2013) Ferritin: the protein nanocage and iron biominerals in health and in disease. *Inorg Chem* 52:12223–12233
- Mohanty A, Subhadarshanee B, Barman P, Mahapatra C, Aishwarya B, Behera RK (2019) Iron Mineralizing bacterioferritin A from Mycobacterium tuberculosis Exhibits Unique Catalase-Dps-like Dual Activities. *Inorg Chem* 58:4741–4752
- Arosio P, Ingrassia R, Cavadini P (2009) Ferritins: a family of molecules for iron storage, antioxidation and more. *Biochim Biophys Acta* 1790:589–599
- Bou-Abdallah F (2010) The iron redox and hydrolysis chemistry of the ferritins. *Biochim Biophys Acta, Gen Subj* 1800:719–731
- Theil EC, Behera RK (2013) The chemistry of nature's iron biominerals in ferritin protein nanocages. In: Ueno T, Watanabe Y (eds) Coordination chemistry in protein cages. John Wiley & Sons, Inc., p 3
- Chasteen ND, Harrison PM (1999) Mineralization in ferritin: an efficient means of iron storage. *J Struct Biol* 126:182–194
- Treffry A, Bauminger ER, Hechel D, Hodson NW, Nowik I, Yewdall SJ, Harrison PM (1993) Defining the roles of the threefold channels in iron uptake, iron oxidation and iron-core formation in ferritin: a study aided by site-directed mutagenesis. *Biochem J* 296:721–728
- Parida A, Mohanty A, Kansara BT, Behera RK (2020) Impact of phosphate on iron mineralization and mobilization in Nonheme Bacterioferritin B from mycobacterium tuberculosis. *Inorg Chem* 59:629–641
- Mohanty A, Mithra K, Jena SS, Behera RK (2021) Kinetics of ferritin self-assembly by laser light scattering: impact of subunit concentration, pH, and ionic strength. *Biomacromolecules* 22:1389–1398
- Chandramouli B, Bernacchioni C, Di Maio D, Turano P, Brancato G (2016) Electrostatic and structural bases of Fe²⁺ translocation through ferritin channels. *J Biol Chem* 291:25617–25628
- Behera RK, Theil EC (2014) Moving Fe²⁺ from ferritin ion channels to catalytic OH centers depends on conserved protein cage carboxylates. *Proc Natl Acad Sci U S A* 111:7925–7930
- Ceci P, Di Cecca G, Falconi M, Oteri F, Zamparelli C, Chiancone E (2011) Effect of the charge distribution along the "ferritin-like" pores of the proteins from the Dps family on the iron incorporation process. *J Biol Inorg Chem: JBIC* 16:869–880
- Maity B, Fujita K, Ueno T (2015) Use of the confined spaces of apo-ferritin and virus capsids as nanoreactors for catalytic reactions. *Curr Opin Chem Biol* 25:88–97
- Mohanty A, Parida A, Raut RK, Behera RK (2022) Ferritin: a promising nanoreactor and nanocarrier for bionanotechnology. *ACS Bio & Med Chem Au* 2:258–281
- Zhao G, Bou-Abdallah F, Arosio P, Levi S, Janus-Chandler C, Chasteen ND (2003) Multiple pathways for mineral core formation in mammalian apoferritin. The role of hydrogen peroxide. *Biochemistry* 42:3142–3150

20. Koochana PK, Mohanty A, Parida A, Behera N, Behera PM, Dixit A, Behera RK (2021) Flavin-mediated reductive iron mobilization from frog M and mycobacterial ferritins: impact of their size, charge and reactivities with NADH/O₂. *J Biol Inorg Chem*: JBIC 26:265–281
21. Koochana PK, Mohanty A, Das S, Subhadarshanee B, Satpati S, Dixit A, Sabat SC, Behera RK (2018) Releasing iron from ferritin protein nanocage by reductive method: the role of electron transfer mediator. *Biochim Biophys Acta GenSubj* 1862:1190–1198
22. Harrison PM, Arosio P (1996) The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim Biophys Acta* 1275:161–203
23. Mehlenbacher M, Poli M, Arosio P, Santambrogio P, Levi S, Chasteen ND, Bou-Abdallah F (2017) Iron oxidation and core formation in recombinant heteropolymeric human ferritins. *Biochemistry* 56:3900–3912
24. Levi S, Yewdall SJ, Harrison PM, Santambrogio P, Cozzi A, Rovida E, Albertini A, Arosio P (1992) Evidence of H- and L-chains have co-operative roles in the iron-uptake mechanism of human ferritin. *Biochem J* 288:591–596
25. Uchida M, Kang S, Reichhardt C, Harlen K, Douglas T (2010) The ferritin superfamily: supramolecular templates for materials synthesis. *Biochim Biophys Acta* 1800:834–845
26. Chiancone E, Ceci P (2010) The multifaceted capacity of Dps proteins to combat bacterial stress conditions: detoxification of iron and hydrogen peroxide and DNA binding. *Biochim Biophys Acta Gen Subj* 1800:798–805
27. Levi S, Rovida E (2015) Neuroferritinopathy: from ferritin structure modification to pathogenetic mechanism. *Neurobiol Dis* 81:134–143
28. Galvez N, Fernandez B, Sanchez P, Cuesta R, Ceolin M, Clemente-Leon M, Trasobares S, Lopez-Haro M, Calvino JJ, Stephan O, Dominguez-Vera JM (2008) Comparative structural and chemical studies of ferritin cores with gradual removal of their iron contents. *J Am Chem Soc* 130:8062–8068
29. Theil EC (2003) Ferritin: at the crossroads of iron and oxygen metabolism. *J Nutr* 133: 1549S–1553S
30. Bradley JM, Moore GR, Le Brun NE (2014) Mechanisms of iron mineralization in ferritins: one size does not fit all. *J Biol Inorg Chem*: JBIC 19:775–785
31. Polanams J, Ray AD, Watt RK (2005) Nano-phase iron phosphate, iron arsenate, iron vanadate, and iron molybdate minerals synthesized within the protein cage of ferritin. *Inorg Chem* 44:3203–3209
32. Mohanty A, Parida A, Subhadarshanee B, Behera N, Subudhi T, Koochana PK, Behera RK (2021) Alteration of Coaxial Heme Ligands Reveals the Role of Heme in Bacterioferritin from *Mycobacterium tuberculosis*. *Inorg Chem* 60:16937–16952
33. Sana B, Poh CL, Lim S (2012) A manganese-ferritin nanocomposite as an ultrasensitive T2 contrast agent. *Chem Commun* 48:862–864
34. Subhadarshanee B, Mohanty A, Jagdev MK, Vasudevan D, Behera RK (2017) Surface charge dependent separation of modified and hybrid ferritin in native PAGE: impact of lysine 104. *Biochim Biophys Acta Proteins Proteom* 1865:1267–1273
35. Toshia T, Hasan MR, Theil EC (2008) The ferritin Fe2 site at the diiron catalytic center controls the reaction with O₂ in the rapid mineralization pathway. *Proc Natl Acad Sci U S A* 105:18182–18187
36. Koochana PK, Mohanty A, Subhadarshanee B, Satpati S, Naskar R, Dixit A, Behera RK (2019) Phenothiazines and phenoxyazines: as electron transfer mediators for ferritin iron release. *Dalton Trans* 48:3314–3326
37. Smith GL, Reutovich AA, Srivastava AK, Reichard RE, Welsh CH, Melman A, Bou-Abdallah F (2021) Complexation of ferrous ions by ferrozine, 2,2'-bipyridine and 1,10-phenanthroline: implication for the quantification of iron in biological systems. *J Inorg Biochem* 220:111460



Chapter 8

A Generalized Method for Metal Fixation in Horse Spleen L-Ferritin Cage

Basudev Maity and Takafumi Ueno

Abstract

The naturally occurring iron storage protein, ferritin, has been recognized as an important template for preparing inorganic nanomaterials by fixation of metal ions and metal complexes into the cage. Such ferritin-based biomaterials find applications in various fields like bioimaging, drug delivery, catalysis, and biotechnology. The unique structural features with exceptional stability at high temperature up to *ca.* 100 °C and a wide pH range of 2–11 enable to design the ferritin cage for such interesting applications. Infiltration of metals into ferritin is one of the key steps for preparing ferritin-based inorganic bionanomaterials. Metal-immobilized ferritin cage can be directly utilized for applications or act as a precursor for synthesizing monodisperse and water-soluble nanoparticles. Considering this, herein, we have described a general protocol on how to immobilize metal into a ferritin cage and crystallize the metal composite for structure determination.

Key words Ferritin protein cage, Metal immobilization, Purification, Crystallization

1 Introduction

Protein cage is one of the most investigated natural protein assembly structures due to their promising applications in chemical, materials, and biomedical science [1–3]. They have unique structural features with a confined environment ranging from 5 nm to 100 nm diameter, which is connected through several channels across which foreign molecules can transport into the cage [4]. The original roles of such confined environments are to store genetic materials or biomineratization. Virus capsids, DPS (DNA binding proteins from starved cells), heat shock proteins, and ferritin are examples of typical protein cages found in nature [5]. Among various protein cages, ferritin is one of the most studied biomolecules due to its unique structural features with exceptional stability at high temperatures and wide pH range of 2–11 [3, 6–8].

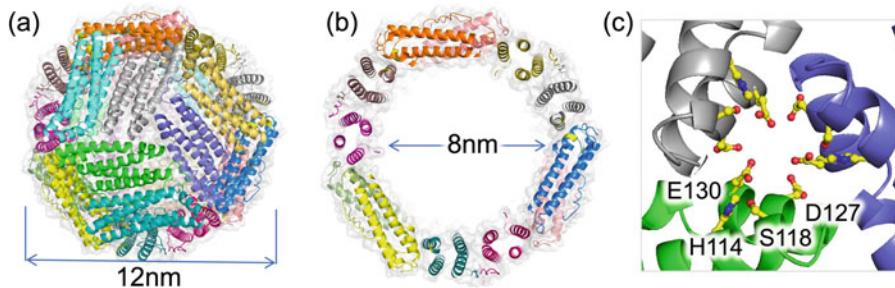


Fig. 1 Structural features of the ferritin cage. **(a)** The whole structure of a 24-mer L-ferritin cage. **(b)** Interior cage structure. **(c)** Symmetric threefold channel commonly known for the entry of metal ions, metal complexes, etc., into the cage. His114, Ser118, Asp127, and Glu130 are located on the surface of the channel. PDB model 1DAT is used to make the figures in Pymol. [24]

The ferritin cage is composed by the self-assembly of 24 sub-units with an outer diameter of 12 nm and an internal cavity diameter of 8 nm (Fig. 1a, b). The cage is connected through several pores, of which eight threefold symmetric channels are known for the transport of metal ions into the cage (Fig. 1c). The internal surface of the threefold channel is composed of Histidine, Cysteine, and anionic residues like Glutamic acid and Aspartic acid, which facilitate the transport of positively charged molecules into the cage (Fig. 1c) [9]. The restricted space of the ferritin cage is utilized for metal coordination reactions and nanomaterials synthesis [3]. Such metal-bounded ferritin cages find various applications in bioimaging, catalysis, and drug delivery [7, 8]. For example, organometallic Rh(norbornadiene) complex bounded ferritin cage showed the phenylacetylene polymerization reaction [10]. Palladium nanoclusters grown in a ferritin cage can catalyze the aerobic oxidation of alcohols in an aqueous medium [11]. Gold nanocluster grown in the cage was utilized for bioimaging [12]. It was found that ferritin can accommodate small biomolecules like proteins and enzymes into the cage [13, 14]. All such examples suggest that ferritin cage is an important bio-template to develop multifunctional bio-nanocage and has a great scope for redesigning the cage to diversify the applications.

One of the most critical steps to design ferritin-based bionano-materials is to fix metal ions or synthetic metal complexes into the cage. Each metal has an individual choice of amino acid residues. Although the wild type of ferritin cage has metal-binding sites, it can be re-designed for proper metal coordination and controlling their coordination structures, properties, and activity of the functionalized ferritin cage [15, 16]. The metal complex-immobilized ferritin cages can be directly utilized for catalysis, structure determination, etc. On the other hand, metal ion-immobilized ferritin cages can be utilized as a precursor for synthesizing monodisperse and water-soluble nanoparticle synthesis. Based on the

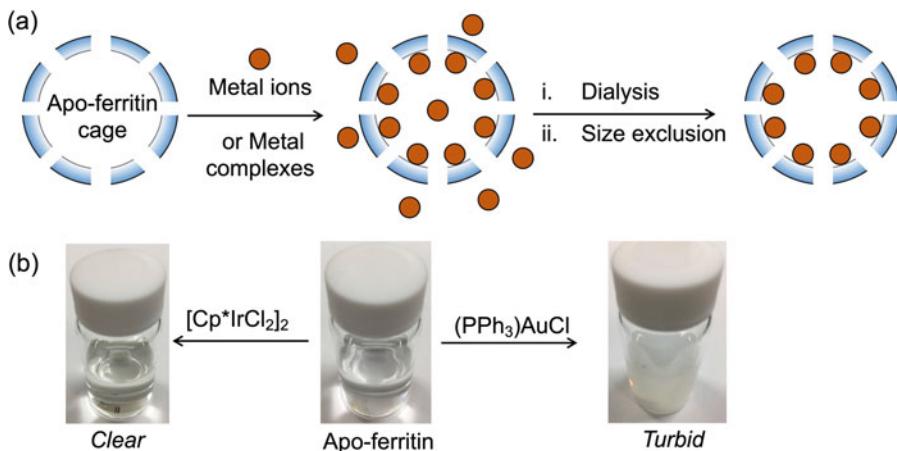


Fig. 2 Preparation of ferritin-metal conjugates. (a) Schematic representation showing the metal immobilization in ferritin cage and purification steps. (b) A typical setup of metal immobilization reaction in which an organometallic $[\text{Cp}^*\text{IrCl}_2]_2$ and $(\text{PPh}_3)\text{AuCl}$ complexes were reacted with apo-ferritin cage (WT) in Tris–HCl buffer (pH 8.0). For Ir complex, it gives a clear solution, whereas turbidity appears due to low solubility for Au complex

coordination properties of the metal ions, strategies can be made to synthesize metal nanoparticles inside the cage [17]. This is also necessary to understand the structural information of protein-nanoparticle interaction [18]. Considering all, it will be advantageous to have a general method for metal immobilization in a ferritin cage, which can be treated as a starting point to design and develop inorganic bionanomaterials.

Herein we describe the detailed methodology of immobilizing synthetic metal complexes based on the representative synthetic metal complexes including $[\text{Pd}(\text{C}_3\text{H}_5)\text{Cl}]_2$, $[\text{Cp}^*\text{IrCl}_2]_2$ [Cp^* = Pentamethyl cyclopentadienyl] and $[\text{Ru}(p\text{-Cymene})\text{Cl}]_2$ [$\text{Rh}(\text{nbd})\text{Cl}_2]_2$ [nbd = Norbornadiene], and metal ions like Au^{3+} and Pd^{2+} into the ferritin cage (Fig. 2a) [10, 15, 17–22]. We discussed the protocol step by step, including various cautions/tricks, so that the methodology can be treated as a general method for metal immobilization. In addition, the method of crystallization of the ferritin-metal conjugates has been included. Therefore, it is expected that this chapter will be helpful for general readers to design new ferritin cage-based inorganic nanomaterials.

2 Materials

2.1 Preparation of Apo-Recombinant Horse Spleen L-Ferritin

1. Luria broth (LB) medium for *E. coli* expression.
2. Nova blue cells containing the expression vector pMK2 with ampicillin resistance.

3. Shaking incubator and a high-speed refrigerated centrifuge for *E. coli* culture and harvesting.
4. High-performance Q-sepharose column for anion exchange and a high-resolution S-300 size exclusion column.
5. Syringe filter (0.2 µm).

2.2 Metal Fixation

Commercial grade metal complexes, such as $[\text{Ru}(p\text{-Cym})\text{Cl}]_2$, $[\text{Cp}^*\text{IrCl}_2]_2$ and $[\text{Pd}(\text{C}_3\text{H}_5)\text{Cl}]_2$ and $[\text{Rh}(\text{nbd})\text{Cl}_2]_2$ and KPdCl_4 , KAuCl_4 were used without any further purification. The solvents acetonitrile, methanol, dimethylformamide, etc., were used as received. Scintillation glass vial was used for reaction.

2.3 Crystallization

1. 24-well plate, cover slip, and sealant.
2. Amicon® ultracentrifugation.
3. Milli-Q water.
4. 2.0 M $(\text{NH}_4)_2\text{SO}_4$: 13.2 g of $(\text{NH}_4)_2\text{SO}_4$ in 50 mL Milli-Q water and pass through a 0.2 µm syringe filter.
5. 0.2 M CdSO_4 : 2.58 g of $3\text{CdSO}_4 \cdot 9\text{H}_2\text{O}$ in 50 mL Milli-Q water and pass through a 0.2 µm syringe filter.
(Caution! CdSO_4 is hazardous. While handling, use proper clothing. All the wastes, washings, contaminates, etc., should be kept in an isolated container for disposal.)
6. Ethylene glycol and glycerol.

2.4 Purification and Characterization

1. Dialysis bag.
2. Size exclusion column chromatography (High Resolution G200).
3. Native polyacrylamide gel electrophoresis (7.5%).
4. UV-visible spectrometer.

3 Methods

3.1 Purification of Apo-rHLFr

1. Prepare the recombinant horse spleen L-ferritin (apo-rHLFr) from lab expression followed by purification through anion-exchange and size exclusion column as follows [20, 23].
2. Take 20 g of *E. coli* pellet in 40 mL of 50 mM Tris–HCl buffer (pH 8), mix well, and spin in high-speed refrigerated centrifuge (17,000 g-force, 10 min). Discard the supernatant and add a fresh 40 mL of 50 mM Tris–HCl buffer (pH 8). Repeat the centrifugation and discard the supernatant.
3. Add 40 mL of 50 mM Tris–HCl buffer (pH 8) into the pellet and stir gently with a magnetic bar for about 15 min for homogeneous mixing. Sonicate to break the *E. coli* cells

under ice cold conditions using Branson sonifier 250. Duty cycle 50, output 8, 1 min sonication. Repeat the process 10 times with an interval of 10 min between two cycles (*see Note 1*).

4. Centrifuge at 27,000 g-force for 15 min. Collect the supernatant in a glass beaker.
5. Place the supernatant at 65 °C water bath for 15 min. Cool down to RT and centrifuge at 39,000 g-force for 25 min.
6. Collect the supernatant and pass through a 0.2 µm syringe filter before anion-exchange column purification.
7. Load the supernatant into an anion-exchange high-performance Q-sepharose column. Pass a total of 300 mL of buffer-A (50 mL Tris-HCl buffer, pH 8) with 10% buffer-B (50 mL Tris-HCl buffer, pH 8-1 M NaCl) with a flow rate of 3 mL/min. Collect the fractions (10 mL) from when increasing the buffer-B gradient from 10 to 40% with a target of 500 mL. The fractions containing ferritin should be checked by native PAGE.
8. Concentrate the ferritin fractions to ~10 mL and pass through a S-300 size exclusion column with a flow rate of 0.5 mL/min using 50 mM Tris-HCl (pH 8)/0.15 M NaCl as an eluting buffer. The ferritin fractions should be checked by both elution profile at 280 nm and native PAGE.
9. The purified protein in 50 mM Tris-HCl (pH 8.0)-0.15 M NaCl should be concentrated to ~15–20 µM, so that it can be diluted easily as per the requirement. The approximate molar extinction coefficient of the apo-rHLFr is 460,000 M⁻¹ cm⁻¹ which can be used to determine the concentration.

3.2 Immobilization of Synthetic Metal Complexes into Apo-rHLFr

A general protocol (Fig. 2) for immobilizing metal complexes like $[\text{Pd}(\text{C}_3\text{H}_5)\text{Cl}]_2$, $[\text{Cp}^*\text{IrCl}_2]_2$, $[\text{Ru}(p\text{-Cymene})\text{Cl}]_2$, $[\text{Rh}(\text{nbd})\text{Cl}_2]_2$, $[\text{Ru}(\text{CO})_3\text{Cl}_2]_2$, and $\text{Mn}(\text{CO})_5\text{Br}$ is given below.

1. Prepare a 10 mL of 10 µM of protein in 50 mM Tris-HCl/0.15 M NaCl (pH 8.0) in a scintillation glass vial (*see Note 2*).
2. Prepare a 10 mM stock solution of the metal complex in a suitable organic solvent like acetonitrile, methanol, dimethylformamide, depending on the solubility of the metal complex (*see Note 3*).
3. Add 1 mL of the metal complex solution drop by drop into the protein solution in 50 mM Tris-HCl (pH 8.0)/0.15 M NaCl to attain a final metal concentration of 200 equivalent to protein. For any other desired metal to protein ratio, the suitable amount of metal complex should be added in a similar way (*see Note 4*).

4. Continue stirring the reaction mixture at RT for 1 h (*see Note 5*).
5. After the desired reaction period, purify the protein by dialysis and size exclusion column chromatography as described in Subheading 3.3.

3.3 Purification of the Apo-rHLFr Composites

1. After the reaction, pass the reaction mixture through a 0.8 µm syringe filter and transfer it into a dialysis bag (MW Cut-off 6–8 kDa), thoroughly washed in boiling water before use.
2. Dialyze the mixture against 1.5 L of 0.15 M NaCl at 4 °C overnight (~12–16 h) (*see Note 6*).
3. Filter the dialyzed mixture through a 0.2 µm syringe filter to obtain a clear solution.
4. Purify the metal-protein composite by size exclusion column chromatography (HR G200). Pre-equilibrate the column with 0.15 M NaCl as eluting buffer before sample loading (*see Note 7*).
5. Identify the fractions containing ferritin-metal composites based on the elution profile for absorbance at 280 nm (*see Note 8*).
6. Collect and store the purified fractions in a refrigerator for characterization, quantitative analysis, crystallization, or other suitable studies.
7. For characterization of ferritin-metal conjugates, usually the UV-visible spectrum is measured and compared with that of free metal precursor. For quantification of protein and accumulated metal, a combination of bicinchoninic acid assay and inductively coupled plasma mass spectrometry can be used, respectively.
8. The ferritin-metal conjugates can be crystallized using the protocol in Subheading 3.5 to determine the coordination structure of immobilized metal ions or metal complexes.

3.4 Immobilization of Metal Ions into Apo-rHLFr

Metal ion solutions such as K_2PdCl_4 and $KAuCl_4$ are acidic in nature. A different protocol should be followed.

1. Prepare a 1 µM of protein solution (30 mL) in 0.15 M NaCl and adjust the pH to 8.5 using 10 mM NaOH (*see Note 9*).
2. Add 150 µL (for 200 eq) metal ion solution ($[K_2PdCl_4]$ or $[KAuCl_4]$) from an aqueous stock solution of 40 mM into the protein. Metal ion solution should be added slowly to ensure a homogeneous distribution.
3. Continue stirring the mixture at RT for 30 min. Since the size of metal ions is small, these can easily pass through the porous channel of the ferritin cage.

- After the desired reaction period, concentrate the reaction mixture to ~10 mL and purify the protein by dialysis and size exclusion column chromatography as described in Subheading 3.3.

3.5 Crystallization of the Apo-rHLFr Composites

- Concentrate the purified Ferritin-metal conjugate to ~15–20 mg/mL using Amicon® Ultra centrifugation. For crystallization, approximate protein concentration is ok and thus, protein concentration might be calculated using the molar extinction coefficient of apo-rHLFr ($460,000 \text{ M}^{-1} \text{ cm}^{-1}$) (see Note 10).
- Take a 24-well plate to prepare the reservoir solutions (Fig. 3). Vary the CdSO_4 concentration vertically from 12.5 mM, 15.0 mM, 17.5 mM, and 20.0 mM and vary the $(\text{NH}_4)_2\text{SO}_4$ horizontally from 0.5 M to 1 M with an interval of 0.1 M. Add Milli-Q water to make the total volume of the reservoir solution to 0.5 mL. Shake the 24-well plate gently to ensure a homogeneous mixture.
- Prepare the crystallization drop by mixing 1.5 μL of protein composite and 1.5 μL of reservoir solution on a coverslip and gently mix with a pipette (Fig. 3c) (see Note 11).
- Transfer the cover slips to the 24-well plate previously fixed with sealant to ensure a close environment.
- Allow the crystallization drop to equilibrate against the reservoir at 20 °C in an undisturbed condition (see Note 12).
- Inspect the appearance of crystals after one day. Usually, it comes within a day (see Note 13).
- Pick a single crystal (Fig. 3d) using Hampton loop, soak into precipitant containing 25% of cryoprotectant (ethylene glycol, glycerol, etc.) for 30–60 s, freeze in liquid nitrogen, and then measure X-ray diffraction (see Note 14).

4 Notes

- This is an important step and ice-cold conditions must be maintained. Final yield of purified protein largely depends on this step.
- It is not compulsory to fix the protein concentration to 10 μM . It can be reduced to even 5 μM as studied for metal complexes like $[\text{Ru}(\text{CO})_3\text{Cl}_2]_2$ and $\text{Mn}(\text{CO})_5\text{Br}$. Metal to protein ratio should be considered as per the requirement.
- The concentration of the stock solution of metal should be in such a way that after achieving desired equivalents like 100, or 200 equiv., the final organic solvent content in the reaction

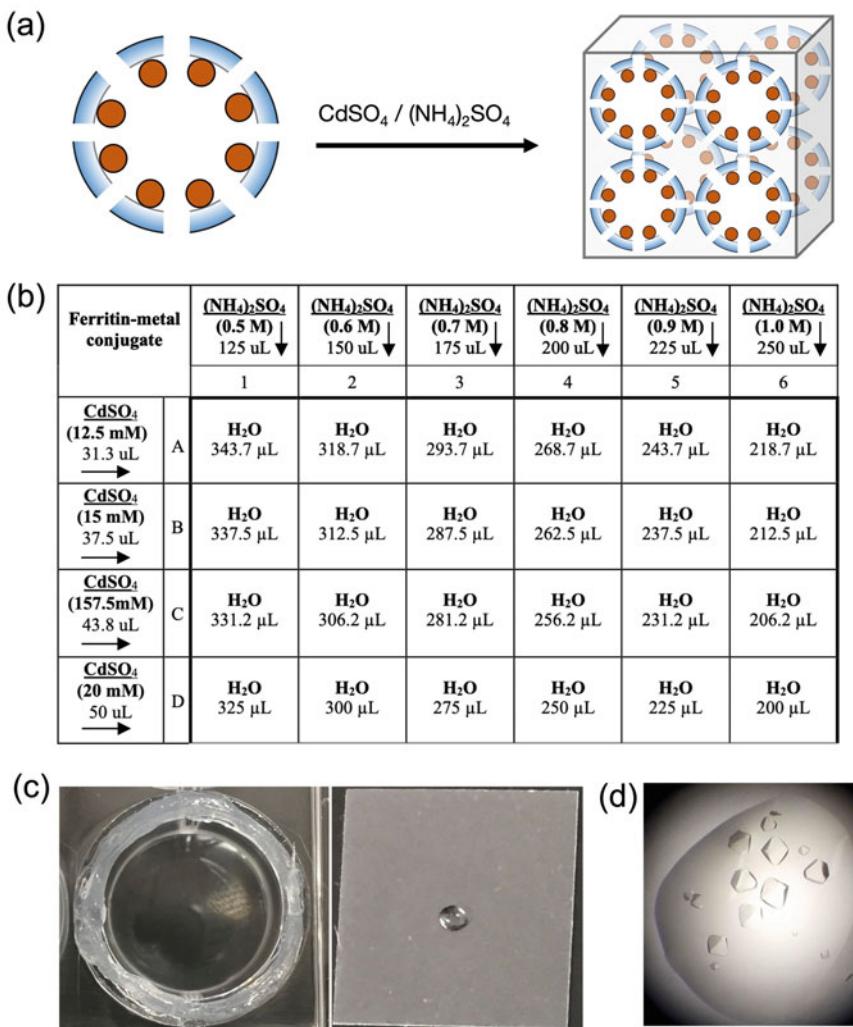


Fig. 3 Crystallization of ferritin-metal conjugates. (a) Schematic representation of the crystallization of metal immobilized ferritin cage. (b) Composition of CdSO_4 and $(\text{NH}_4)_2\text{SO}_4$ in the reservoir solutions for ferritin crystallization. The CdSO_4 and $(\text{NH}_4)_2\text{SO}_4$ are from 0.2 M and 2.0 M stock solutions, respectively. (c) An example of reservoir well and crystallization drop on the coverslip. (d) A typical example of ferritin crystals appeared in a single drop

mixture should not exceed 10% for better yield. A freshly prepared metal complex solution should be used.

4. Depending on the solubility, the addition of metal complex into protein solution might lead to turbidity or precipitation in the reaction mixture (Fig. 2b). In such situations, the reactions should be carried on as usual. We observed metal binding in the crystal structure for $[\text{Ru}(p\text{-cymene})\text{Cl}_2]_2$ even though precipitation appeared in the reaction mixture.

5. If the metal accumulation is low, the reaction temperature can be raised to 50 °C. The ferritin cage is stable up to 100 °C. We used 50 °C for the reaction with $[\text{Ru}(\text{CO})_3\text{Cl}_2]_2$, $\text{Mn}(\text{CO})_5\text{Br}$, $[\text{Cp}^*\text{IrCl}_2]_2$ complexes.
6. We did not find any changes in the number of accumulated metal ions into the ferritin cage if we exchange the buffer ($1.5 \text{ L} \times 2$) with time during dialysis. Different dialysis buffer such as 50 mM Tris–HCl (pH 8)–0.15 M NaCl or other can be used depending on the uses in next steps.
7. After the purification, the column should be adequately cleaned by passing the excess eluting buffer. The residual metal in the column could damage the matrix. For Au^{3+} ions, the column turned light purple/pink upon repeated use.
8. Although the UV profile is enough to choose fractions containing the protein, checking the fractions by native PAGE is advisable. This procedure ensures the purity of the fractions.
9. It was found that the presence of Tris–HCl buffer leads to precipitation, and therefore the reaction and crystallization was carried out in 0.15 M NaCl (pH 8.5).
10. As a quick guide for ferritin-metal conjugate crystallization, an absorption value of 0.8–1.2 for the 20th dilution of most metals is enough to obtain good diffractable single crystals ($>200 \mu\text{m}$) within a day. It is worth mentioning that at low protein concentration, the crystals appear large and take a little longer time, whereas at high protein concentration, crystals appear as quick as within 12 h, and depending on metal, it may vary. For some metal complexes or metal ions, precipitation might appear at high concentration. In those cases, pH 8 buffer containing 0.15 M NaCl should be used instead of only 0.15 M NaCl.
11. It is advisable to prepare 1–3 coverslips at a time because if many drops are prepared together, there is a chance to dry up while placing on the 24-well plate. Depending on the protein conjugates, turbidity might appear in the crystallization drop after the mixing of protein and precipitant. In such a case, crystallization should be continued, and crystals appeared in time. Otherwise, the protein concentration should be reduced.
12. Temperature affects the speed of crystallization due to variation in diffusion rate. We found 20 °C as the best condition to crystallize ferritin.
13. Sometimes, depending on metal conjugates, it may need some time to appear the crystal. The temperature of the crystallization may be varied to control the size of the crystals. Usually, at a higher metal concentration (500 equivalents to protein, etc.), the excess metal complexes bind to the surface of the protein cage and inhibit the crystallization.

14. If in house diffractometer such as Rigaku XtaLAB Synergy or other is available, single crystal can be used without freezing in liquid N₂.

Acknowledgment

This research was funded by JSPS KAKENHI Grant Nos. JP19H02830, Grant-in-Aid for Scientific Research on Innovative Areas “Molecular Engines” (JP18H05421) and 22H00347 to T.U. and 19K15695 and 20H05438 to B.M.

References

1. Uchida M, Klem MT, Allen M, Suci P, Flenniken M, Gillitzer E, Varpness Z, Liepold LO, Young M, Douglas T (2007) Biological containers: protein cages as multifunctional nanoplatforms. *Adv Mater* 19:1025–1042
2. Zhang Y, Ardejani MS, Orner BP (2016) Design and applications of protein-cage-based nanomaterials. *Chem Asian J* 11:2814–2828
3. Maity B, Fujita K, Ueno T (2015) Use of the confined spaces of apo-ferritin and virus capsids as nanoreactors for catalytic reactions. *Curr Opin Chem Biol* 25:88–97
4. Rother M, Nussbaumer MG, Renggli K, Bruns N (2016) Protein cages and synthetic polymers: a fruitful symbiosis for drug delivery applications, bionanotechnology and materials science. *Chem Soc Rev* 45:6213–6249
5. Aumiller WM, Uchida M, Douglas T (2018) Protein cage assembly across multiple length scales. *Chem Soc Rev* 47:3433–3469
6. Zhang C, Zhang X, Zhao G (2020) Ferritin nanocage: a versatile nanocarrier utilized in the field of food, nutrition, and medicine. *Nanomaterials* 10:1894
7. Bhaskar S, Lim S (2017) Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Mater* 9:e371–e371
8. Jutz G, van Rijn P, Santos Miranda B, Böker A (2015) Ferritin: a versatile building block for bionanotechnology. *Chem Rev* 115:1653–1701
9. Honarmand Ebrahimi K, Hagedoorn P-L, Hagen WR (2015) Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin. *Chem Rev* 115:295–326
10. Abe S, Hirata K, Ueno T, Morino K, Shimizu N, Yamamoto M, Takata M, Yashima E, Watanabe Y (2009) Polymerization of phenylacetylene by rhodium complexes within a discrete space of apo-ferritin. *J Am Chem Soc* 131:6958–6960
11. Kanbak-Aksu S, Nahid Hasan M, Hagen WR, Hollmann F, Sordi D, Sheldon RA, Arends IWCE (2012) Ferritin-supported palladium nanoclusters: selective catalysts for aerobic oxidations in water. *Chem Commun* 48:5745–5747
12. Sun C, Yang H, Yuan Y, Tian X, Wang L, Guo Y, Xu L, Lei J, Gao N, Anderson GJ, Liang X-J, Chen C, Zhao Y, Nie G (2011) Controlling assembly of paired gold clusters within apoferritin nanoreactor for in vivo kidney targeting and biomedical imaging. *J Am Chem Soc* 133:8617–8624
13. Tetter S, Hilvert D (2017) Enzyme encapsulation by a ferritin cage. *Angew Chem Int Ed* 56: 14933–14936
14. Hestericová M, Heinisch T, Lenz M, Ward TR (2018) Ferritin encapsulation of artificial metalloenzymes: engineering a tertiary coordination sphere for an artificial transfer hydrogenase. *Dalton Trans* 47:10837–10841
15. Abe S, Niemeyer J, Abe M, Takezawa Y, Ueno T, Hikage T, Erker G, Watanabe Y (2008) Control of the coordination structure of organometallic palladium complexes in an apo-ferritin cage. *J Am Chem Soc* 130: 10512–10514
16. Wang Z, Takezawa Y, Aoyagi H, Abe S, Hikage T, Watanabe Y, Kitagawa S, Ueno T (2011) Definite coordination arrangement of organometallic palladium complexes accumulated on the designed interior surface of apo-ferritin. *Chem Commun* 47:170–172
17. Suzuki M, Abe M, Ueno T, Abe S, Goto T, Toda Y, Akita T, Yamada Y, Watanabe Y (2009) Preparation and catalytic reaction of Au/Pd bimetallic nanoparticles in apo-ferritin. *Chem Commun* 32:4871–4873

18. Maity B, Abe S, Ueno T (2017) Observation of gold sub-nanocluster nucleation within a crystalline protein cage. *Nat Commun* 8:14820
19. Maity B, Fukumori K, Abe S, Ueno T (2016) Immobilization of two organometallic complexes into a single cage to construct protein-based microcompartments. *Chem Commun* 52:5463–5466
20. Fujita K, Tanaka Y, Sho T, Ozeki S, Abe S, Hikage T, Kuchimaru T, Kizaka-Kondoh S, Ueno T (2014) Intracellular co release from composite of ferritin and ruthenium carbonyl complexes. *J Am Chem Soc* 136:16902–16908
21. Takezawa Y, Böckmann P, Sugi N, Wang Z, Abe S, Murakami T, Hikage T, Erker G,
- Watanabe Y, Kitagawa S, Ueno T (2011) Incorporation of organometallic Ru complexes into apo-ferritin cage. *Dalton Trans* 40:2190–2195
22. Abe S, Hikage T, Watanabe Y, Kitagawa S, Ueno T (2010) Mechanism of accumulation and incorporation of organometallic Pd complexes into the protein nanocage of apo-ferritin. *Inorg Chem* 49:6967–6973
23. Iwahori K, Yoshizawa K, Muraoka M, Yamashita I (2005) Fabrication of zinc nanoparticles in the apoferritin cavity by designing a slow chemical reaction system. *Inorg Chem* 44: 6393–6400
24. <https://pymol.org/2/>



Chapter 9

Dual Modification of Artificial Protein Cage

Norifumi Kawakami, Erika Nasu, and Kenji Miyamoto

Abstract

Chemical modifications of proteins confer new functions on them or modulate their original functions. Although various approaches are developed for modifications, modifications of the two different reactive sites of proteins by different chemicals are still challenging. In this chapter, we show a simple approach for selective modifications of both interior and exterior surfaces of protein nanocages by two different chemicals based on a molecular size filter effect of the surface pores.

Key words Protein nanocage, Drug delivery, Dual surface modification, Molecular size filter effect, Protein nanoparticles

1 Introduction

Chemical modifications of proteins are approaches to confer new functions on them or modulate their original functions [1]. A broad range of chemicals is used to modify proteins, for example, small organic compounds [2–4], metal complexes [5–7], and polymer molecules [8–11]. These chemicals often contain reactive substituents to form a covalent linkage with reactive amino acids in proteins, such as cysteine and lysine residues [12]. One of the major reactive substituents in chemicals for protein modifications is the maleimide group that selectively reacts with a thiol group of a cysteine residue at neutral pH [13, 14]. Although these can react with all cysteine residues exposed at the surface of protein molecules, the replacement of cysteine residues to serine residues avoids off-target modifications. Therefore, modification of the single cysteine residue on protein is relatively easy with high specificity.

The protein nanocages [15, 16], such as ferritins and virus-like particles, are structures that have inner space. Their use as nanocapsule for drug delivery has, thus, been attracted attention for a long time [2–4, 17, 18]. The introduction of the mechanisms for precise delivery of nanocages to targeted cells and controlled release

of cargos are indispensable for this purpose. The chemical modification enables conferring these properties. For example, the chemical modification of the exterior surface by DNA aptamer or peptide that tightly bind to the receptors of targeted cells allows precise transport [19, 20], and modification of the interior surface by drug molecules with reversibly releasable conjugation allows controlled release of them [21]. However, the selective modifications of two surfaces by different molecules are still challenging for the following reasons: (1) The incorporation of drug molecules often needs decompositions of cage structures. (2) Numbers of reactive amino acids prevent modification of specific positions. (3) Each modification step requires different reaction conditions if the two chemicals have different reactive substituents.

We have developed a dual surface modification system of protein nanocage using our designed porous, hollow protein nanoparticle TIP60 [22, 23]. The method is based on the molecular size filter effect by surface pores of TIP60. This method only uses a single reactive substituent, the maleimide group, to modify cysteine residues displayed at both surfaces. The controlled release is also possible if the chemicals are modified through disulfide bonds. The principle of this method is simple and would be applicable for a wide variety of known protein nanocages.

2 Materials

2.1 Proteins and Materials for Their Purifications

1. TIP60 is a 60-mer protein nanocage with 15 and 23 nm inner and outer diameter, respectively. The molecular weight of the monomer protein is 17.8 kDa. There are 20 triangle pores at the molecular surface. The triangle pore has 4 nm edges and the inscribed circle with a diameter of 2.3 nm. The concentration of TIP60 is determined based on theoretical molar extinction coefficient $8480 \text{ M}^{-1} \text{ cm}^{-1}$ at 280 nm. The mutants S20C and S50C have cysteine residue at the exterior and interior surfaces of the nanocage, respectively. In addition to these two mutants, double mutants S20C/S50C are used for modification experiments (Fig. 1). (Any of the protein nanocages

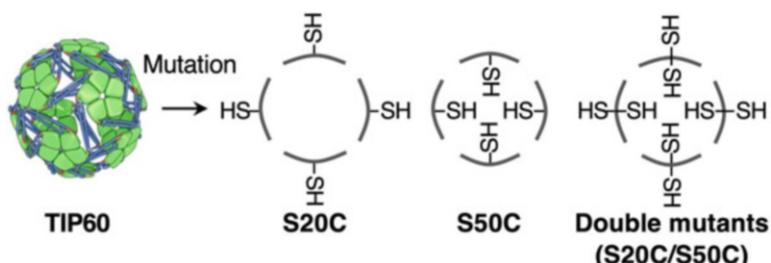


Fig. 1 Model structure of TIP60 and images of mutants

can be used instead of TIP60 if the nanocages have cysteine residues at both surfaces.)

2. LB medium: 10 g of tryptone, 5 g of yeast extract, and 10 g of NaCl are dissolved in water with adjusted volume to 1 L.
3. Chloramphenicol: 309 mM Chloramphenicol in ethanol is added to LB medium to final concentration 309 µM.
4. Isopropyl β-D-1-thiogalactopyranoside (IPTG): 0.1 M IPTG is sterilized by filtration through a membrane with 0.20 µm pores. The solution is added to LB medium to final concentration 0.1 mM.
5. Dithiothreitol (DTT): 1 M DTT.
6. Lysis buffer: 20 mM Tris–HCl, 10 mM DTT (pH 8.0).
7. Wash buffer A: 20 mM Tris–HCl, 10 mM DTT and 20 mM Imidazole (pH 8.0).
8. Wash buffer B: 20 mM Tris–HCl and 20 mM Imidazole (pH 8.0).
9. Elution buffer: 20 mM Tris–HCl, 500 mM imidazole (pH 8.0).
10. Dialysis buffer (also used for purified protein storage): 20 mM Tris–HCl, 1 mM EDTA (pH 8.0 or pH 7.4) (*see Note 1*).
11. Dialysis membrane: the molecular size cut off 12–14 kDa.

2.2 Chemicals for Modification Reactions

1. Methyl-PEG-maleimide (MM(PEG)_ns): The molecular weights are adjustable by the chain length of PEGs. In this experiment, we use MM(PEG)_ns with molecular weights of 700, 1200, 2000, 5000, 10,000, 20,000 (Table 1). These MM(PEG)_ns are dissolved in Dimethylformamide (DMF).

Table 1
Characteristics of MM(PEG)_n and modification ratios

Molecular weight	Stokes diameter (nm)	Modification ratios of	
		S50C (%)	S20C (%)
700	1.3	82 ± 3	94 ± 2
1200	1.8	65 ± 7	90 ± 1
2000	2.3	51 ± 3	91 ± 1
5000	3.8	26 ± 6	86 ± 3
10,000	5.7	ND	89 ± 1
20,000	8.3	ND	65 ± 2

Reprinted from Ref. [23] under Creative Commons license 4.0 CC BY-NC-ND

250 mM (700 and 1200: PEG molecular weight), 125 mM (2000), 50 mM (5000), 20 mM (10,000), 10 mM (20,000) are stored at -20°C .

2. 5,5-dithiobis(2-nitrobenzoic acid) (DTNB) [24]: 1.8 mM DTNB in the dialysis buffer (pH 8.0). They are prepared immediately before reactions.
3. Tris-2-carboxyethyl phosphine (TCEP): 0.5 M TCEP solution (neutral pH adjusted by NaOH).
4. 2-mercaptoethanol (2-ME): 7.2 M 2-ME in the dialysis buffer (pH 7.4).

3 Methods

3.1 Preparation of Three Mutants

1. *E. coli* BL21(DE3) is used as an expression host. The gene encoding TIP60 is introduced into the pACYCDuet vector (any of the T7 expression systems can be used).
2. Transformed cells are inoculated into 2 mL of LB medium and cultured for 16 h at 37°C at 200 rpm, then 1 mL cell culture is inoculated into 100 mL LB medium. The medium used for cell culture contains chloramphenicol.
3. After shaking at 37°C with 200 rpm for 2 h, IPTG was added for induction of TIP60 with the additional 6 h culture at the same conditions.
4. Cells are harvested by centrifugation (10,000 g, 10 min).
5. Cells harvested are suspended in the lysis buffer.
6. The cells are sonicated for 10 min with keeping the temperature at 4°C .
7. The disrupted cells are again centrifuged to remove the insoluble fraction.
8. The supernatant is collected and applied to Ni immobilized column.
9. Bound TIP60 are washed at least 6 column volumes of the wash buffer A.
10. Bound TIP60 are washed at least 5 column volumes of the wash buffer B.
11. The bound proteins are then eluted by the elution buffer. Then, 1 mM TCEP is added to the fraction containing TIP60 (*see Note 2*).
12. The TCEP-treated TIP60 is stored at 4°C and dialyzed to the dialysis buffer immediately before use. The dialysis buffer is replaced twice every 3 h, and the last one is used for 18 h (*see Notes 3 and 4*).

3.2 Optimization of the Reaction Conditions for Estimation of Pore Size of TIP60

The molecules with sizes larger than the surface pores of nanocages cannot enter the interior space. Accordingly, modification of interior cysteine residues by a series of MM(PEG)_ns is possible to estimate the average pore sizes in solution. The mutant S50C is used for the estimation of the surface pore size. Initially, the reaction conditions are optimized at the sizes and concentrations of MM(PEG)_ns and incubation times. The optimization of the reaction conditions must be performed for each protein nanocage.

1. The concentration of S50C is adjusted to 0.22 μM with a volume of 27 μL (pH 7.4).
2. MM(PEG)_ns with molecular weights of 700, 1200, 2000, 5000 are used for optimization. The 3 μL of 2–20 mM solution is added to the TIP60 solution. The concentration of MM(PEG)_ns is varied from 0.2 to 2 mM. In any case, the final concentration of DMF must be less than 10% to avoid denaturation (*see Note 5*).
3. After incubation of the samples with various times 0.5, 1, 1.5, 2, 3, 4, 5 min at 30 °C, the reaction is terminated by the addition of 3 μL of 50% 2-ME.
4. The samples are separated by SDS-PAGE (Fig. 2a). The gel is stained by CBB. The PEG-modified band appears at positions with higher molecular weight than the unmodified band.
5. The modified and unmodified bands are analyzed to quantify the modification ratio by densitometry using image analysis

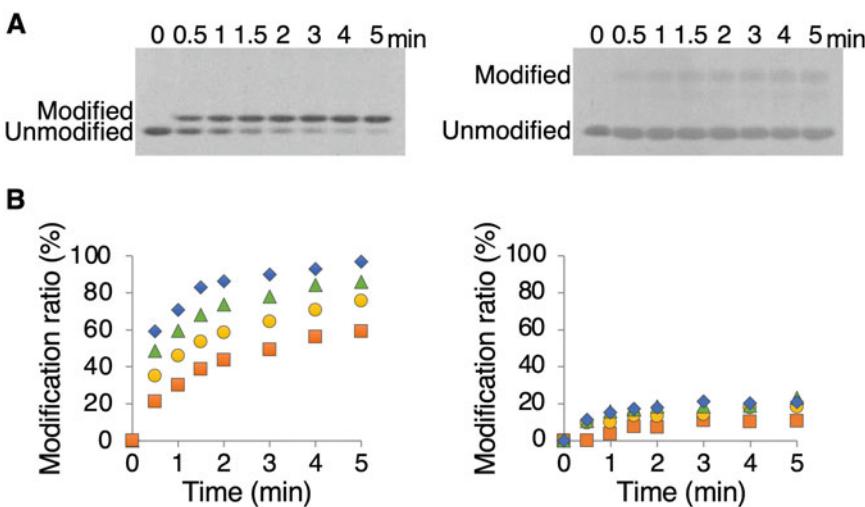


Fig. 2 (a) Examples of time-dependent pattern changes of S50C reacted with 1 mM MM(PEG)_n with molecular weights of 700 (left) and 5000 (right). **(b)** Plots of modification ratios of S50C mutants reacted with MM(PEG)_n with molecular weights of 700 (left) and 5000 (right). Blue diamonds, green triangles, yellow circles, and orange squares represent MM(PEG)_n concentrations of 2, 1, 0.5, and 0.2 mM, respectively

software such as ImageJ (Fig. 2b) (*see Note 6*). Based on the modification ratio, the optimized reaction conditions are determined (*see Note 7*).

3.3 Estimation of the Actual Pore Size of Protein Nanocages in Solution

The reactions are performed similarly to the optimization processes.

1. The concentration of S50C is adjusted to 0.22 μM with a volume of 27 μL (pH 7.4).
2. MM(PEG)_ns with molecular weights 700, 1200, 2000, 5000, 10,000, 20,000 are used for modifications. The 3 μL of 10 mM solution is added to the TIP60 solution. The final concentration of MM(PEG)_ns is adjusted to 1 mM.
3. After incubation of the samples for 5 min at 30 °C, the reaction is terminated by the addition of 3 μL of 50% 2-ME.
4. The samples are separated by SDS-PAGE. The gel is stained by CBB. The PEG-modified band appears at positions with higher molecular weight than the unmodified band (Fig. 3a) (*see Note 8*).
5. The modified and unmodified bands are analyzed to quantify the modification ratio by densitometry using image analysis software such as ImageJ (Table 1).
6. The modification ratio is then plotted against the estimated Stokes diameter of MM(PEG)_ns (Fig. 3b) (*see Note 9*).

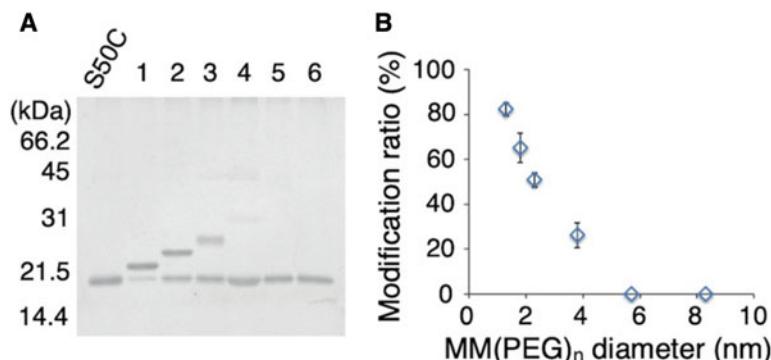


Fig. 3 (a) SDS-PAGE results of the modification of S50C. The left-most lane (“S50C” at the top) is the sample without MM(PEG)_n. Lanes 1–6 are S50C samples modified with MM(PEG)_n with molecular weights of 700, 1200, 2000, 5000, 10,000, and 20,000 Da, respectively. The numbers to the left of the gel show the molecular weights of the protein markers. (b) Relation between the MM(PEG)_n Stokes diameter and modification ratio of S50C calculated from the SDS-PAGE results. (Reprinted from Ref. [23] under Creative Commons license 4.0 CC BY-NC-ND)

3.4 Dual-Surface Modifications

- The double mutant (S20C/S50C) protein is adjusted its concentration to 1 μM (pH 7.4).
- According to the optimization experiment described, the double mutant is initially reacting with 1 mM MM(PEG)_ns larger than the pore size. In the case of TIP60, MM(PEG)_ns with molecular weight higher than 10,000 is used for exterior surface modification. The sample is incubated for 5 min at 30 °C.
- Incubated samples are ultrafiltrated to remove excess MM(PEG)_n and 2-ME and to exchange buffer to dialysis buffer (pH 8.0) (*see Note 10*). Protein concentrations of samples after ultrafiltration are confirmed by absorption at 280 nm. The solution is used as a stock protein solution.
- An aliquot of the stock protein solution is diluted to 0.2 μM then subjected to SDS-PAGE to confirm the modification ratio.
- Remained stock protein solution is adjusted to its concentration to 0.6 μM and reacted with 360 μM DTNB for 20 min at 30 °C. The reaction is monitored by the UV-Vis absorption spectrum. The peak 412 nm attributable to produced free TNB is quantified based on the molar extinction coefficient $14.15 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ [25].
- The excess DTNB and produced free TNB molecules are removed by ultrafiltration.
- The bound TNB to interior Cys is confirmed by absorption maxima at 323 nm.

3.5 Release of Bound TNB to Cys Residue on the Interior Surface of TIP60

- The dual-surface modified TIP60 is transferred to a quartz cell.
- Modified TIP60 is then reacted with 1 mM TCEP that cleaves the disulfide bonds between interior Cys residues and TNB.
- The absorption of 412 nm attributable to free TNB is monitored for 40 min.
- The release rate is determined based on the release rate curve (*see Note 11*).

4 Notes

- The protein solution is dialyzed against the dialysis buffer pH 7.4 and 8.0 before the modification experiments by MM(PEG)_n and DTNB, respectively.
- TCEP is used for reducing inner- or intermolecular disulfide bonds. Removal of TCEP again initiates disulfide bond formation. Thus, the mutant proteins must be stored before the dialysis step. Although TCEP does not contain thiol groups,

it decelerates the conjugation reaction of cysteine thiols by the maleimide group [26]. Therefore, TCEP is removed immediately before modification reaction.

3. The freeze-thaw cycle induces denaturation of TIP60. The 60-mer structure is kept for at least one month if the protein is stored at 4 °C.
4. All the mutants can be purified by the same methods.
5. DMF is used to dissolve MM(PEG)_ns in this experiment. If the proteins are denatured by DMF, DMSO is the possible alternative.
6. A total of 1 µg protein is applied for SDS-PAGE analysis. The modification ratios are calculated from the following equation.

$$\text{Modification ratios (\%)} = \frac{I_m}{I_n + I_m}$$

where I_n and I_m are the intensity of the non-modified band and modified band, respectively. In the case of MM(PEG)_n with molecular weights of 2000 and 5000, the intensities of multiple minor bands at ~45 kDa are included in the intensity of modified bands.

7. The optimized reaction conditions are determined based on the modification ratio of MM(PEG)_ns with molecular weight 5000. If the MM(PEG)_ns access to the interior space freely, the modification ratio is concentration-dependent in addition to the incubation time. The MM(PEG)_ns with molecular weight 5000 gives the imperfect modification ratio, 20% in maxima, at 5 min incubation with 1 mM MM(PEG)_ns. Interestingly, the modification ratio is not increased at the longer incubation time with higher concentrations. Thus, we decided the reaction conditions would be appropriate for estimating the pore sizes of TIP60 in the solution.
8. In the case of MM(PEG)_n with a molecular weight larger than 10,000, excess MM(PEG)_n should be removed by ultrafiltration before SDS-PAGE analysis because they disturb the electrophoresis band patterns.
9. Stokes radii of PEGs (r in cm) are calculated from their molecular weight (M) from the following equation [27]

$$r = 16.73 \times 10^{-10} M^{0.557}$$

10. The 2-ME interrupting subsequent modification reactions should be thoroughly removed by dialysis buffer (pH 8.0).
11. If the release rate from interior space is essential, a control experiment that modification of exterior cysteine by DTNB should be performed simultaneously.

Acknowledgment

This work was supported by JSPS KAKENHI Grants JP18K05324 and JP19H02522.

References

1. Spicer CD, Davis BG (2014) Selective chemical protein modification. *Nat Commun* 5:4740
2. Uchida M, Kosuge H, Terashima M et al (2011) Protein cage nanoparticles bearing the LyP-1 peptide for enhanced imaging of macrophage-rich vascular lesions. *ACS Nano* 5:2493–2502
3. Ren D, Kratz F, Wang S-W (2011) Protein nanocapsules containing doxorubicin as a pH-responsive delivery system. *Small* 7:1051–1060
4. Du B, Jia S, Wang Q et al (2018) A self-targeting, dual ROS/pH-responsive apoferritin nanocage for spatiotemporally controlled drug delivery to breast cancer. *Biomacromolecules* 19:1026–1036
5. Anderson EA, Isaacman S, Peabody DS et al (2006) Viral nanoparticles donning a paramagnetic coat: conjugation of MRI contrast agents to the MS2 capsid. *Nano Lett* 6: 1160–1164
6. Kawano T, Murata M, Kang JH et al (2018) Ultrasensitive MRI detection of spontaneous pancreatic tumors with nanocage-based targeted contrast agent. *Biomaterials* 152:37–46
7. Meier-Menches SM, Casini A (2020) Design strategies and medicinal applications of metal-peptidic bioconjugates. *Bioconjug Chem* 31: 1279–1288
8. Matsumoto NM, Prabhakaran P, Rome LH, Maynard HD (2013) Smart vaults: thermally-responsive protein nanocapsules. *ACS Nano* 7: 867–874
9. Cobo I, Li M, Sumerlin BS, Perrier S (2015) Smart hybrid materials by conjugation of responsive polymers to biomacromolecules. *Nat Mater* 14:143–149
10. Chen C, Ng DYW, Weil T (2020) Polymer bioconjugates: modern design concepts toward precision hybrid materials. *Prog Polym Sci* 105:101241
11. Pires IS, Savla C, Palmer AF (2020) Poly(ethylene glycol) surface-conjugated apohemoglobin as a synthetic heme scavenger. *Biomacromolecules* 21:2155–2164
12. Ding S, Zhang N, Lyu Z et al (2021) Protein-based nanomaterials and nanosystems for biomedical applications: a review. *Mater Today* 43:166–184
13. Kim Y, Ho SO, Gassman NR et al (2008) Efficient site-specific labeling of proteins via cysteines. *Bioconjug Chem* 19:786–791
14. Nair DP, Podgórski M, Chatani S et al (2014) The Thiol-Michael addition click reaction: a powerful and widely used tool in materials chemistry. *Chem Mater* 26:724–744
15. Zhang Y, Ardejani MS, Orner BP (2016) Design and applications of protein-cage-based nanomaterials. *Chem Asian J* 11:2814–2828
16. Chen H, Tan X, Fu Y et al (2021) The development of natural and designed protein nanocages for encapsulation and delivery of active compounds. *Food Hydrocoll* 121:107004
17. Lee EJ, Lee NK, Kim IS (2016) Bioengineered protein-based nanocage for drug delivery. *Adv Drug Deliv Rev* 106:157–171
18. Ahn B, Lee SG, Yoon HR et al (2018) Four-fold channel-nicked human ferritin nanocages for active drug loading and pH-responsive drug release. *Angew Chem Int Ed* 57:2909–2913
19. Stephanopoulos N, Tong GJ, Hsiao SC, Francis MB (2010) Dual-surface modified virus capsids for targeted delivery of photodynamic agents to cancer cells. *ACS Nano* 4:6014–6020
20. Ashley CE, Carnes EC, Phillips GK et al (2011) Cell-specific delivery of diverse cargos by bacteriophage MS2 virus-like particles. *ACS Nano* 5:5729–5745
21. Niikura K, Sugimura N, Musashi Y et al (2013) Virus-like particles with removable cyclodextrins enable glutathione-triggered drug release in cells. *Mol Biosyst* 9:501–507
22. Kawakami N, Kondo H, Matsuzawa Y et al (2018) Design of hollow protein nanoparticles with modifiable interior and exterior surfaces. *Angew Chem Int Ed* 57:12400–12404
23. Nasu E, Kawakami N, Miyamoto K (2021) Nanopore-controlled dual-surface modifications on artificial protein nanocages as nano-carriers. *ACS Appl Nano Mater* 4:2434–2439
24. Ellman GL (1959) Tissue sulphydryl groups. *Arch Biochem Biophys* 82:70–77

25. Eyer P, Worek F, Kiderlen D et al (2003) Molar absorption coefficients for the reduced Ellman reagent: reassessment. *Anal Biochem* 312:224–227
26. Tyagarajan K, Pretzer E, Wiktorowicz JE (2003) Thiol-reactive dyes for fluorescence labeling of proteomic samples. *Electrophoresis* 24:2348–2358
27. Singh S, Khulbe KC, Matsuura T, Ramamurthy P (1998) Membrane characterization by solute transport and atomic force microscopy. *J Memb Sci* 142:111–127



Chapter 10

Modification and Production of Encapsulin

Sandra Michel-Souzy and Jeroen J. L. M. Cornelissen

Abstract

Encapsulins are a class of protein nanocages that are found in bacteria, which are easy to produce and engineer in *E. coli* expression systems. The encapsulin from *Thermotoga maritima* (Tm) is well studied, its structure is available, and without modification it is barely taken up by cells, making it promising candidates for targeted drug delivery. In recent years, encapsulins are engineered and studied for potential use as drug delivery carriers, imaging agents, and as nanoreactors. Consequently, it is important to be able to modify the surface of these encapsulins, for example, by inserting a peptide sequence for targeting or other functions. Ideally, this is combined with high production yields and straightforward purification methods. In this chapter, we describe a method to genetically modify the surface of Tm and *Brevibacterium linens* (Bl) encapsulins, as model systems, to purify them and characterize the obtain nanocages.

Key words Nanocages, Encapsulin, Targeting, Drug encapsulation, Genetic modification

1 Introduction

Encapsulins are protein cages found in some bacteria and archaea [1], with possible viral origin [2–4]. Nanocages and especially protein nanocages are used for different applications such as for vaccines, drug delivery, imaging agents, and nanoreactors. Encapsulins are divided in different classes with the most studied examples present in the family containing redox-active enzymes [5]. Because of their bacterial origin, encapsulins are easy to produce in *E. coli* expression systems and, therefore, are also easy to be engineered. Encapsulins are found in different sizes as the constituent proteins self-assemble in the cell in a different symmetry depending of the specie [5, 6]. This chapter focuses on the encapsulin from *Thermotoga maritima* (Tmenc) and the one from *Brevibacterium linens* (Blenc) which both form a 24 nm cage of 60 monomers assembled in a $T = 1$ symmetry [4, 7]. Different parts of the encapsulin can be modified, depending on the application, such as the modifications of the pores in the protein shell, in order to work as a bionanoreactor [8, 9], or on the surface to

produce vaccines, imaging agents, or for targeting drug delivery [10–14]. These modifications can be either chemical or genetic, with the latter one having the advantage that a peptide sequence can be inserted at a controlled position and in a known amount; that is, a single insertion in the monomer will lead to 60 functional groups for a $T=1$ symmetric encapsulin (120 for a $T=2$, 180 for a $T=3\dots$). Moreover, for targeting drug delivery purposes, Tmenc is an asset because the introduction of a targeting peptide is needed to be taken up by cells [10, 15]. Finally, encapsulins are good candidates for developing carriers as they can encapsulate polypeptides by fusing it with a short C-terminal sequence that is present on the physiological encapsulin cargo [4, 16].

The modifications of different locations on the encapsulin are important for multiple applications, but can lead of a destabilization of the cage structure; thereupon, it is crucial for each new modifications to check the formation and stability of the formed cage. Consequently, the library of encapsulin variants is growing quickly. For Tmenc, the majority of the studies cover genetic surface functionalization for vaccine, targeting, or enhanced peptide expression [11, 15, 17–20], and there are also studies about the modifications of the pores to enable transport across the cage shell [8, 9].

The monomer of Tmenc consists of three domains; P, E, and A. The reported modifications for vaccine development or cell targeting are located on the N-terminus and in loop positions 42–43 (which are in the P-domain); 127–128, 138–139, and the C-terminus (A-domain); and in loop positions 57–58, 60–61, 64–65, 71–72 (E-domain). Among these nine modifications, six lead to a stable or soluble cage, while the reported studies show that modifications after residues 57, 60, and 71 are not the best options [11, 15, 17–20]. Note that modifications in C-terminus need a linker to be accessible and the cage is less stable over long-time storage [15, 17]. Concerning the pore modification, no additional peptide sequence was introduced, but amino acids were substituted or deleted to observe the effect on the dynamics of the pore. These studies focused on the loop between residues 184 and 193, and a total of 35 variants were constructed, resulting in 21 stable variants [8, 9].

In this chapter, we describe a methodology to genetically modify an encapsulin from Tm and how to purify and characterize it.

2 Materials

2.1 Cloning

1. Primers.
2. Polymerase for cloning (e.g., Q5 polymerase) and its buffers.
3. dNTPs.

4. Plasmid containing encapsulin gene.
5. Hosting vectors. Preferably, use the pDUET vectors which possess dual multiple cloning sites (MCS), and there are four which are compatible and allow a certain modularity.
6. Agarose.
7. Electrophoresis system.
8. Gel Red/Sybr safe.
9. UV lamp/Imager.
10. Restriction enzymes (*see Note 1*).
11. DpnI enzyme (New England Biolabs).
12. PCR cleaning kit.
13. T4 DNA ligase.
14. Competent *E. coli* strain for cloning such as NovaBlue or TG1.
15. Luria Broth (LB) media.
16. Agar.
17. Antibiotics correspondent to the plasmid resistance.
18. Miniprep kit.
19. Polymerase for PCR colony.

2.2 Production

1. Competent *E. coli* strain for production such as Rosetta or BL21.
2. LB agar.
3. LB.
4. Antibiotics correspond to the plasmid resistance.
5. Isopropyl β -D-1-thiogalactopyranoside (IPTG).
6. Centrifuge.

2.3 Purification Steps

Affinity Chromatography

1. Ultracentrifuge.
2. Sonicator or French Press.
3. Fast Protein Liquid Chromatography (FPLC) system.
4. Column: HisTrapTM HP from Cytiva for an His-tagged encapsulin and StrepTrapTM HP for a Strep-tagged encapsulin can be used.
5. Lysis buffer: 50 mM Hepes pH8, 150 mM NaCl, 15 mM Beta-mercaptoethanol (β me), 1 mM EDTA, 20 mM MgCl₂, 1 protease inhibitor tablet/7 mL (cCompleteTM), 0.5 mg/mL Lysosyme, 20 μ g/mL DNase, 30 μ g/mL RNase. If it is a purification on nickel matrix, add 30 mM Imidazole.
6. Washing buffer: 50 mM Hepes pH8, 150 mM NaCl, 15 mM β me. For purification on a nickel matrix add 30 mM Imidazole.

7. Elution buffer: 50 mM Hepes pH8, 150 mM NaCl, 15 mM β me + 500 mM imidazole for Nickel column or 2.5 mM desthiobiotin for StrepTactin Column.

Sucrose Gradient

1. Encapsulin buffer: 20 mM Tris-HCl pH7.5, 150 mM NH₄Cl, 20 mM MgCl₂, 1 mM β me.
2. Encapsulin lysis buffer: 20 mM Tris-HCl pH7.5, 150 mM NH₄Cl, 20 mM MgCl₂, 1 mM β me, 20 μ g/mL DNase, 30 μ g/mL RNase.
3. Storage enc buffer: 20 mM Tris-HCl pH7.5, 150 mM NH₄Cl, 1 mM β me.
4. Sucrose: 38% (w/v), 10% and 50%.
5. Peristaltic pump.
6. Swinging rotor (e.g., Thermo Scientific SureSpin 630 (36 mL) Rotor).

Size Exclusion Chromatography

1. Amicon® Ultra filter unit, 100 kDa.
2. Fast Protein Liquid Chromatography (FPLC) system.
3. Column: Superose® 6 Increase 10/300 GL can be used.
4. Storage buffer: 50 mM Hepes pH8, 150 mM NaCl, 15 mM β me.

2.4 Characterization

1. Nanotrac Wave (Microtrac) particle analyzer.
2. Transmission electron microscope.
3. Uranyl acetate (1% w/v).
4. Formvar carbon-coated copper grid (Electron Microscopy Sciences).

3 Methods

3.1 Cloning Steps

1. Design the primers needed to add the modification at the desired position. Be sure to have at least 10 bases overlapping with the gene and with the vector.
2. Prepare the hosting vector with the good restriction enzyme. For example, the couple NdeI-EcoRV can be used to clone in pDUET MCS2. Cut 1 μ g of vector in 100 μ L of final volume with 1 μ L of each enzyme.
3. Purify the cut vector with the PCR cleaning kit.
4. To incorporate a peptide coding sequence inside the encapsulin gene, proceed in two PCR steps (Fig. 1). Use as a matrix the native DNA sequence of encapsulin gene and do PCR 1 with

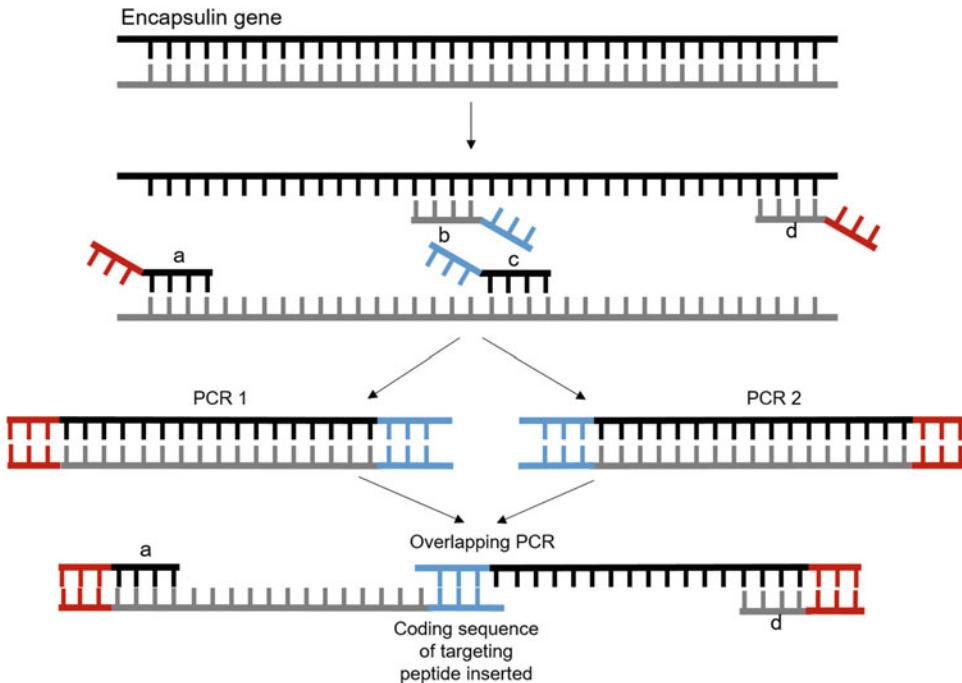


Fig. 1 Steps to perform an overlapping PCR to insert a sequence in a middle of a gene (see text)

primer **a** (with the complementary sequence to the vector) and primer **b** (possessing the DNA sequence of the peptide to be inserted). Do PCR 2 with primer **c** (possessing the sequence of the peptide and complementary to the primer **b**) and primer **d** (with the complementary sequence to the vector).

5. Check the success of the PCR and the size of the products obtained on agarose gel 1%. If it is good, digest the product with 1 μ L of DpnI to remove the remaining matrix. Then purify the PCR product with the PCR cleaning kit (*see Note 2*).
6. Use the PCR1 and PCR2 as a matrix and perform the overlap PCR with primer **a** and **d**.
7. Check the product of the PCR on agarose gel.
8. If it is good, purify the PCR product with the PCR cleaning kit.
9. Perform the SLIC method [21] to ligate the vector with your modified encapsulin gene. Mix a ratio of 1 vector for 4 inserts with a fixed amount of 100 ng of vector in the tube. When you add the T4 DNA polymerase, the reaction immediately starts and it is really important to respect the 2'30" of reaction time (*see Note 3*).
10. Put in ice immediately for 10 min.

11. Transform the cloning *E. coli* strain with the SLIC product. Put in contact for 30 min. in ice, give a thermal shock for 40 s at 42 °C, add 400 µL of LB, and express for 1 h at 37 °C.
12. Spread the bacteria on an agar plate containing the antibiotic corresponding to the used vector. Incubate overnight at 37 °C.
13. Patch between 20 and 30 colonies on agar plates with antibiotic. Incubate overnight at 37 °C.
14. Perform a colony PCR to discriminate between bacteria containing plasmid with and without an insert.
15. Select the good colonies and inoculate two positive colonies in separate culture tubes in 4 mL LB-antibiotics. Incubate overnight at 37 °C.
16. Perform the extraction of the plasmid with a miniprep kit, following the kit instructions and send 30 µL of sample to sequencing.

Until here the modification of the encapsulin monomer gene was described, to obtain the desired shell. Steps 17 and 18 describe the (potential) modification of the cargo which is desired to be encapsulated.

17. Take the sequence of the desired cargo and design the primer to insert the C-terminal extension allowing the cargo encapsulation [4].
18. Repeat all the steps used for the modification of the encapsulin gene.

3.2 Production Steps

1. Get the plasmid(s) in the good cells for production. Transformation of Rosetta or BL21 competent cells for example, with the plasmids of interest (*see Note 4*).
2. Prepare an overnight (ON) culture of the desired strain. Take 3–5 colonies from the agar plate or scratch a bit of the glycerolate at –80 °C. Prepare 30 mL of LB/antibiotics per liter of culture.
3. Incubate ON at 37 °C 200 rpm.
4. Measure the optic density at 600 nm (OD₆₀₀).
5. Inoculate the desired volume of culture (usually 2–4 L of culture to obtain 20 mg of proteins) at 0.1 uOD₆₀₀.
6. Incubate at 37 °C 200 rpm.
7. Monitor the growing; when the OD₆₀₀ reaches between 0.4 and 0.6 (exponential phase) add isopropyl β-D-1-thiogalactopyranoside (IPTG) 0.1 to 1 mM to induce the genes expression.
8. Reduce the temperature to 25 °C and incubate ON (*see Note 5*).

9. Collect the cells by centrifugation at 4000 g for 10 min. Discard the supernatant (*see Note 6*).

3.3 Purification Steps

1. Resuspend the pellet in lysis buffer. Use 10 mL per liter of culture.
2. Sonicate 2 × 1 min in ice with 4 s pulse 1 s pause, amplitude 80% (*see Note 7*).
3. Centrifuge 10 min at 4 °C at 4000 g to remove the bigger debris.
4. Collect the supernatant. Discard the pellet.
5. Ultracentrifuge 30 min at 4 °C at 100,000 g to remove the membrane and small debris.
6. Collect the supernatant (*see Note 8*).

Purification by Affinity via a Tag

7. Use appropriate affinity column to separate the lysate depending on the Tag present on encapsulin. Different methods are possible, using either Fast Protein Liquid Chromatography (FPLC) system or a gravity column (*see Note 9*). Only the methods using FPLC will be described.
8. Wash the column with 2 column volumes of water.
9. Equilibrate the column with 3 column volumes of washing buffer.
10. Load the sample on the column and keep 30 µL to analyze it by SDS-PAGE.
11. Follow the absorbance at $\lambda = 280$ nm and in case of a fluorescent cargo the absorbance according to the excitation of the cargo.
12. Collect the flow through for analyses.
13. Wash with washing buffer and wait until the absorbance stabilizes and is close to 0 (*see Note 10*). Take a sample of the wash at this point.
14. Elute with elution buffer and collect fractions of 0.5 to 2 mL (*see Note 11*).
15. Take a sample of 30 µL of six elution fraction which correspond to the peak and its borders.
16. Make an SDS-PAGE to check the purification and the purity of the samples (Fig. 2).
17. Mix together all the elution fractions containing your proteins (*see Note 12*).

For a higher degree of purity, but also to separate the cages from the remaining monomers, it is highly recommended to perform a size

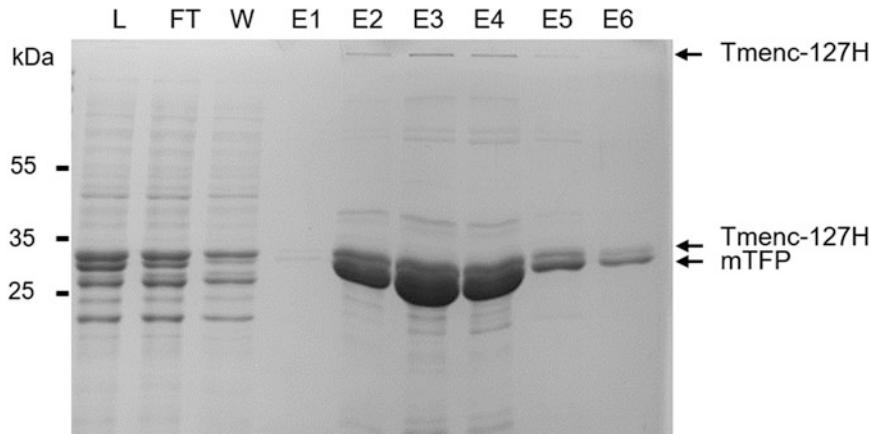


Fig. 2 SDS-PAGE profile example of encapsulin affinity purification. Loading 10 µL of sample. SDS-PAGE 15% acrylamide

exclusion chromatography. This size exclusion step will be the first step in the characterization of the cage (see below).

18. In order to prepare the sample and reduce its volume to allow the loading on the size exclusion column, concentrate the sample using Amicon® Ultra filter unit, 100 kDa. Centrifuge at 1000 g by rounds of 2 min until you reach 1 or 2 mL (see Note 13).
19. Collect the sample and load it on an equilibrated size exclusion column to separate the proteins by size. A Superose® 6 Increase 10/300 GL is good and enough for the separation of the encapsulins. On this column, the cages will elute between 12 and 13 mL. Follow the absorbance at $\lambda = 280$ nm and collect the peak containing the nanocages (Fig. 3).
20. Facultative step: the purity can be check by SDS-PAGE.

By Sucrose Gradient Without Tag

All steps are the same until the Subheading 3.3, step 6.

7. Place 10 mL of sucrose 38% in an ultracentrifuge tube.
8. Add 10 mL of the supernatant gently on the top (see Note 14).
9. Ultracentrifuge 17 h at 10 °C at 100,000 g for Blenc and at 165,000 g for Tmenc.
10. For Tmenc, remove the supernatant and resuspend the pellet in 1 mL storage enc buffer. The sample is ready to be loaded on the size exclusion column (see Subheading 3.3, step 19) in storage enc buffer.

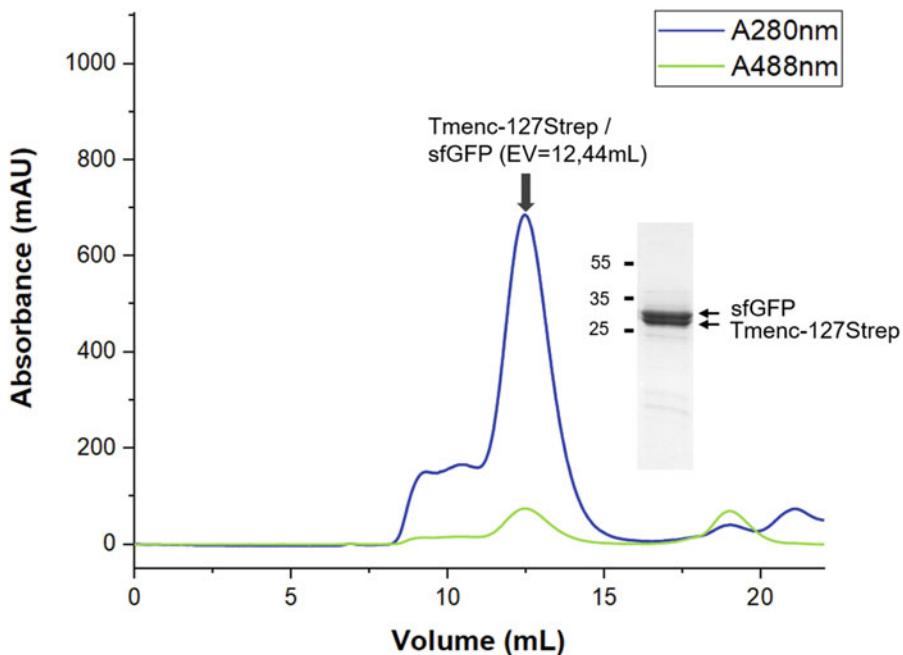


Fig. 3 Example profile and SDS-PAGE analysis of a SEC of Tmenc127strep/sfGFP [15]

11. For Blenc collect 3 mL from the bottom of the tube (*see Note 15*).
12. Concentrate the sample using an Amicon® Ultra filter unit. Centrifuge at 1000 g by rounds of 2 min. until you reach 1 mL (*see Note 13*).
13. Prepare 34 mL of a 10–50% sucrose gradient in a tube fitting with a swinging rotor (*see Note 16*).
14. Add the 1 mL sample from step 12 carefully to the top.
15. Centrifuge 17.5 h at 4 °C at 145,000 g.
16. Remove the first 20 mL and collect the fraction from 21 to 29 mL.
17. Put this 9 mL in a dialysis bag and dialyze at 4 °C against 1 L of storage enc buffer for minimum 6 h.
18. Concentrate the sample using an Amicon® Ultra filter unit. Centrifuge at 1000 g by rounds of 2 min. until you reach 1 mL (*see Note 13*).
19. The sample is ready to be loaded on the size exclusion column (*see Subheading 3.3, step 19*) in storage enc buffer.

3.4 Characterization

1. The size exclusion is already a good experiment to determine the size of the encapsulins (*see* Subheading 3.3, step 19). Every column is slightly different and can have a little variation on the elution volume. It is therefore recommended to establish a calibration curve to calculate the estimated size.
2. Another method to determine if the structure is still intact is Dynamic Light Scattering (DLS). It has the advantage to not denature or dilute the sample. Perform 5 runs of 120 s on the Nanotrac Wave (Microtrac) particle analyzer. Intensity results will give the size and allow to check the degree of aggregation present in the sample. Number transformation will allow to see what is the distribution of the particles.
3. Transmission electron microscopy (TEM) analyses will allow visualization of the cages and determine their size and shape. Apply 5 µL sample to a Formvar carbon-coated copper grid. Incubate on the grid for 2 min, then remove any excess buffer with filter paper. Negatively stain the sample by applying 5 µL uranyl acetate (1% w/v) onto the grid and incubating for 40 s. Remove the excess with filter paper, leave to dry for 10 min before imaging.

4 Remarks and Conclusion

This chapter presents an efficient method to modify encapsulins by tools from molecular biology. It also describes how to purify tagged and untagged encapsulins; however, as shown in Fig. 4, the purification by affinity column leads to an improved purity, which is essential for biomedical application. Therefore, if it is possible, it is advised to add a purification tag to encapsulin. For Tmenc, it has been shown that two positions can be modified at the same time, the position 127 and the C-terminal of the monomer [15].

5 Notes

1. NEB High fidelity restriction enzymes have the advantage to cut all in the same buffer.
2. If the PCR does not succeed, play with the hybridization temperature *and* double check if the sequence used for designing the primers matches with the sequence of the gene (genomic sequence or optimized sequence).
3. With the SLIC method, it is possible to directly ligate PCR1 and PCR2, without doing the overlapping PCR; however, this is a little more difficult, but definitely faster.

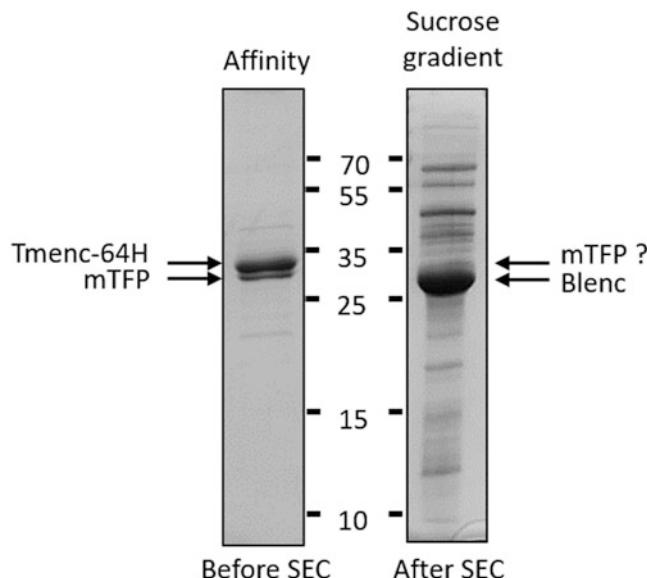


Fig. 4 SDS-PAGE data showing the difference of purity after affinity purification (left) and after sucrose gradient purification (right)

4. Possibility of a double transformation to have the encapsulin and the cargo. Don't forget to prepare a plate with both antibiotics. Note that the numbers of colonies will be reduced on agar plate. The correct antibiotic concentrations to use are as follows: Ampicillin 50 µg/mL, Kanamycin 30 µg/mL, and Streptomycin 30 µg/mL.
5. Reducing the temperature allows the reduction of protein degradation activities.
6. At this point, a break is possible. Put the pellet at -20 °C and use it within a month.
7. Pay attention to the noise made by the sonicator. It should stay the same during the whole process and no foam should appear. The probe should be as deep as possible, but without touching tube walls. If foam is appearing, stop the sonication, replace the probe, and restart.
8. At this stage, the pellet is "loosy." Be careful when collecting the supernatant to not take the pellet, better to let some supernatant above it.
9. Columns which can be used for FPLC are HisTrap™ HP from Cytiva for an His-tagged encapsulin and StrepTrap™ HP for a Strep-tagged encapsulin. For a gravity column, Econo-Pac® Chromatography Columns from Biopac can be used associated with Ni-NTA agarose beads or Strep-Tactin sepharose beads.
10. In general, about 4–5 column volumes are needed to wash a HisTrap™ column and 3 to wash a StrepTrap™.

11. The smaller the fractions are the easier it will be to select and collect just the fractions corresponding to your protein peak. However, the more fractions you have the more you have to analyze on SDS-PAGE. A 1 mL fraction is a workable volume.
12. If a gravity column is used, the SDS-PAGE is essential to determine in which fractions the proteins of interest are.
13. The concentration step is critical; there is a high risk of precipitation, which means the loss of protein. Check every 2 min to be sure that there is no precipitation by pipetting the sample up and down. If it starts to become a little turbid, stop the concentration process IMMEDIATELY and continue with the protocol. In case of concentrating a sample with imidazole, dilute the sample 2 times before starting the concentration protocol. The imidazole increases the chance of precipitation.
14. Be careful, 2 layers should be visible the sample being lighter than the sucrose.
15. Fill another ultracentrifuge tube with 3 mL of water. Carefully remove all the liquid in the Blenc ultracentrifuge tube from the top, until the same amount is left as in the reference tube (3 mL).
16. If there is not a peristaltic pump, the gradient can be done by hand. Prepare 9 solutions (10, 15, 20, 25, 30, 35, 40, 45, 50% sucrose) and start by adding 3.78 mL to the tube and continue to add gently the decreasing concentrations of sucrose solution.

References

1. Jones JA, Giessen TW (2020) Advances in encapsulin nanocompartment biology and engineering. *Biotechnol Bioeng* 118:491–505
2. McHugh CA, Fontana J, Nemecek D, Cheng N, Aksyuk AA, Heymann JB, Winkler DC, Lam AS, Wall JS, Steven AC, Hoiczyk E (2014) A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J* 33:1–16
3. Akita F, Chong KT, Tanaka H, Yamashita E, Miyazaki N, Nakaishi Y, Suzuki M, Namba K, Ono Y, Tsukihara T, Nakagawa A (2007) The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. *J Mol Biol* 368:1469–1483
4. Sutter M, Boehringer D, Gutmann S, Günther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol* 15: 939–947
5. Wiryaman T, Toor N (2022) Recent advances in the structural biology of encapsulin bacterial nanocompartments. *J. Struct. Biol.* X 6: 100062
6. Giessen TW (2016) Encapsulins: microbial nanocompartments with applications in biomedicine, nanobiotechnology and materials science. *Curr Opin Chem Biol* 34:1–10
7. Putri RM, Allende-Ballesteros C, Luque D, Klem R, Rousou KA, Liu A, Traulsen CHH, Rurup WF, Koay MST, Castón JR, Cornelissen JJLM (2017) Structural characterization of native and modified encapsulins as nanoplatforms for in vitro catalysis and cellular uptake. *ACS Nano* 11:12796–12804
8. Williams EM, Jung SM, Coffman JL, Lutz S (2018) Pore engineering for enhanced mass transport in encapsulin nanocompartments. *ACS Synth Biol* 7:2514
9. Adamson LSR, Tasneem N, Andreas MP, Close W, Jenner EN, Szyszka TN, Young R, Cheah LC, Norman A, MacDermott-Opeskin

- HI, O'Mara ML, Sainsbury F, Giessen TW, Lau YH (2022) Pore structure controls stability and molecular flux in engineered protein cages. *Sci Adv* 8:1–13
10. Moon H, Lee J, Kim H, Heo S, Min J, Kang S (2014) Genetically engineering encapsulin protein cage nanoparticle as a SCC-7 cell targeting optical nanoprobe. *Biomater Res* 18:1–7
11. Lagoutte P, Mignon C, Stadthagen G, Potisopon S, Donnat S, Mast J, Lugari A, Werle B (2018) Simultaneous surface display and cargo loading of encapsulin nanocompartments and their use for rational vaccine design. *Vaccine* 36:3622–3628
12. Van de Steen A, Khalife R, Colant N, Mustafa Khan H, Deveikis M, Charalambous S, Robinson CM, Dabas R, Esteban Serna S, Catana DA, Pildish K, Kalinovskiy V, Gustafsson K, Frank S (2021) Bioengineering bacterial encapsulin nanocompartments as targeted drug delivery system. *Synth Syst Biotechnol* 6: 231–241
13. Edwardson TGW, Levasseur MD, Tetter S, Steinauer A, Hori M, Hilvert D (2022) Protein cages: from fundamentals to advanced applications. *Chem Rev.* <https://doi.org/10.1021/acs.chemrev.1c00877>
14. Gabashvili AN, Vodopyanov SS, Chmelyuk NS, Sarkisova VA, Fedotov KA, Efremova MV, Abakumov MA (2021) Encapsulin based self-assembling iron-containing protein nanoparticles for stem cells MRI visualization. *Int J Mol Sci* 22:12275
15. Michel-Souzy S, Hamelmann NM, Zarzuelo-Pura S, Paulusse JMJ, Cornelissen JJLM (2021) Introduction of surface loops as a tool for encapsulin functionalization. *Biomacromolecules* 22:5234–5242
16. Rurup WF, Snijder J, Koay MST, Heck AJR, Cornelissen JJLM (2014) Self-sorting of foreign proteins in a bacterial nanocompartment. *J Am Chem Soc* 136:3828–3832
17. Moon H, Lee J, Min J, Kang S (2014) Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplateform. *Biomacromolecules* 15:3794–3801
18. Choi B, Moon H, Hong SJ, Shin C, Do Y, Ryu S, Kang S (2016) effective delivery of antigen-encapsulin nanoparticle fusions to dendritic cells leads to antigen-specific cytotoxic T cell activation and tumor rejection. *ACS Nano* 10:7339–7350
19. Bae Y, Kim GJ, Kim H, Park SG, Jung HS, Kang S (2018) Engineering tunable dual functional protein cage nanoparticles using bacterial superglue. *Biomacromolecules* 19:2896–2904
20. Lee T, Carpenter TS, D'haeseleer P, Savage DF, Yung MC (2019) Encapsulin carrier proteins for enhanced expression of antimicrobial peptides. *Biotechnol Bioeng* 117:603–613
21. Jeong JY, Yim HS, Ryu JY, Lee HS, Lee JH, Seen DS, Kang SG (2012) One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics Studies. *Appl Environ Microbiol* 78:5440–5443

Part III

Characterization and Simulation of Protein Cage Systems



Chapter 11

Cryo-electron Microscopy of Protein Cages

Raymond N. Burton-Smith and Kazuyoshi Murata

Abstract

Protein cages are one of the most widely studied objects in the field of cryogenic electron microscopy—encompassing natural and synthetic constructs, from enzymes assisting protein folding such as chaperonin to virus capsids. Tremendous diversity of morphology and function is demonstrated by the structure and role of proteins, some of which are nearly ubiquitous, while others are present in few organisms. Protein cages are often highly symmetrical, which helps improve the resolution obtained by cryo-electron microscopy (cryo-EM). Cryo-EM is the study of vitrified samples using an electron probe to image the subject. A sample is rapidly frozen in a thin layer on a porous grid, attempting to keep the sample as close to a native state as possible. This grid is kept at cryogenic temperatures throughout imaging in an electron microscope. Once image acquisition is complete, a variety of software packages may be employed to carry out analysis and reconstruction of three-dimensional structures from the two-dimensional micrograph images. Cryo-EM can be used on samples that are too large or too heterogeneous to be amenable to other structural biology techniques like NMR or X-ray crystallography. In recent years, advances in both hardware and software have provided significant improvements to the results obtained using cryo-EM, recently demonstrating true atomic resolution from vitrified aqueous samples. Here, we review these advances in cryo-EM, especially in that of protein cages, and introduce several tips for situations we have experienced.

Key words Cryo-electron microscopy, Image processing, Apoferritin, Viruses, Giant viruses, 3D reconstruction, Single particle analysis

1 Introduction

The study of protein cages by electron microscopy begins at almost the same time as electron microscopy itself, with some of the first reports studying Poxviruses [1], tobacco mosaic virus (TMV) [2], and bacteriophage [3]. The development of sample vitrification [4] to trap the subject of interest in a thin film of amorphous ice paved the way for improving the biological relevance of the studied samples, as the sample is preserved in a condition that is hoped to be as close to a native state as possible. Indeed, subjects of interest can often be captured in multiple distinct conformations, lending structural information to mechanistic studies of biological systems. Cryo-electron microscopy (cryo-EM) is a complementary

structural technique to nuclear magnetic resonance (NMR) and X-ray crystallography, both of which excel at the study of smaller objects, but which are much more difficult to utilize as the subject increases in size. Further, very small quantities of sample can be used for cryo-EM, which can be of great help with biological studies where large quantities of sample can be difficult or costly to obtain.

A paradigm shift occurred in the field of cryo-EM with the development of direct electron detectors, which can directly observe the transmission of a single electron through the sample to the detector as an “electron event,” rather than using a scintillator or photographic film to record micrographs [5]. Combining this with advances in computer technology has permitted acquisition of micrograph “movies,” grouped into individual frames which allows correction of sample drift [6]. Sample drift can be caused by both the physical hardware (“stage drift”) and the electron beam warming the vitrified ice before sample destruction occurs. Some direct detectors can use so-called “super resolution” acquisition. In this acquisition mode, the algorithm estimates where an electron entered the detector with sub-pixel accuracy [7]. The paradigm shift was termed the “resolution revolution” [8], the period of transition from resolutions where it was difficult or impossible for single particle cryo-EM to determine secondary structure to so-called “near” atomic resolution. True “atomic resolution” has finally been achieved with two recent reports studying apoferritin [9, 10].

In cryo-EM, there are distinct methods for studying protein cages: single particle analysis (SPA) [11], tomography [12], and a hybrid of the two called sub-tomogram averaging (STA) [13]. While sample preparation is common between the methods, acquisition and analysis share only broad similarities.

SPA takes tens or hundreds of thousands of 2D projections on separate micrographs imaged via cryo-EM, and separates them into individual particles within “boxes” which are usually smaller than the original micrograph [11]. The boxed particles are aligned to build a 3D reconstruction of the subject of interest. There are many software suites for this purpose, using different techniques and with varying degrees of user-friendliness (https://en.wikibooks.org/wiki/Software_Tools_For_Molecular_Microscopy). For most SPA, the sample is kept perpendicular to the electron beam, and the variant views are obtained by the random orientations of the subject when the solution was vitrified. Sometimes this can result in a lack of diversity in orientation, which in turn causes difficulties with and errors in 3D reconstructions. To compensate for this orientation bias, it is often possible to tilt the sample, which improves the sampling around the Euler sphere and removes the artifacts caused by a preferred orientation [14].

Tomography, on the other hand, requires tilting the subject [12]. A tomographic 3D reconstruction is created from different views of the same subject of interest. It is a powerful technique, particularly for study of objects *in vivo*, but cannot compete with SPA for absolute resolution of a single subject. This is mainly due to the observation of a single subject, rather than tens of thousands. This can help in situations where the subject of interest is highly heterogeneous, as is the case in some natural protein cages. One limitation of tomography is caused by the requisite design of cryo-EM sample grids. To get the best penetration of the sample, the vitrified ice layer needs to be as thin as possible. Because tilting the sample for tomography increases the thickness of the ice in the direction of the electron beam, it leads to decreased sample penetration by electrons. Further, the sample on the grid cannot be tilted 90°. As a result, tomographic reconstructions suffer from what is called the “missing wedge” effect, where a section of the tomogram cannot be reconstructed. Specially designed sample holders and dual-tilt strategies can decrease, but not completely eliminate this effect [15].

A hybrid of the two techniques, sub-tomogram averaging (STA), uses multiple tomograms for further 3D averaging and reconstruction [13]. Presuming that the sample distribution is random and there is no significant orientation bias, the signal-to-noise ratio and the resolution improve, and tilt-based artifacts can be minimized or overcome. Because of this, some STA is reaching resolutions competitive with SPA [16]. Tomography and STA can permit localized viewing of the subject *in situ*, allowing for visualization within a cell. This is especially beneficial for bacterial micro-compartments, chaperonins, and viruses in the cell. Examining the environment and interactions on a viral/host scale can reveal details that cannot be obtained by examining the dynamic interactions of two individual proteins between a virus capsid and a host receptor [17, 18].

Radiation damage is a constant challenge in cryo-EM, particularly of biological samples. Vitrified cryo-EM samples are much more easily damaged by the electron beam than samples prepared with heavy metal compounds that act as contrast agents. Because vitrified samples are critically sensitive to damage from the electron beam, areas that have suffered radiation damage can often be easily identified and discarded (*see Note 1*). Modern cryo-EM uses “low-dose” acquisition techniques to minimize the damage the beam causes to the subject. A significant body of work has been put into optimizing a balance between image contrast and minimizing damage to the subject. Higher accelerating voltages can also be used to decrease (relative) damage to the subject; however, their modern advantage is in sample penetration [19–21], and thus far only a subset of the very largest protein cages may benefit from this [22]. Our recent report used Tokyovirus, a 250 nm giant virus,

to examine the utility of high-voltage cryo-electron microscopy (cryo-HVEM) at 1 MV for the study of the larger protein cages at sub-nanometer resolution [23]. While it shows promise, particularly in reducing the defocus gradient caused by the Ewald sphere, the cryo-HVEM has more limiting factors than modern mainstream 200 or 300 kV cryo-EM [23] (*see Note 2*).

Once data have been acquired, image processing takes place on (normally) Linux-based computers. These are often equipped with powerful graphics cards (GPUs) at present. Cryo-EM data processing is mathematically intensive and can be almost embarrassingly parallelized with sufficient resources. This means modern GPUs are ideal for the task; however, some processing suites are still CPU focused while remaining competitive in terms of processing speed. Popular programs for cryo-EM image processing include cisTEM [24], RELION [25], and cryoSPARC [26] (*see Note 3*). Other options available include Bsoft [27], EMAN2 [28], SPHIRE [29], Scipion (incorporating XMIPP) [30], cow-EM (<https://www.cow-em.de/>), and Warp/M [31]. Historical programs include FreAlign [32], IMAGIC [33], and SPIDER [34]. Supporting programs, which are focused on a single aspect of the cryo-EM image processing pipeline, are UNBLUR [35], MotionCor2 [36], CTFFIND [6], and Gctf [37].

Here we will describe the processing of two very different examples of protein cage by SPA: apoferritin, which has become the de facto benchmark protein for cryo-EM and requires little to no previous experience to achieve a good result, and the giant virus Melbournevirus, which requires some special techniques developed specifically for objects of that scale when studied by cryo-EM.

2 Materials

For preparation of biological samples, most of the necessary equipment will already be present in a “wet” laboratory which can isolate or produce the subject of interest. Other items necessary will most likely be provided (at least for exploratory studies) by the cryo-EM facility with which the reader is collaborating.

2.1 Cryo-EM Data Collection

Access to an electron microscope, sample grids, a manual plunge-freezing setup or an automated cryo-plunging device such as a Vitrobot (Thermo Fisher Scientific), and a sample is required. Facility staff will always be happy to help. For readers not yet involved with cryo-EM but wishing to test it yourself, a cryo-EM dataset is required to experiment with. The electron microscopy public image archive (EMPIAR) [38] database contains many datasets of varying sizes and levels of difficulty to experiment with and learn how different data behaves when processed (*see Note 4*). Here, we will use the EMPIAR dataset 10337, consisting of

micrograph movies of a mouse heavy chain apoferritin, which was acquired with a 200 kV microscope equipped with a Gatan K2 Summit using super resolution mode for acquisition [39]. The Melbournevirus dataset we examine will be made public as soon as possible. It was acquired with a 300 kV Titan Krios G2 equipped with a Gatan K2 Summit direct detector using counting mode for acquisition [40].

2.2 Image Processing

For image processing a Linux PC is required, which is preferably equipped with a pair of recent nVIDIA GPUs supported by recent CUDA releases. A pair of GTX1080 cards should be considered the minimum, although video memory (VRAM) is arguably more important than absolute processing power, therefore 8GB of VRAM should be considered the minimum (*see Note 5*). For working on whole particle giant viruses, GPUs with >40GB of VRAM will be required. Different processing suites have different requirements: cisTEM does not require GPUs, cryoSPARC does, and RELION has both a GPU and CPU pathway, although the CPU pathway is dependent on Intel Xeon CPUs with AVX-512 (*see Note 6*). An affordable (but generally capable) system for processing most data would be akin to an AMD Ryzen 3900X CPU, 128GB RAM, 2× nVIDIA GTX1080Ti GPUs, a large hard disk drive (HDD) for storing raw data, and a large solid-state drive (SSD) for data processing. This was the system we processed EMPIAR-10337 on. Dedicated facilities may provide computer clusters, etc. (*see Notes 7 and 8*).

3 Methods

3.1 Sample Preparation

A stable and homogeneous sample is always desired. Unfortunately, biology often has different ideas. β -galactosidase and apoferritin are popular standard samples and benchmarks because they are generally well-behaved proteins that can be mistreated in a variety of ways and still be a good sample (*see Note 9*). Other samples may be heterogeneous, or easily damaged, or light or oxygen sensitive. It is a good idea to have a solid understanding of your subject of interest in normal laboratory conditions before thinking about cryo-EM. Unlike X-ray crystallography or NMR, concentrations do not need to be particularly high.

3.2 Grid Selection, Preparation, and Sample Application

Grid selection can play a dramatic effect on how the sample behaves. In the days of manual data acquisition, grids were often hand-made and described as “lacey carbon film,” which meant that the holes in which you hoped to capture the thin layer of ice (and your subject of interest) were random shapes and sizes. This (at least in theory) improved the chances of finding an area of the sample grid with conditions that were “just right.” With automated

acquisition, however, a much more uniform type of grid, where all the holes are a defined size and in a defined spacing, is preferred. Different models of grid may have holes with different diameters and spacing (*see Note 10*).

The material from which the porous grid is made also plays a role, where copper and molybdenum grids are often used. Gold-deposited porous grids have become popular as they demonstrate exceptional stability [41]. Furthermore, an additional thin carbon film may be applied to the porous carbon grid. This can help with issues such as samples avoiding the holes or occasionally taking preferred orientations. “Glow discharging” a grid will change the surface of the grid to increase or decrease hydrophilicity depending on the gas mixture present during operation. If left too long, or set too high, however, this can become “plasma cleaning” which will strip all the carbon off the grid. This is a bad thing. Applying the sample to a grid is now often done with mechanical systems in the name of reproducibility. Despite this, no two grids will ever be the same. Manual blotting is still carried out in some laboratories, with superb results.

3.3 Data Acquisition

3.3.1 Physical Hardware

While a choice may not be available, the type of sample may influence what microscope and/or detector best suits the use-case of a particular user. A detector with a wider field of view can provide an advantage for larger samples. But most of the time, getting the data is more important than worrying about precisely what detector will be used.

3.3.2 Accelerating Voltage

This is not usually a difficult choice. High voltage microscopes are few, and access to them limited. Further, unless your subject of interest exceeds 300 nm diameter, they will provide no advantage over mainstream electron microscopes. Meanwhile, access to 200 kV systems at individual institutions is usually possible, and access to cutting-edge 300 kV systems from Thermo Fisher or JEOL is likely via a working group, consortium, or national facility upon application. The 200 kV systems are now often referred to as “grid screening” or “optimizing” systems, where samples and grid preparation can be tested before acquiring the primary dataset on a 300 kV system. For some samples, 200 kV can be competitive with 300 kV [39].

3.3.3 Magnification

Magnification selected for SPA will depend on the subject observed and the target resolution. Larger subjects, such as giant viruses, will require lower magnifications just to get a single complete particle on a micrograph. Smaller subjects, such as apoferritin, can be imaged at high magnification and still have several hundred particles on each micrograph. Magnification will also depend on the physical size of the detector. Detectors with a larger number of

physical pixels and thus a physical area, such as the DE-64 (Direct Electron LP), can be used at higher magnifications without particle count dropping too low.

What we care about here is pixel scale or Angstrom-per-pixel ($\text{\AA}/\text{pixel}$) rather than magnification specifically. Different mounting systems and physical dimensions of the detector for different cameras will mean the pixel scale is different between detectors at the same “magnification.” So “magnification” is not such a useful descriptor as it once was (*see Note 11*). Having said that, it is a balancing act. No one wants to have to go back to collect data again, but too low a magnification will not result in high-resolution reconstructions. Except for those groups aiming for record-breaking resolutions of a given subject, the majority of SPA for aiming at building atomic models of proteins is carried out at magnifications that will correspond to a pixel scale of $\sim 0.8\text{--}1.5 \text{\AA}/\text{pixel}$ (not including super resolution data).

3.3.4 Electron Dose

Due to radiation damage, lower doses of electrons are preferable. Modern detectors, particularly when using counting mode or super resolution acquisition, require a low number of electron events per sensor readout so that electron counting is possible. The speed at which the detector records is termed, “frames per second,” which, depending on the detector, may not actually be what is output as a micrograph movie after processing by the detector computer controller. However, if there are too few electrons, the signal-to-noise ratio will be too low. A target electron dose of $\sim 50 \text{ e}^-/\text{\AA}^2$ total dose (for each micrograph) is reasonable.

3.3.5 Defocus

A greater defocus gives higher contrast images, which in turn make it easier to determine the quality of the grid and the applied sample “by eye.” However, it potentially limits the maximum resolution obtainable from data as the envelope function that determines the intensity of high-resolution signals falls (close) to zero more quickly. If aiming for “record-breaking” resolutions, it is now common to target defocus values between -250 nm and -750 nm ($-0.25 \mu\text{m}$ to $-0.75 \mu\text{m}$). For general data acquisition, however, a reasonable target defocus range is -1000 nm to -3000 nm ($-1 \mu\text{m}$ to $-3 \mu\text{m}$). Due to differences in particle position within the ice layer, ice thickness, and grid flatness, defocus will vary naturally between micrographs and between particles. Particularly high defocus particles can be removed individually if necessary.

3.3.6 Quantity of Data

This will depend on the hardware used for acquisition. Most modern electron microscopes come with automation software developed by the manufacturer (e.g., the JEOL CRYOARM series or the Thermo-Fisher Scientific (formerly FEI) Arctica, Glacios, and Titan). Many microscopes that do not have official manufacturer

support for automation can be automated through the installation of Leginon [42] or SerialEM [43] software. The detector used will have a significant impact on the speed of data acquisition; the Gatan K3 and Thermo-Fisher Falcon 4 are very fast detectors, capable of recording multiple micrograph movies every minute when combined with “beam-image shift” (beam tilt) acquisition to reduce the need for physical movement of the sample stage within the microscope [44]. Because of this, acquisition of thousands of micrographs in less than a day is now possible. Even with older detectors, acquisition of 1000–5000 micrographs for a single dataset is not unreasonable. However, if using cryo-HVEM, automated acquisition is not currently available [23]. Further, cryo-HVEM micrograph movies typically required longer acquisition times due to decreased contrast at higher accelerating voltages. When we acquired Tokyovirus data in super resolution mode, the rate of acquisition was slowed to somewhere between six and ten micrograph movies per hour (*see Note 12*). As such, quantity of data acquired will depend on available microscope time, equipped detector, quality of sample, and storage and processing capacity of the user.

3.4 Image Processing

3.4.1 Importing Micrograph Movies

Assuming a dataset has been collected and acquisition went smoothly, all modern cryo-EM SPA processing broadly follows the same path: motion correction, contrast transfer function (CTF) estimation, particle selection, particle classification, initial model generation, refinement, and reconstruction (Fig. 1a). As we primarily use the RELION processing suite [25], we will predominantly explain options and functions as viewed through the options available in RELION (Fig. 1b). Occasionally, we have experienced better results at one step or another by using one of the other cryo-EM image processing suites. It is in the later stages of processing where the treatment of data may be different depending on the dataset.

RELION 3.1 [45] and RELION 4.0 [46] can carry out refinement of electro-optical parameters for the dataset, but for this to work best, the micrograph movies need to be divided into different subsets, called “optics groups.” Defining individual optics group is not necessary until later stages of image processing where these subsets will play a role, but it can be easier to sort this step out now, rather than later. The import step can be carried out multiple times, importing a different subset of micrographs each time, though this can be laborious and slow (*see Note 13*). Everything can be imported at once and the resulting `movies.star` file can be edited by hand to create groups manually (*see Note 14*). The “Subset Selection” and “Join Star” functions can also be used with a small amount of manual editing after the motion correction step. Further, datasets that require different gain references of the camera

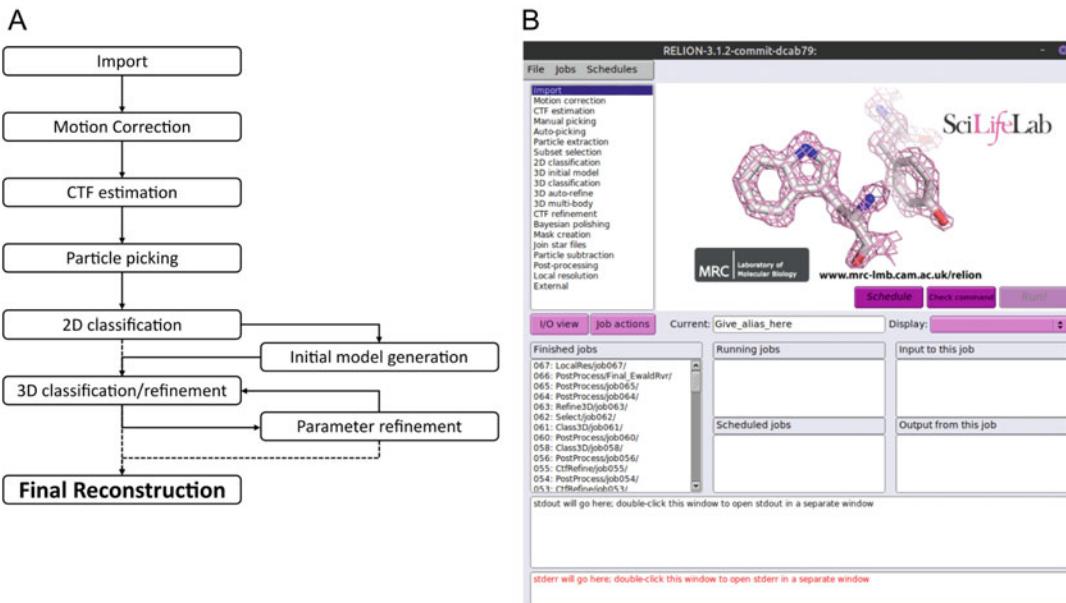


Fig. 1 (a) Cryo-EM SPA image processing workflow. (b) The RELION 3.1 interface [45]. When RELION is opened, a window like this one will be displayed. Different tasks are listed in the left-hand column, with their various options appearing in tabs to the right. The “pipeline” of the processing will be listed under “Finished jobs,” “Running jobs,” and “Scheduled jobs” depending on the run state. Two logs, one of general output and one of error messages, are displayed at the bottom

(*see Note 15*) will need to be treated separately during the next step, motion correction, and will therefore require separate import runs.

3.4.2 Motion Correction

The advent of dose fractionated (“movie mode”) acquisition for cryo-EM data provided a marked and dramatic improvement in quality of micrographs when used to estimate the relative motion of the image between frames [47]. “Whole frame” motion correction can correct for global drift, which is often caused by the stage and sample holder in the microscope. However, it is less capable in correcting for more subtle motion caused by the ice warming in the electron beam, which will vary depending on ice thickness, position relative to the edge of the grid hole, and sample concentration. Here, patch correction [36], which divides the micrograph up into smaller areas and calculates the movement of each segment before reassembling the whole micrograph with a polynomial motion, is better (*see Note 16*).

The second part of this first step, which is carried out at the same time as motion correction on supported data (*see Note 17*), is dose weighting [47]. Dose weighting reduces the relative percentage (weight) in the image sum of the later acquired frames, which will have been damaged by excessive exposure to the electron beam.

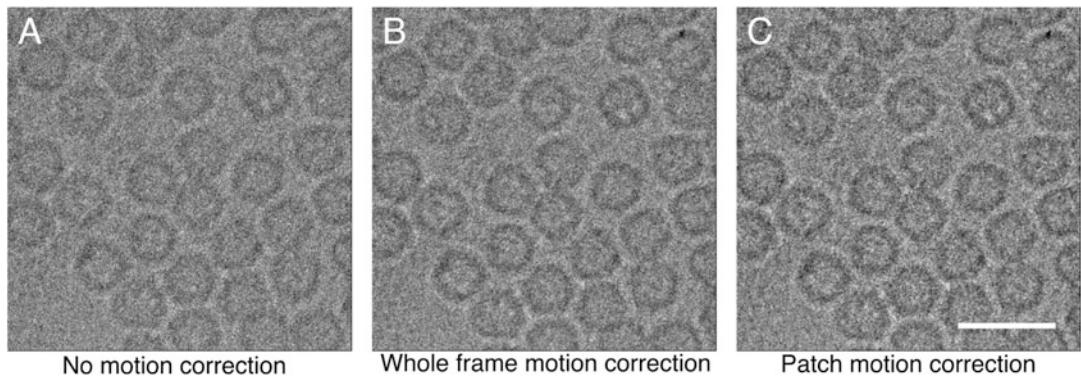


Fig. 2 The impact of motion correction on micrographs will vary, but a correctly motion corrected micrograph is always superior to an uncorrected micrograph. Here you can see the same section of a micrograph of apo ferritin from EMPIAR 10146 [24] treated with no motion correction (**a**), whole-frame motion correction (**b**), and patch motion correction (**c**). While the difference between whole-frame and patch motion correction can be subtle, the improvement in clarity versus the uncorrected image should be evident in the edges of the apo ferritin particles. Scale bar equals 20 nm

As high-resolution information is the first to be lost (and that is what we really want to keep), dose weighting can result in improvements in the quality and resolution of the result.

Figure 2 shows a section of a micrograph of apo ferritin taken from EMPIAR-10146 [24] which has undergone different motion correction schemes from simply summing all frames without motion correction to patch correction. The micrograph movie has low drift, but the apo ferritin particles are visibly clearer with motion correction. However, in this section of micrograph, there is little visible difference between whole-frame motion correction and patch correction. If a dataset requires multiple gain references, either because it was acquired at different times or because of gain drift during acquisition (*see Note 18*), motion correction runs must be carried out individually as only one gain reference can be set per run. An improper gain reference is usually immediately evident upon visual inspection of the micrograph (e.g., during manual particle selection). For this reason, we recommend a brief visual check of a few micrographs at the beginning of the dataset, and a few at the end.

3.4.3 Contrast Transfer Function Estimation

The next step of processing is CTF estimation (*see Note 19*). This is vital for high-resolution cryo-EM as the signal is delocalized across the image in an oscillating and decaying waveform, where discontinuously the low-resolution information is strongest and high resolution is weakest. By estimating the CTF, a simulated fit can be made, thus compensating for this oscillation and decay. A good estimate is vital for high-resolution reconstructions, which can be adjusted in the later stages of processing to fully optimize parameters. For many datasets, the default parameters suffice to give

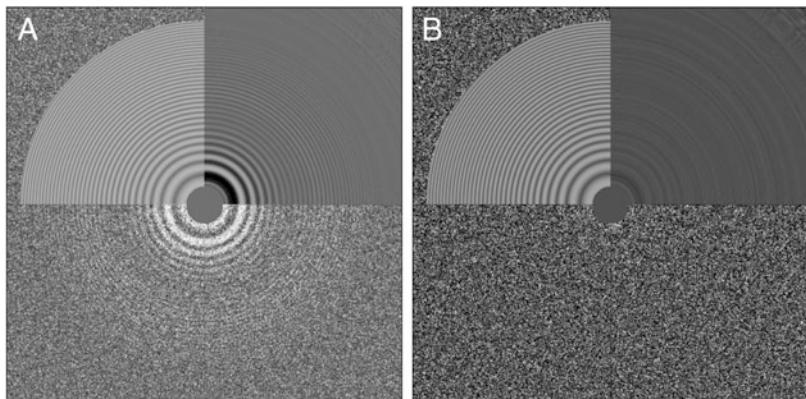


Fig. 3 A good CTF fit for a micrograph (**a**) and a poor CTF fit for a micrograph (**b**). Both micrographs of a mouse heavy-chain apoferritin are from EMPIAR 10337. A power spectrum like that of (**a**) is desired. Micrographs with power spectra like (**b**) should be discarded

good CTF estimates. It is standard practice to discard all micrographs which fail to give a good CTF estimate, as poor estimates will cause problems with later processing. Recently, it has become more common to discard all micrographs which fail to give a good fit beyond some user-determined cut-off. For the highest resolution work, this will often be somewhere with an estimated maximum resolution between 3 and 5 Å (*see Note 20*). Figure 3 shows examples of a good CTF fit, and a poor CTF fit, although in this case the CTF fit is poor because the Fourier transform shows no clear Thon rings [48]. Therefore, the micrograph is not usable.

3.4.4 Particle Selection

There has been considerable effort expended in the last few years to improve the speed and accuracy of particle selection (often called “picking”), while also trying to minimize potential bias (*see Note 21*). The strategy chosen will depend on processing suite used and dataset studied. The least-biased methods are based on Gaussian (e.g., cisTEM) or Laplacian-of-Gaussian (e.g., RELION) “blob” selection methods, although this can also pick a lot of undesired areas if the micrographs are less than perfect (*see Note 22*) or parameters are not optimized (*see Note 23*). Template picking, using a set of 2D templates which have been generated either from 2D classification of a manually or automatically selected subset of the data, can yield good results.

Whether we use fully manual, manually trained automatic, or fully automatic picking, depends on the dataset. Apoferritin is approximately spherical and is a featureless blob when treated with a 20 Å default low-pass filter for automated picking. Therefore, it can be sufficient to manually pick a dozen or so particles from each of five micrographs of varying defoci, to classify them into three classes, and to use all three classes for template-based autopicking. We followed this strategy for the reprocessing of EMPIAR-10337. Figure 4 shows a picked micrograph.

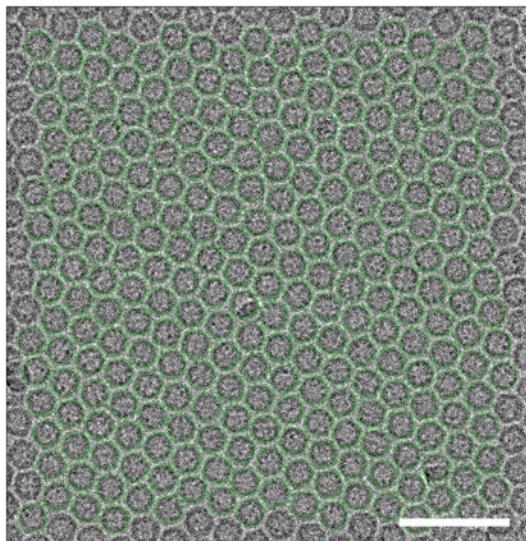


Fig. 4 A micrograph of mouse heavy-chain apoferritin from EMPIAR 10337, with picked particles shown by a surrounding green circle. Particles were picked using a template from the three classes described in the text. Scale bar equals 40 nm

For Melbournevirus, approximately 180 particles were manually picked and classified for use in template autopicking [40]. Since then, we have experimented with Laplacian-of-Gaussian (LoG) autopicking for Melbournevirus with mixed success. The LoG autopicking, even with parameters to center the Melbournevirus particle well, tends to over pick particles, particularly at edges, necessitating their removal manually (*see Note 24*). While machine learning or artificial intelligence picking has been the subject of much investigation [49], we have not tried any of those available enough to form an opinion. In our experience, automatic particle selection does not perform well on cryo-HVEM data [23]. As cryo-HVEM data must be acquired manually, the fully manual particle selection is a little tedious, but it is not unreasonable for picking to be manual as well. It also helps that with giant viruses, there are not so many particles per micrograph. Manually picking from 300 micrographs is easy, when each one contains less than a dozen particles. It is a different matter if each micrograph has more than 300 particles.

After particles have been selected, they are extracted into square boxes (of even size) (*see Note 25*), which fully contain the particle and some of the surrounding micrograph. Depending on defocus used, the box size should be varied: higher defoci need larger box sizes to account for the high-resolution signals delocalized by the CTF. At this stage in RELION, particles can be downsampled (often called “binning”), which functionally decreases the sampling

size while simultaneously reducing box size. This can both dramatically speed up and reduce storage and memory requirements for earlier stages of processing. This can be particularly useful for large datasets that are more heterogeneous than may be ideal, or for large subjects that may require very large box sizes which make processing difficult. The level of downsampling used will depend on the sampling scale originally used during data acquisition. In our reprocessing of the mouse heavy-chain apoferritin movies from EMPIAR-10337, we initially downsampled the extracted particles by a factor of five. From the non-downsampled super resolution micrographs, this corresponded to a pixel scale of 1.405 Å/pixel (*see Note 26*). Downsampling has another advantage during 2D classification. The effective X/Y translations (for a given setting of the value in RELION) of the particle are increased with respect to the level of downsampling. If particles show some variations in centering, this can improve class convergence.

3.4.5 Particle Classification

Two-dimensional (2D) classification is used to sort particles into different types, with the intent to be grouping similar particles together. This increases the signal-to-noise ratio and allows selection of clear classes showing various features expected of a given subject. It also allows removal of poor-quality particles, mis-picks (where, e.g., the autopicking algorithm got confused) or “junk”—areas of ice or particles too close to the edge of the micrograph or hole edge. Giant particles are particularly sensitive to this.

Depending on the sample, it can be advantageous to run multiple rounds of 2D classification, selecting the classes between each classification run. This can prevent the loss of good particles in classes that are otherwise less-than-perfect and allow reclamation of good particles from the subpopulations. It can also remove bad particles hiding in good classes. Depending on the quality of the dataset, and depending on the number of particles obtained, a second or even third round of 2D classification can make later 3D processing a little smoother.

One option in RELION which can have a particularly large impact on 2D classification is that of “ignore CTFs until first peak.” Figure 5 shows a 2D classification of the same particle set using the same parameters for classification with and without this option enabled. The 2D classification of cisTEM works in a similar manner to that of enabling “ignore CTFs until first peak,” or at least gives similar looking results. By default, cisTEM (and cryoSPARC) begins 2D classification with a subset of the data, which can dramatically increase processing speed on larger datasets. A similar option, inspired by the cisTEM implementation, is available in RELION 3.1 but is not enabled by default (*see Note 27*). Beginning with the beta version of RELION 4, another method of 2D classification is available; “variable metric gradient descent with

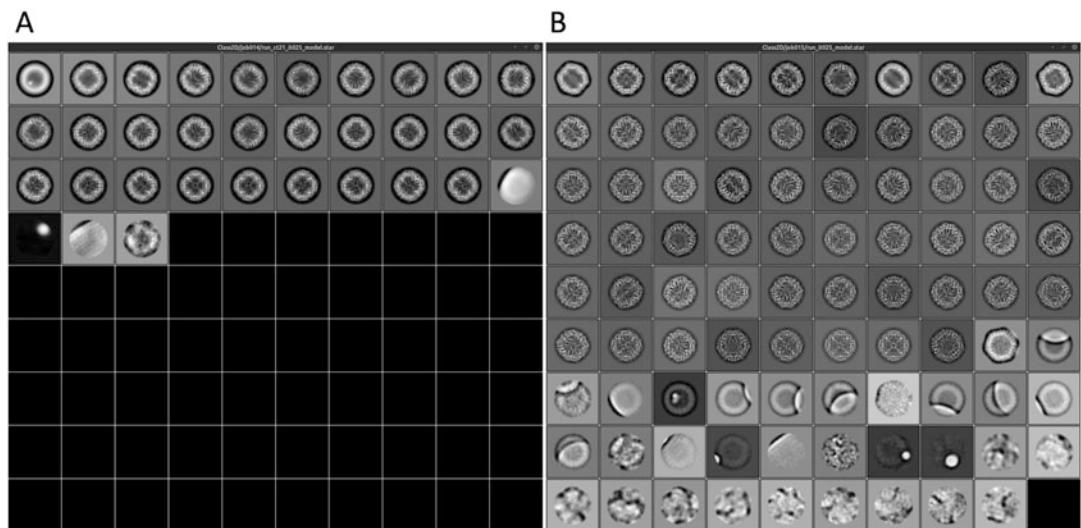


Fig. 5 The impact of a single option on 2D classification of an apoferitin. **(a)** 2D classes processed with full CTF correction. **(b)** 2D classes processed with “Ignore CTFs until first peak” enabled. The first class (top left) of **(a)** contains many good particles, but also contains many poor particles

adaptive moments” (VDAM) which in our testing is considerably faster than the classical expectation/maximization algorithm for large datasets. This has so far shown comparable results in several datasets tested, although in one dataset (which, to be fair, is an extremely difficult dataset to process) it resulted in very poor 2D classes. The *cisTEM*-inspired subset Expectation/Maximization processing is not available in RELION 4.

Sometimes, selection of good classes and re-classification of other classes can yield further good classes. This is particularly true for apoferitin (Fig. 6) where classes that contain a large number of particles can lack clarity and thus many good particles could be discarded by accident. We recovered nearly 50,000 particles that would have otherwise been discarded (approximately 15% more particles) by doing this. Two or three extra rounds of 2D classification can pay dividends here, but the fourth usually recovers so few particles that it is not worth it. There were just 653 particles selected in the fourth round of 2D classification in our reprocessing of EMPIAR-10337 (Fig. 6).

Masking during 2D classification in RELION is a simple circular mask. This mask should be loose enough to contain the particle with some clear background around it, but tight enough to remove anything in the box which is not the particle of interest. It is not always possible if the particles are oddly shaped and sample concentration is high. Protein cages like apoferitin and icosahedral viruses fill a circular mask well. Apoferitin can have a remarkably tight mask fit around it (Fig. 6).

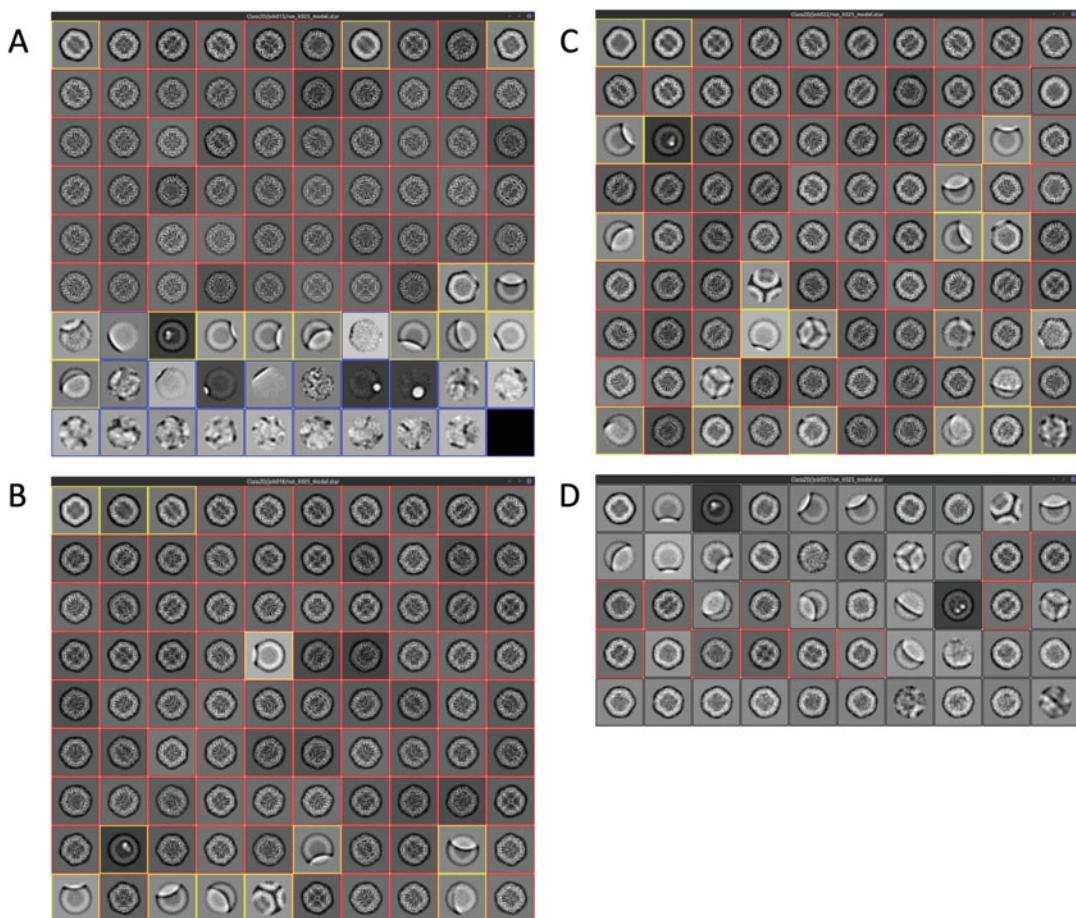


Fig. 6 Multiple rounds of 2D classification of EMPIAR 10337 (mouse heavy-chain apoferitin) to isolate good particles from less clear classes. **(a)** First 2D classification, red-boxed classes were selected as good, blue-boxed classes were discarded, and yellow-boxed classes were taken forward to another round of 2D classification. **(b)** Second round of 2D classification using the yellow classes from **(a)**. **(c)** Third round of 2D classification using yellow classes from **(b)**. **(d)** Fourth round of 2D classification using the yellow boxed classes from **(c)**. This time, there were fewer classes as there were few particles carried forward

3.4.6 Initial Model Generation

For 3D classification and refinement, a starting reference model is required in RELION (and cistem) (*see Note 28*). Initial model generation remains one of the more challenging aspects of bias-free cryo-EM image processing. It is also possible to use a previously reported map (perhaps downloaded from the electron microscopy databank (EMDB) [50] or from a preliminary dataset) or generate one from a PDB model (*see Note 29*). Success or failure of a single run of initial model generation does not always mean the dataset is poor. We have had datasets where we have needed to run initial model generation a few times to get a model which looks appropriate to the 2D class averages. Often increasing the number of particles used for initial model generation will help. For small

symmetric objects, increasing the start and end resolutions may result in a better initial model. For larger objects, this can be lowered. RELION 3.1 uses a stochastic gradient descent algorithm for initial model generation similar to RELION 3.0, where symmetry is imposed throughout all steps, while RELION 4.0 uses the newly implemented VDAM algorithm and provides the option to impose symmetry throughout model generation, or process without symmetry and apply desired symmetry as the final step.

On occasion, we have preferred to move initial model generation to another processing suite, usually cisTEM (*see Note 30*). If exporting the dataset from RELION into cisTEM for initial model generation, a few things need to be observed. Firstly, export the particles that have already been classified and selected. This can be carried out with the following command:

```
relion_stack_create --i Selected/good1/particles.star --o cisTEMstackExport --one_by_one
```

Where the input (“*--i*”) describes the location of the .star file defining particle positions, and output (“*--o*”) describes the filename used for the particle stack (.mrcs file) and particle definitions (.star file). The “*--one_by_one*” option reduces memory load for generation of the stack and does not dramatically slow writing the file output. One important point is that the current release of cisTEM does not support RELION 3.1 format .star files and the first section (the table of optics parameters) of the .star file must be removed before cisTEM can read the file (*see Note 31*). As cisTEM requires manual input of acceleration voltage, pixel size, and amplitude contrast, no information it uses is lost. Once the particle stack has been imported into cisTEM, one may either jump straight to initial model generation, or run a quick 2D classification in cisTEM as a check. As initial model generation is extremely sensitive to poor quality particles, we would normally do a quick 2D classification run and select the clearest classes for initial model generation. Upon completion of initial model generation, copy the initial model to the RELION processing directory for use as a reference. At this point, the cisTEM processing directory should not be too large, so it can be left alone or archived. Use of cisTEM for initial model generation may well not be necessary.

For EMPIAR-10337 (mouse heavy-chain apoferritin), a correct (but inverted handedness) initial model was generated in RELION. It is impossible to determine the chirality or “handedness” of a particle only by the method of acquiring cryo-EM data. The “correct” 3D object gives rise to the same set of projections (the 2D images captured in the micrograph) as its mirror image (*see Note 32*). During processing, whether the handedness of the reconstruction is “correct” or inverted makes no impact on attained

resolution (*see Note 33*), but the handedness must be corrected for model building. Fortunately, changing the handedness is easy, for example, in RELION, a single command will invert the handedness of a 3D reconstruction:

```
relion_image_handler --i run_class001.mrc --o
run_class001_invertedHand.mrc --invert_hand
```

3.4.7 3D Reconstruction and Refinement

Once an initial model has been generated, one can either move directly to 3D refinement, or first carry out 3D classification. Historically, we have pursued 3D classification first, and then refined the best resulting classes. However, in recent years, we have found moving straight to 3D refinement with all particles selected after 2D classification, followed by optical parameter refinement and 3D classification afterwards to yield the same or better results, although it depends on the dataset.

For a first 3D refinement, the downsampled particles may still be used. A mask can be generated from the initial model or not, depending on the shape of the particle. Early refinement of apo-ferritin and giant viruses can be carried out using a simple, spherical mask defined in the RELION graphical user interface (GUI). Later refinements or particles that fill a sphere poorly will benefit from a custom mask. As RELION and cisTEM deal with normalization and greyscale levels differently, remember to set the “Ref. map is on absolute greyscale?” option to “No,” if using an initial model not generated by RELION. Low-pass filtering of the initial model (and subsequently limiting the resolution used for alignment) is used to avoid aligning noise [51]. However, for highly symmetric subjects like apo ferritin or icosahedral complexes, this can result in poor particle alignment in the Expectation/Maximization algorithm used by RELION, resulting in a bad final model and a poor resolution at the same time (Fig. 7). Sometimes decreasing the initial angular sampling increment to 3.7° (or even 1.8°) can help avert this issue, while otherwise a 3D classification to align particles before starting with local 3D refinement may be needed.

Once a good reconstruction is obtained, electro-optical parameter refinement can begin in earnest. This requires some post-processing-related information, so cycles are carried out after an initial post-processing run. This post-processing run estimates the solvent-corrected global “gold standard” Fourier Shell Correlation (FSC) resolution [52]. For electro-optical parameter refinement, exactly how to proceed is left largely up to the user. We normally proceed with magnification anisotropy estimation first, as it can be done quickly. However, it will not always have a significant effect on estimated resolution. For the apoferritin processing shown here, it did not improve resolution. For the Melbournevirus reconstruction, magnification anisotropy correction improved the estimated

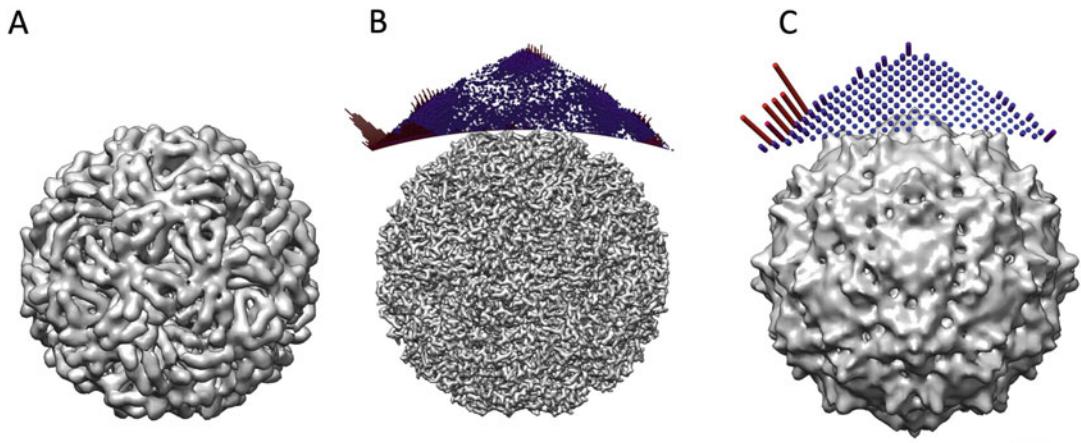


Fig. 7 An example of the effect that an incorrect local minimum may have on a reconstruction, using lumazine synthase from EMPIAR 10384 [71]. **(a)** The initial model used. With a maximum resolution of 8 Å imposed, α -helices can be visualized as “sausage” shapes. **(b)** A good 3D reconstruction, with the assigned angles of particles shown above as cylinders colored from blue (few particles) to red (many particles). This reconstruction achieved 2.2 Å from ~10,000 particles. **(c)** A reconstruction where the alignment settled on a poor local minimum, resulting in a reconstruction that looks visibly wrong for the estimated resolution (~5.5 Å). Scale bar equals 2 nm

resolution of the reconstruction from 7.1 Å to 6.6 Å [40]. Defocus and astigmatism, which were the parameters originally calculated during CTF estimation, may be adjusted at this point on a per-micrograph or per-particle basis. For a dataset with many particles across a micrograph, per-particle defocus and astigmatism refinement can dramatically improve achieved resolutions (*see Note 34*). We never use per-micrograph defocus refinement. Beam tilt estimation in RELION 3.1 and 4.0 can individually estimate second-, third-, and fourth-order electro-optical distortions [45]. The impact this will have on reconstructions depends on the resolution already obtained, how well aligned the microscope was before acquisition began, and method of acquisition (beam shift acquisition is fast, but requires tilting the beam, affecting results at the highest resolutions [44]). Estimating per-particle defocus, per-particle astigmatism, and all three beam tilt orders for the apoferritin dataset EMPIAR 10337 improved resolution from 2.17 Å to 1.8 Å (*see Note 35*).

At this point, we moved to Bayesian polishing. Bayesian polishing (often just “particle polishing”) permits improvements in estimating the motion of individual particles in each micrograph. As ice thickness will vary across a micrograph, particles may move differently under the effect of the electron beam [53]. For datasets that are very stable and have little motion (like EMPIAR 10337), this can have a negligible effect. For data where drift is greater, Bayesian polishing can have a more marked effect on improving obtained

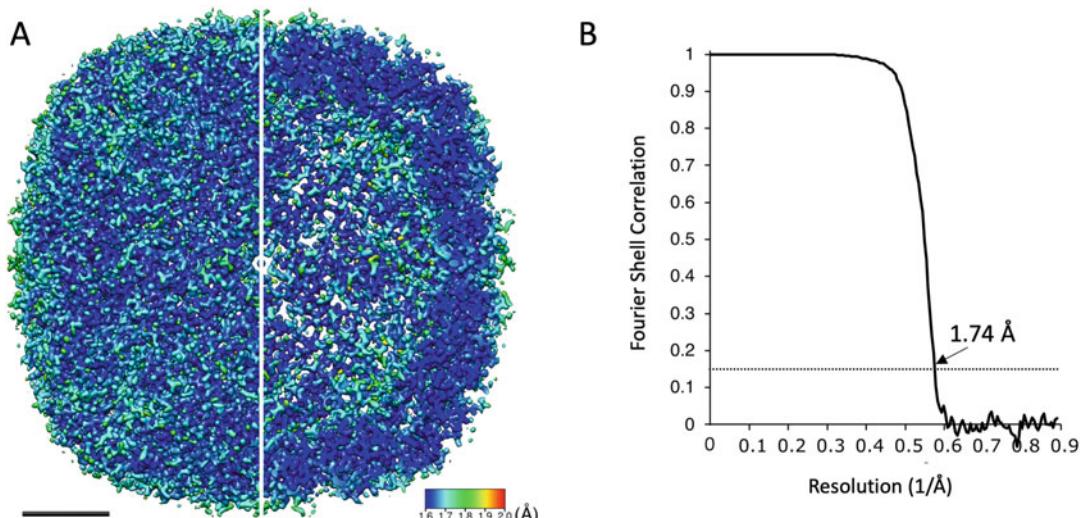


Fig. 8 The result of processing EMPIAR 10337 (mouse heavy-chain apoferritin). **(a)** Apoferritin, colored by local resolution which extends from $\sim 1.8 \text{ \AA}$ (estimated with the blocres module of Bsoft) to $\sim 1.6 \text{ \AA}$. The color code is indicated. The map is viewed from the outside (left) and inside (right). **(b)** The corresponding Fourier shell correlation (FSC) curve estimating a global resolution of 1.74 \AA . Scale bar equals 2 nm

resolution. Once Bayesian polishing is completed, a final refinement of electro-optical parameters may yield a further small improvement.

Since Ewald sphere correction was implemented into RELION [54], the final step is often a final pass of `relion_reconstruct`, as Ewald sphere correction cannot currently be applied during 3D refinement. One must test both forward and reverse curvature of the Ewald sphere, as one can often result in resolution improvements while the opposite curvature results in resolution loss (or no change). Application of Ewald sphere correction is easy to run but must be carried out in the command line. A mask slightly larger than the object of interest should be applied, although otherwise the default parameters normally give good results. Here, the final resolution of apoferritin obtained after Ewald sphere correction was further improved to 1.74 \AA (Fig. 8) with local resolution extending to $\sim 1.6 \text{ \AA}$. Since the recent implementation of Ewald sphere correction in cryoSPARC, beginning in release 3.3, it is possible to apply Ewald sphere correction during 3D refinement. The algorithm used in cryoSPARC is the same as implemented in FreAlign by Wolf et al. [55] while the RELION implementation is based around the method described by Russo and Henderson [56].

Masks for 3D refinement can be better optimized to the particle of interest than 2D masks and are a separate job within RELION rather than a single option in the relevant classification menu. They are normally generated from low-pass-filtered 3D reconstructions which have already been carried out, for example, the first 3D mask

would normally be generated from the initial model. The different processing suites treat masking differently; the long-standing recommendation in RELION is to low-pass-filter to 15 Å, binarize the map, then apply a user-defined global expansion and soft edge of the resulting volume. CryoSPARC low-pass-filters the map to 12 Å and attempts to optimize other parameters automatically (while still permitting user customized adjustment if desired), while cisTEM automasking binarizes the map before applying a soft edge of defined width.

3.4.8 Analysis and Validation of Cryo-EM Maps

Analysis of cryo-EM reconstructions can be carried out in a range of different programs, depending on user preference. We tend to favor UCSF Chimera [57, 58] for visualization of 3D reconstructions. For the large 3D volumes and high polygon count of giant virus reconstructions, UCSF ChimeraX [59] can be easier to use as the rendering engine has been updated, and from version 1.3 has had several functions added which were extremely useful for cryo-EM visualization, but had not been implemented in earlier releases of UCSF ChimeraX. Some basic analysis and validation may be carried out within the processing suites themselves. CisTEM and cryoSPARC report FSC curves and angular assignment visually throughout each iteration of refinement, while to see detailed reports for each iteration in RELION requires examination of intermediate output files (*see Note 36*).

Analysis in RELION is carried out in the Postprocessing and Local Resolution sections. A separate mask should be generated for post-processing from the final output of the 3D refinement. This will be a softer mask than the one(s) used for 3D refinement, but still obscure the non-protein regions of the volume (*see Note 37*). Local resolution estimation and filtering can be useful in visualizing better ordered and less flexible regions of the map, as some regions will be higher resolution than the global estimate from the FSC curve, while other regions will be lower resolution.

Local resolution estimation is somewhat tricky, however, with different programs estimating local resolution in different ways (or permitting multiple ways of estimating). The RELION local resolution estimation does not obey symmetry imposition, so can estimate varying local resolution in areas of a reconstruction which should be symmetric and therefore identical. The Bsoft [27] local resolution estimate may have symmetry imposed, so may be preferable for high symmetry local resolution estimation (Fig. 8a). ResMap [60], one of the first programs to estimate local resolution, can be perhaps a little too generous with local resolution estimates, estimating resolutions to be higher than other competing local resolution programs when compared [61].

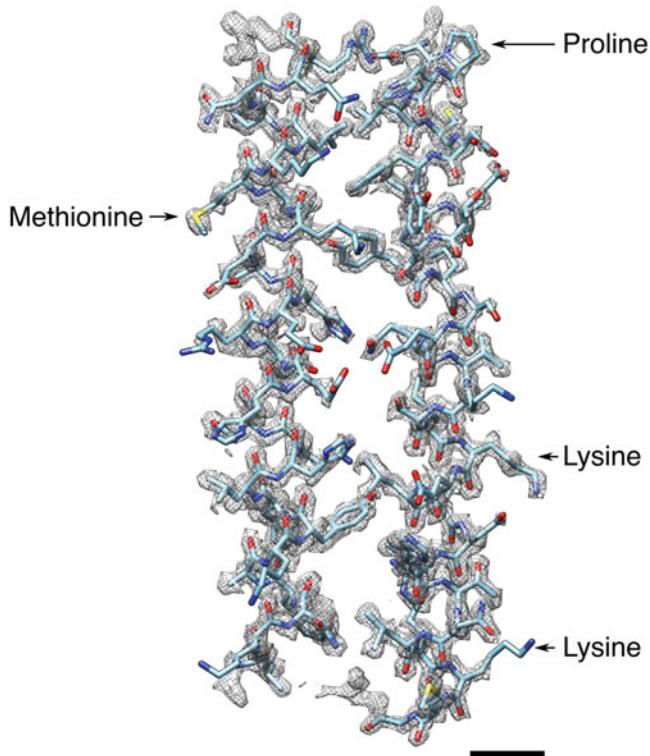


Fig. 9 Two helices of an apoferitin monomer extracted from the map in Fig. 8a with a fitted atomic model (PDB 6V21). The central “hole” of the proline ring is clearly defined, along with many other residue side chains. Lysine is often a tricky residue, as can be seen since one lysine residue (upper, marked) has clear and contiguous density for the full residue, while for another (lower, marked) much of the side chain density is not resolved. Scale bar equals 0.5 nm

One thing to note is that even if the initial model has the correct handedness, we have seen apoferitin reconstructions in RELION proceed perfectly as expected but the final output has the handedness inverted. This is easily fixed once a final post-processed map is obtained and has no impact on any stage of processing, but it is due to the initial low-pass filtering carried out to prevent model bias (*see Note 38*). If necessary, the initial low-pass-filter (which by default is 60 Å) can be set to a higher resolution, although we do not generally recommend this.

Model building, the process of either fitting and modifying an existing PDB model into the cryo-EM density map or building one de novo, can be carried out in a range of other programs. COOT [62, 63] and PHENIX [64] are favored for model building and refinement, respectively. ISOLDE allows real-time interactive optimization by molecular dynamics [65]. Figure 9 shows a pair of helices from the apoferitin EMPIAR 10337 reprocessing with a fitted atomic model from PDB 6V21 [39]. However, a detailed examination of the options available is outside of the scope of this chapter.

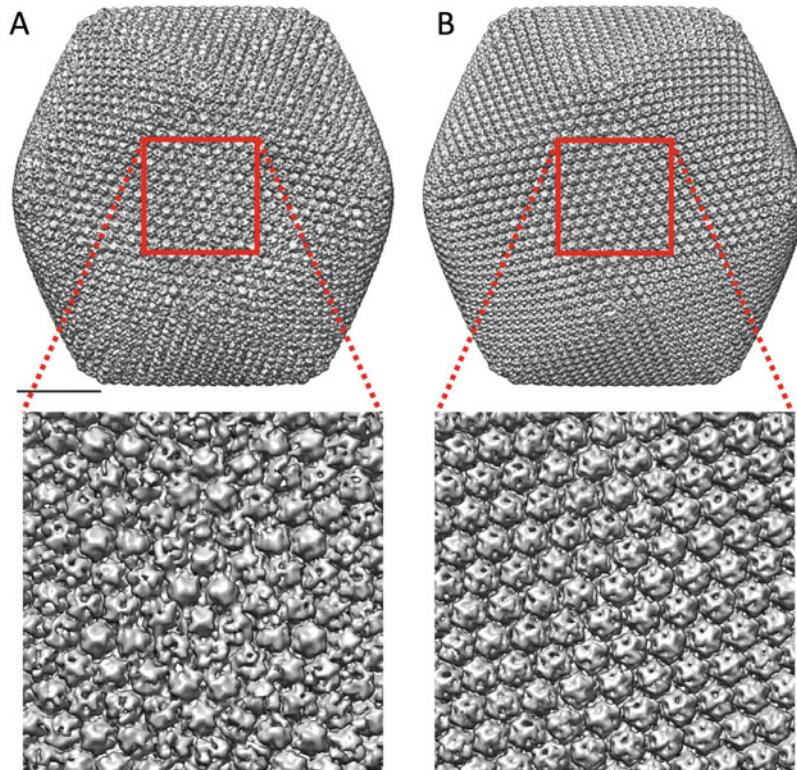


Fig. 10 A mixed handedness reconstruction (**a**) and uniform handedness reconstruction (**b**) of Melbournevirus. This is a curiosity that we have thus far only seen exhibited with giant viruses. Sometimes a refinement that started from global search parameters will converge on a model where a subset of the particles is of inverted handedness to other particles in the reconstruction. While not difficult to fix, it does not appear to have a negative impact on the resolution estimated by the FSC curve, as it does with a poor local minimum. Scale bar equals 50 nm

3.4.9 Special Considerations for Giant Particles

One curiosity we encountered when carrying out further experimental analysis using the Melbournevirus dataset [40] was as follows. When filtering down to 60 Å for the initial stage of 3D refinement or classification, if beginning particle alignment with a global search, sometimes a subset of the particles would be aligned with inverted handedness to the others. The result of this is that model looks extremely odd. It is the right shape, but the characteristic capsid protein pattern of the virus is disturbed (Fig. 10). This appears to be a completely random occurrence, as running three 3D refinements from global parameters resulted in a correct refinement twice and a “mixed handedness” refinement once. The mixed handedness particles can be classified out by a 3D classification with alignment disabled, but they are likely to be perfectly good particles which would be a waste to dispose of. Selection of the “wrong” (when compared to the other classes) handedness 3D class and 3D classification with global alignment enabled with a higher resolution (20–30 Å) reference will solve this. At that point, the class can

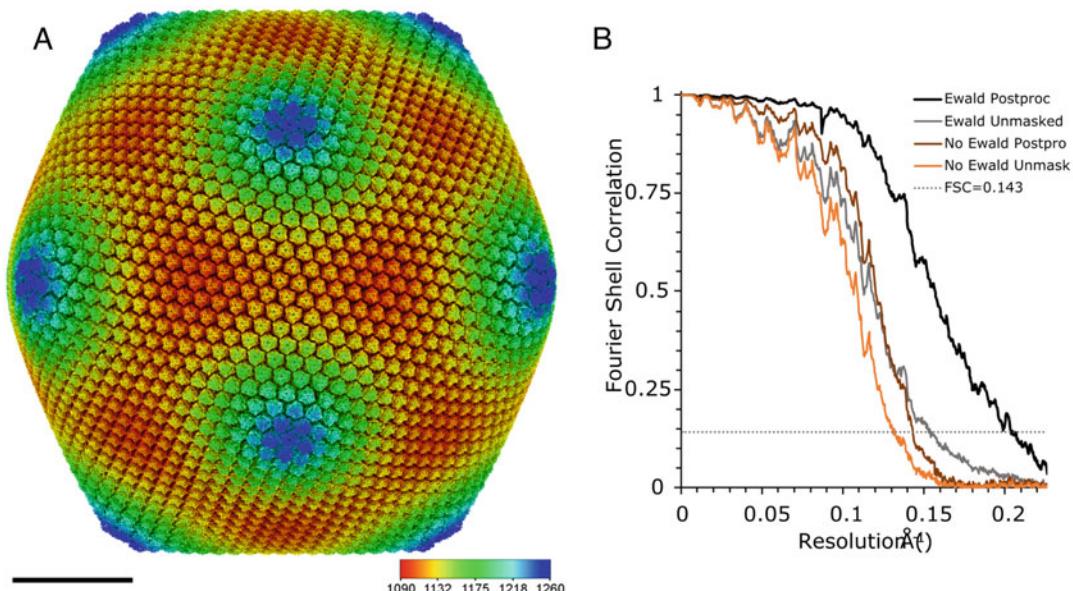


Fig. 11 (a) The final Ewald sphere corrected reconstruction of the whole Melbournevirus particle, colored by radius. Scale bar equals 50 nm. The color code of the particle radius is indicated. (b) FSC curve of the post-processed Ewald sphere corrected, solvent corrected reconstruction (black curve) and the unmasked Ewald sphere corrected reconstruction. The brown and orange curves are the same without Ewald sphere correction. Global estimated resolution of the final map was 4.9 Å

be recombined with the particles showing correct handedness from the previous classification. After recombining, future refinements should be started with local refinement. As the Melbournevirus capsid loses all features beyond some wave-like bumps at 60 Å, we attribute this to the Expectation/Maximization algorithm of RELION becoming stuck in a local minimum. We have not seen this behavior with any other icosahedral reconstruction (e.g., lumazine synthase, Murine norovirus) so it may be just one more thing for those studying giant structures by cryo-EM to be aware of.

For giant particles, the resolution loss caused by the effect of Ewald sphere curvature can be significant (Fig. 11). Once the user has determined they are unable to improve the resolution of the reconstruction further, the `run_data.star` file of the final 3D refinement should be passed to `relion_reconstruct` to reconstruct the two independent half-maps with Ewald sphere correction. For Melbournevirus, we used the following four commands to reconstruct both forward and reverse curvature half-maps (*see Note 39*):

```
relion_reconstruct --i Refine3D/job073/run_data.star --sym I3 --pad 2 --mask_diameter 2580 --width_mask_edge 5 --sectors 2 --ewald --reverse_curvature false --subset 1 --o Refine3D/job073/run_half1_class001EwaldFwd_unfil.mrc
```

```

relion_reconstruct --i Refine3D/job073/run_data.
star --sym I3 --pad 2 --mask_diameter 2580 --
width_mask_edge 5 --sectors 2 --ewald --reverse_-
curvature false --subset 2 --o Refine3D/job073/
run_half2_class001EwaldFwd_unfil.mrc

relion_reconstruct --i Refine3D/job073/run_data.
star --sym I3 --pad 2 --mask_diameter 2580 --
width_mask_edge 5 --sectors 2 --ewald --reverse_-
curvature true --subset 1 --o Refine3D/job073/run_-
half1_class001EwaldRvr_unfil.mrc

relion_reconstruct --i Refine3D/job073/run_data.
star --sym I3 --pad 2 --mask_diameter 2580 --
width_mask_edge 5 --sectors 2 --ewald --reverse_-
curvature true --subset 2 --o Refine3D/job073/run_-
half2_class001EwaldRvr_unfil.mrc

```

The half-maps should have something explanatory added to the “class001” section if you wish to make selection in the RELION GUI easier. While modern cryo-EM suites support estimation and correction of the effect of Ewald sphere curvature on reconstructions, analysis of the giant viruses has proven to further require a special adaptation to achieve optimum resolutions. This is because of their size. This technique is termed “block-based reconstruction” [66] and was developed and pioneered by Michael G. Rossmann and his collaborators to study the 190 nm *Paramecium burseria* Chlorella virus 1 (PBCV-1) [67]. Since then, it (or processing strategies based upon it) has been used for African swine fever virus (ASFV) [68, 69] and Melbournevirus [40] (see Note 40). The basic idea of block-based reconstruction is elegant in its simplicity. The defocus gradient caused by the curvature of the Ewald sphere causes a loss in accuracy of the CTF estimation due to the size of the particle, so the particle is broken down into smaller blocks. They permit localized defocus and astigmatism refinement, like per-particle defocus refinement for smaller subjects. This means the resulting reconstruction has optimized defocus parameters for the different regions of the giant virus (Fig. 12a). This works best on high symmetry particles (see Note 41) as the method calls for symmetry expansion and focusing refinement on a small section of the resulting asymmetric reconstruction.

The original block-based reconstruction scripts required icosahedral particles to be in I3 symmetry. RELION 3.1 [45] contains (almost) all the tools necessary for carrying out block-based reconstruction (see Note 42). This is useful as the electro-optical parameter refinement of RELION 3.1 is a great improvement above that of RELION 3.0 [54], which is what it required if utilizing the block-based reconstruction scripts directly. The first step is to

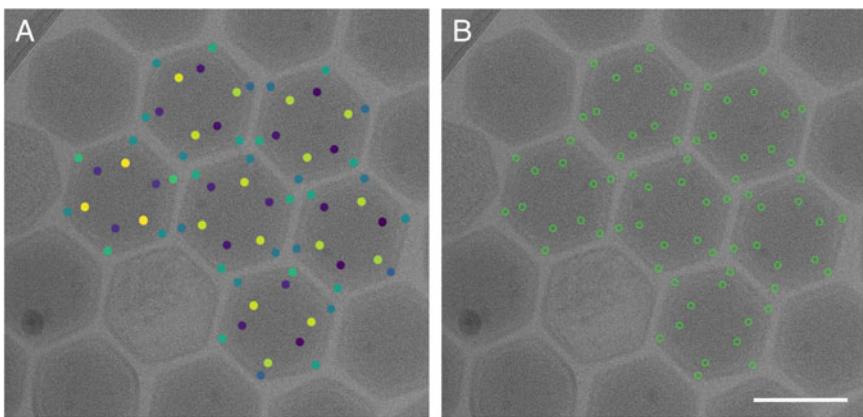


Fig. 12 **(a)** After per-particle defocus refinement has been carried out. The defocus gradient is approximately 2530 Å, ranging from purple to yellow. **(b)** A Melbournevirus micrograph showing the recentered points focusing on the fivefold block, indicated by small green circles. Scale bar equals 200 nm

carry out symmetry expansion of the final reconstruction. For example, with Melbournevirus this was carried out with the following command:

```
relion_particle_symmetry_expand --i
Refine3D/job073/run_data.star --o
Refine3D/job073/run_data_expanded.star --sym I3
```

Then, the shifts necessary to focus each block of interest in the center of a smaller box must be calculated. There are several ways to do this, but the easiest we found was to load the Melbournevirus final map into UCSF Chimera [57] and use the “Volume Eraser” tool (with the “erase outside sphere” option) with the eraser sphere centered on the penton on a fivefold axis, the MCP centermost of the threefold axis, or the middle of the MCPs along the twofold axis (Fig. 13). Once the position has been checked, the eraser sphere can be expanded before erasure outside of it. Other ways include segmentation of the map using SEGGER [70] and selection of desired regions (*see Note 43*) or fitting of a PDB model and generation of a density using the molmap function of UCSF Chimera.

This results in three maps the same size as the original, but with a small region of density and the rest of the volume empty. Using the relion_mask_create module, these deletion maps had a pixel expansion and soft edge applied, before the particle subtraction module was used with the symmetry expanded particle set to calculate the shifts. As particle subtraction itself was not desired, as soon as the shifts were calculated (which takes just less than a minute) the particle subtraction runs were terminated (*see Note 44*). The particle subtraction module requires the final

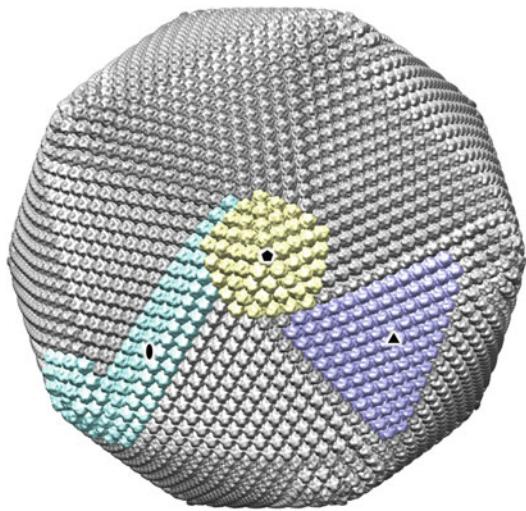


Fig. 13 Visualizing the positions of the five- (light yellow), three- (lilac), and two- (cyan) fold blocks for block-based reconstruction of Melbournevirus. Five-, three-, and twofold axes are labeled with a pentagon, triangle, and double-teardrop, respectively

optimiser.star file as an input. For example, the command to estimate shifts for the fivefold block was as follows:

```
relion_particle_subtract_mpi --i Refine3D/job073/
run_it023_optimiser.star --mask MaskCreate/job091/
mask.mrc --data Refine3D/job073/run_data_expanded.
star --o Subtract/job094/ --recenter_on_mask --
new_box 440 --pipeline_control Subtract/job094/
```

Then, the resulting output we are interested in was as follows:

```
> The subtracted particles will be re-centred on
projections of 3D-coordinate: (0.161785,
0.0771528, 515.33)
```

Using the symmetry expanded particles.star file and the calculated shifts, particle extractions were carried out in 440-, 500-, and 700-pixel boxes for the five-, three-, and twofold blocks, respectively. Then, relion_reconstruct was used with these newly extracted particles to check positioning and simultaneously generate a reference for further processing. A brief check of positioning is possible using the output from the extraction run to visualize particle box locations on the micrograph (Fig. 12b).

With careful thresholding, a mask focused on the capsid can be generated, and 3D local refinement carried out. The resulting resolution was not a significant improvement over that of the

Ewald-sphere corrected whole virus reconstruction, but a single pass of defocus refinement brought the global estimated resolution to the maximum resolution possible with the data we had (Fig. 14). Examining the defocus of the selected particles via the RELION output shows the estimated defocus gradient across the particles to be what we would expect for a subject of ~250 nm diameter (Fig. 12a).

3.5 Summary and Outlook

Protein cages are a critical component of all organisms. The study of protein cages by cryo-EM will only grow as more viruses are discovered and chaperone proteins across a wider range of organisms studied. By walking through reprocessing of the publicly available EMPIAR 10337 dataset of the protein cage apoferritin, we hope to have given those without experience in cryo-EM an example to try for themselves. Through demonstrating the processing of Melbournevirus, we show the impact of Ewald sphere correction on the whole particle reconstruction of a giant virus (Fig. 11a). For Melbournevirus, Ewald sphere correction improved the estimated global resolution by approximately 1.6 Å (Fig. 11b). By utilizing block-based reconstruction on focused segments of the viral capsid, we were able to improve resolution further after defocus refinement. This improved estimated global resolution of the blocks and achieved the maximum possible resolution with the original data (Fig. 14d, e, and f).

3.6 Data Availability

The mouse heavy-chain apoferritin data is available on EMPIAR with accession code 10337 [39]. The lumazine synthase data is available on EMPIAR with accession code 10384 [71]. The Melbournevirus dataset will be made available on EMPIAR in the future upon approval of all authors. The 1.74 Å resolution reprocessing of EMPIAR 10337, together with half maps, masks, and calculated local resolution, is available by contacting the authors.

4 Notes

1. While damage becomes evident very quickly, high-resolution information in the image is lost before the area is visibly burned. It is also obvious when an area of grid has been previously examined for too long at high magnification.
2. Indeed, cryo-HVEM is currently at a state not entirely dissimilar to that of mainstream cryo-EM just over a decade ago. Limiting factors include no automated acquisition, less stable and less coherent electron sources, and weaker software support that includes no dose weighting. The technical demands, expense, and relative rarity of cryo-HVEM will likely bottleneck further development.

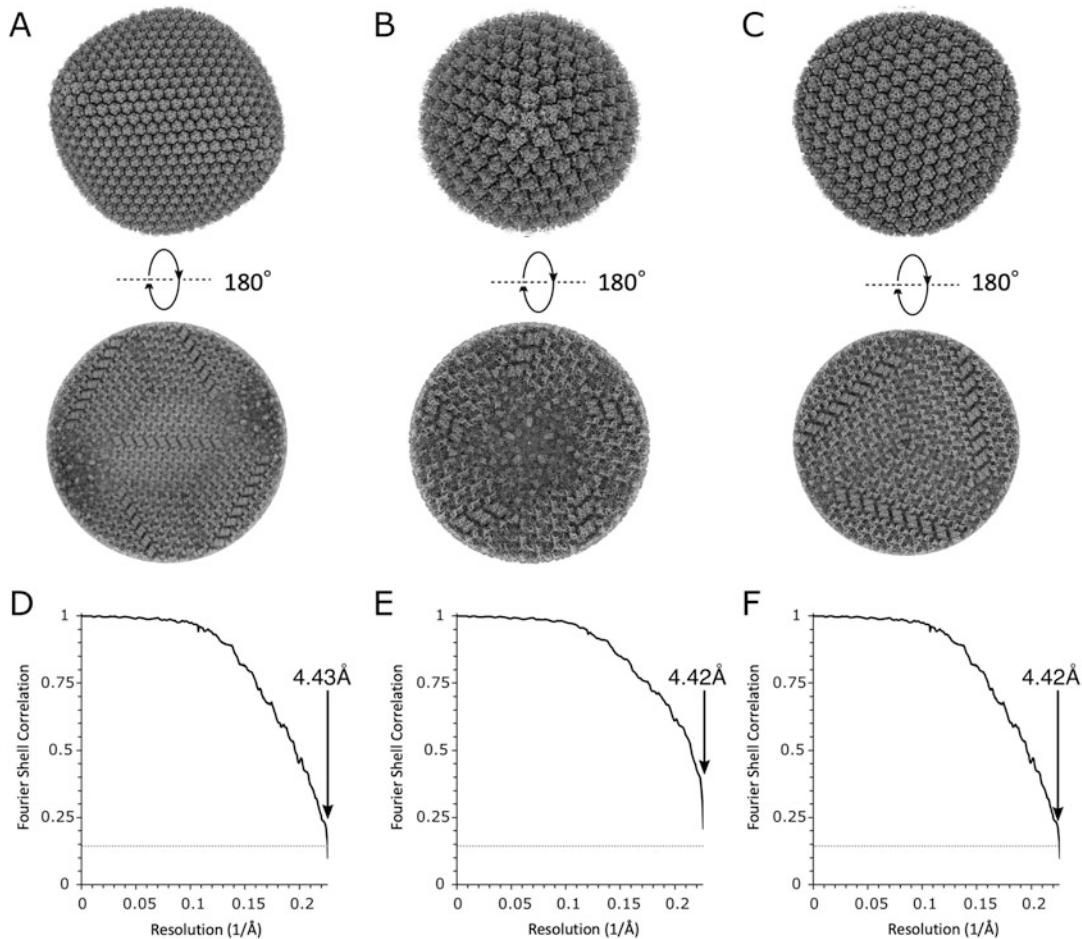


Fig. 14 The (a) twofold, (b) fivefold, and (c) threefold blocks of Melbournevirus shown viewed from external (upper) and internal (lower) of the capsid. (d–f) Corresponding FSC curves, showing that all block reconstructions reached the maximum resolution possible with the data

3. RELION and cisTEM are open source and freely available. cryoSPARC is a commercial, closed source suite that is freely available to academic users upon application. To our knowledge, except for IMAGIC and cow-EM, other cryo-EM processing suites are open source and freely available also.
4. We heartily recommend employing the EMPIAR database to “get a feel” for how different data behaves across the different suites, as the different methods used for particle selection, classification, and reconstruction can and will give different results. There are some EMPIAR datasets with which the published structure can be reliably matched or improved (as image processing software has improved since publication) while others demonstrate just how difficult some analysis can be.

5. VRAM creates a hard limit on box size, although this can be reduced by disabling Fourier padding during 3D refinement, which effectively doubles the usable box size. It should be noted that disabling Fourier padding can create artifacts in the corners of the 3D volume, however [72].
6. The vector-extensions-based CPU acceleration in RELION can be compiled and run on AVX2 hardware (e.g., AMD Ryzen or Epyc); however, the performance leaves much to be desired when compared to either the AVX-512 or GPU code paths. AMD has brought AVX-512 support to their next-generation “Zen 4” CPUs, which may level the playing field somewhat. However, compiler choice and optimizations will still affect results.
7. If access to large-scale compute resources is not possible, what are often considered “gaming PCs” with some extra system memory added can serve. Ultimately, however, it will depend on budget: all the way from a few thousand dollars (or local equivalent) to if-you-have-to-ask-you-can’t-afford-it. There are now many companies that will build systems to user specifications.
8. Analysis and processing of giant viruses means large box sizes, even when using some of the specially developed strategies for processing. Workstation/server grade GPUs (nVIDIA Quadro or Tesla cards) with 40GB+ of VRAM are basically a requirement, as is a large quantity of system memory. Due to this, the AVX-512 CPU accelerated route may prove more attractive.
9. Lyophilized protein which reconstitutes back into a functional enzyme will always astonish me. Resolubilizing crystals to regain active protein is one thing, but to completely desiccate protein is something else entirely.
10. Our recent Melbournevirus report, the processing of which we discuss here, was prepared on lacey carbon grids but acquired with automation. It does mean, however, that there may be a random carbon support film in the middle of your micrograph. See Fig. 15 for examples of worst-case, usable, or ideal micrographs from lacey carbon.
11. Currently, the detector magnification is often used rather than the number the microscope states (the nominal magnification).
12. We briefly experimented with various settings. The Tokyovirus dataset [23], where micrographs consisted of 160 frames acquired over 32 s in super resolution, caused the K2 Summit controller PC to become unresponsive to user input for several minutes post micrograph acquisition. Thus, we could acquire only six micrographs per hour. With the same settings but with only 16 s of acquisition (80 frames per micrograph), the non-responsive period was less than 30 s. Acquisition in

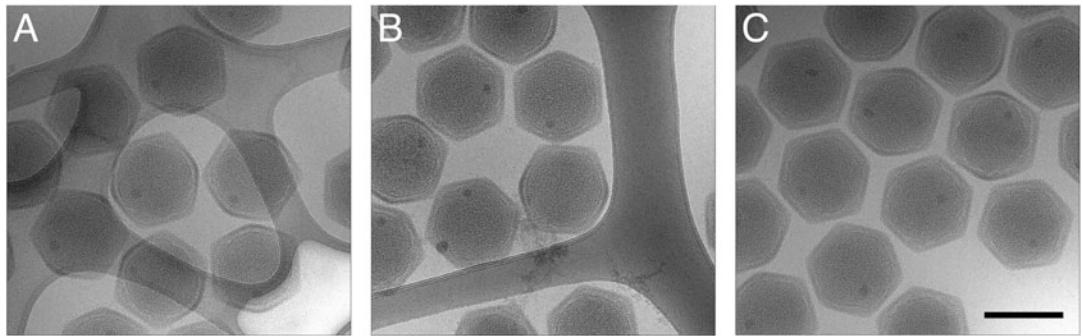


Fig. 15 Three micrographs from the Melbournevirus dataset [40], showing some of the possible micrographs obtained when using lacey carbon grids with automation. **(a)** An unusable micrograph. This is a shame as there are some nice particles here, but all overlap the carbon edge, which should be avoided. **(b)** A usable micrograph. Initially it may appear there are four usable particles present; however, three are either too close to the edge of the lacey carbon or very slightly overlapping. **(c)** A good micrograph with seven potentially usable particles, although one is closer to the edge of the micrograph than we would prefer. Scale bar equals 200 nm

counting mode (25% of the size of the super resolution micrograph frames) never became unresponsive, and finding the next good hole, positioning, and focusing became the speed limiting factor.

13. A descriptive naming scheme for acquired micrograph movies can help a lot here, although it is up to the user to decide what is or isn't appropriate. Don't make them too long, though, as different filesystems handle excessively long filenames and paths in different (and sometimes unexpected) ways.
14. A text editor capable of column selection is extremely useful for this. We recommend no specific editor, but most Linux distributions should come with either `emacs` or `vim`, both of which may be used. For those preferring a GUI, there are many options.
15. Normally this can happen when datasets are acquired at different times, sometimes over a period of a few months. A separate gain reference means that the file sizes of micrograph movies are smaller as they are often saved in lower precision formats, but gain-corrected micrograph movies can be an advantage at times. A full treatment of this subject will spend considerable time discussing signed and unsigned integers, floating point, bit depth, and other computer-related technicalities.
16. For giant viruses, which include just a few particles in each micrograph, we found whole frame motion correction to be better than patch correction. We attribute this to the size of the giant virus particles relative to the size of the micrograph. For smaller subjects, there is really no reason not to use patch correction.

17. Cryo-HVEM currently does not have calculated dose-response curves, therefore dose weighting is not available for cryo-HVEM data [23].
18. The Gatan K3 and Direct Electron DE-64 can both demonstrate troublesome gain drift, although there are no direct detectors that do not demonstrate gain drift. For this reason, fresh gain and dark references are usually collected immediately prior to each period of acquisition.
19. For years, the go-to program for CTF estimation has been CTFFIND [6], which has received regular bug fixes and improvements, and is used by cisTEM [24] and RELION [25]. It is CPU based. A GPU-based equivalent, Gctf [37], was released in 2016; however, it is not open source and has received no updates for several years, which sadly means it is becoming more difficult to run, as it depends on CUDA libraries which are updated and may not maintain full backwards compatibility. CryoSPARC [26], on the other hand, has both its own CTF estimation and is also able to use CTFFIND or Gctf if desired.
20. We normally use a cut-off for estimated resolution in the 8–12 Å range. This can help (particularly with giant virus data) to avoid throwing away particles that are perfectly usable once some electro-optical parameter refinement has been carried out.
21. With good reason: the low signal-to-noise ratio of cryo-EM can make seeing particles difficult, which can lead to spurious results [51]. Many checks for problems of this type are now built in [73].
22. There are a few public datasets that are this perfectly clean; however, the challenges of obtaining perfect grids for cryo-EM data acquisition are many and varied.
23. The Gaussian blob picking in cisTEM is interactive, and the user can adjust settings controlling the size of the blob, threshold criterion, and maximum resolution used, while observing the effect directly on a micrograph. RELION cannot: settings must be chosen, run, and examined, and adjusted and run again. Non-optimized parameters can manifest themselves as mis-centered particles, excessive selection of junk, or missing good particles.
24. This is not as onerous as it may seem at first glance. The `relion_display` program allows sorting by X and Y coordinates, so particles can be sorted by X-coordinate and particle boxes showing the “barcode edge” can be deselected before the selected particles are again sorted. This time, after sorting by Y-coordinate, the same “barcode edge” particle boxes can be deselected.

25. Box sizes need to be an even number for 2D Fourier transformation.
26. We could have downsampled further, but we normally use 2 to $4\times$ downsampling so $5\times$ seemed adequate. Early processing was still acceptably fast. As a rule of thumb, regardless of original size, we will not downscale normal data beyond ~ 4 Å/pixel. For giant viruses, the rule of thumb is ignored, and the boxes are downscaled to a size where processing can be done more quickly. For example, Melbournevirus was down-scaled from 1440-pixel boxes to 360-pixel boxes, resulting in 8.84 Å/pixel [40].
27. We use this option if the dataset is homogeneous and contains more than $\sim 200,000$ particles. If a dataset is heterogeneous, we would not use it. If a dataset is very large ($>1,000,000$ particles) and heterogeneous, we have had success with splitting the particle sets into different randomized subsets and running multiple 2D classifications, then recombining the selected classes before proceeding.
28. It is also required in cryoSPARC, although cryoSPARC can generate a reference “on the fly” if necessary. Regardless, a good initial model is a requirement.
29. It can still be easier to have a self-consistent initial model, generated from the data being processed.
30. The different processing suites use different methods for initial model generation. We have previously found the cisTEM implementation of initial model generation to be more likely to give a “good” initial model, although problems can still occur, particularly with datasets with a strong orientation preference and no symmetry. Initial model generation has no built-in protection from overfitting (overestimation of the resolution and fitting to noise), so the resolution of initial models should be limited to ~ 8 Å. If an initial model looks “spiky,” it is probably overfitted and should be discarded to generate another—possibly with lower start and end resolutions.
31. Figure 16 highlights the section which needs to be removed. As particle stack .star files can often be hundreds of megabytes in size, opening the file in an editor may be extremely slow. A quicker method uses two ubiquitous Linux programs – head and tail. head should be used to identify where the optics table ends and the particle table begins (with RELION 3.1, you want to look for the second occurrence of “# version 30001” and count how many lines there are above that). The number of lines will vary depending on the number of optics groups you have defined and whether magnification anisotropy correction and beam tilt estimation have been carried out. When you know the number of lines, for example,

```

particles.star (~/data/fast2/10337/Extract/job013)
File Edit View Search Tools Documents Help
particles.star x
1
2 # version 30001
3
4 data_optics
5
6 loop
7 _rlnOpticsGroupName #1
8 _rlnOpticsGroup #2
9 _rlnMicrographOriginalPixelSize #3
10 _rlnVoltage #4
11 _rlnSphericalAberration #5
12 _rlnAmplitudeContrast #6
13 _rlnImagePixelSize #7
14 _rlnImageSize #8
15 _rlnImageDimensionality #9
16 opticsGroup1 1 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
17 opticsGroup2 2 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
18 opticsGroup3 3 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
19 opticsGroup4 4 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
20 opticsGroup5 5 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
21 opticsGroup6 6 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
22 opticsGroup7 7 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
23 opticsGroup8 8 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
24 opticsGroup9 9 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
25 opticsGroupA 10 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
26 opticsGroupB 11 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
27 opticsGroupC 12 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
28 opticsGroupD 13 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
29 opticsGroupE 14 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
30 opticsGroupF 15 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
31 opticsGroupG 16 0.281000 200.000000 2.700000 0.100000 1.405000 144 2
32
33
34 # version 30001
35
36 data_particles
37
38 loop
39 _rlnCoordinateX #1
40 _rlnCoordinateY #2
41 _rlnClassNumber #3
42 _rlnAutopickFigureOfMerit #4
43 _rlnAnglePsi #5
44 _rlnImageName #6
45 _rlnMicrographName #7
46 _rlnOpticsGroup #8
47 _rlnCtfMaxResolution #9
48 _rlnCtfFigureOfMerit #10
49 _rlnDefocusU #11
50 _rlnDefocusV #12
51 _rlnDefocusAngle #13
52 _rlnCtfBfactor #14
53 _rlnCtfScalefactor #15
54 _rlnPhaseShift #16
55 1535.200063 1215.366716 1 8.882123 250.000000 000001@Extract/job013/Micrographs/-
18oct04b_00077h16_00004edhiiiframes.mrcs MotionCorr/job002/Micrographs/-
18oct04b_00077h16_00004edhiiiframes.mrc 1 2.834435 0.281164 13337.375000 13221.450195

```

Fig. 16 A RELION 3.1 run_data.star file opened in Xed. The green highlighted section, defining optics parameters, must be deleted before cisTEM can import data processed by RELION 3.1

25 lines, run the following command: `tail -n +25 particles.star > particles_noOptics.star` to create a copy of the particles.star file without the optics table, which can be used for importing into cisTEM. The + sign is very important, without it, `tail` will only output the last 25 lines!

32. This is dependent on the resolution of the 3D reconstruction. Without prior knowledge, at resolutions worse than approximately 4 Å it is still not possible to determine handedness.
33. Unless you are unlucky enough to be working on something where mixed handedness minima may happen (Fig. 10).

34. The diagnostic plots output by RELION as PDF files for defocus are particularly useful for visualizing ice thickness or micrograph tilt and can also show if any potentially low-quality particles are still hiding in the dataset.
35. You may notice that we did not discuss how the electro-optical parameter refinements were checked between each step. This is because there are two options: direct reconstruction of the half-maps with the refined parameters, or a full re-refinement. The first is quicker, the second may be more reliable if adjustment of the parameters means that the particles can be aligned better, and in the process increase resolution. We favor the second as we think allowing particles to potentially optimize their alignment is safer, although we have used the first without any notable loss in “performance” of the dataset. For now, it largely comes down to how fast something needs to be completed, or whether we need it to “take care of itself.” It is easier to “set and forget” multiple rounds of parameter refinement cycling with 3D refinement cycling with the RELION pipeline [45, 54, 74], while calling “`relion_reconstruct`” requires user interaction or scripting outside of RELION.
36. Having an extra terminal window running and examining `.star` files as the run proceeds is a little less flashy than graphical output, but no less informative.
37. This is not always possible for viruses, which will probably contain a disordered viral genome. That said, careful thresholding of the map in UCSF Chimera or similar will often permit the user to obscure the low-resolution disordered regions while still containing all the ordered regions. A larger extension (e.g., 8–15 pixels rather than 4–8 pixels) of the initial threshold may be required to maintain an appropriately soft mask, however. The soft edge may also need to be adjusted.
38. This was very confusing the first time we saw it happen. Now we just ignore it and fix in post processing. It happened with the EMPIAR 10337 reprocessing carried out for this text; we manually checked the handedness of the initial model against PDB 2CIH [75], but after 3D classification the handedness was inverted.
39. You don’t need to define “true” and “false” for `reverse_-curvature`, for example. Simply using the flag implies “true” while not using it implies “false.” We just find it a little easier to track in a long script by being explicit.
40. We, and we are sure others, have also tested the utility of block-based reconstruction on smaller particles. We have seen no improvement on smaller icosahedral particles tested (and a large increase in processing time!) but the results for giant viruses are unquestionable.

41. There is no reason why it would not work on lower symmetry particles as well, but more blocks would have to be reconstructed and relative particle counts would also be decreased, which in turn would impact attainable resolution. As such, it may not actually improve the reconstruction.
42. You will also need UCSF Chimera to make block mask generation a lot easier.
43. We initially used this method but have found the volume eraser method to be much faster and no less accurate, so now use that instead.
44. The “Abort now” option in the RELION GUI can be used but can be slow to respond. A quicker, but less graceful method is to “killall relion_particle_subtract_mpi” in a new terminal. The subtract jobs can then be manually marked as successful (because what we needed from them was the shift calculations, which we now have) or as failed (because they did not run to completion).

Acknowledgments

We thank Kenta Okamoto, Martin Svenda, and Chantal Abergel for giving us the opportunity to process the Melbournevirus data. This study was supported by MEXT KAKENHI (JP19H04845 to K. M.), the Joint Research Program of ExCELLS (20-004 to K.M.), and the Cooperative Study Program of NIPS.

References

1. von Borries B, Ruska E, Ruska H (1938) Bakterien und Virus in Übermikroskopischer Aufnahme. *Klinische Wochenschrift* 17:921–925
2. Kausche GA, Pfankuch E, Ruska H (1939) Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop. *Naturwissenschaften* 27: 292–299
3. Ruska H (1940) Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop. *Naturwissenschaften* 28:45–46
4. Adrian M, Dubochet J, Lepault J, McDowall AW (1984) Cryo-electron microscopy of viruses. *Nature* 308:32–36
5. McMullan G, Faruqi AR, Henderson R (2016) Direct electron detectors. *Methods Enzymol* 579:1–17
6. Rohou A, Grigorieff N (2015) CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192:216–221
7. Ruskin RS, Yu Z, Grigorieff N (2013) Quantitative characterization of electron detectors for transmission electron microscopy. *J Struct Biol* 184:385–393
8. Kühlbrandt W (2014) Biochemistry. The resolution revolution. *Science* 343:1443–1444
9. Nakane T, Kotecha A, Sente A, McMullan G, Masilis S, Brown P, Grigoras IT, Malinauskaitė L, Malinauskas T, Miehling J, Uchanski T, Yu L, Karia D, Pechnikova EV, de Jong E, Keizer J, Bischoff M, McCormack J, Tiemeijer P, Hardwick SW, Chirgadze DY, Murshudov G, Aricescu AR, Scheres SHW (2020) Single-particle cryo-EM at atomic resolution. *Nature* 587:152–156
10. Yip KM, Fischer N, Paknia E, Chari A, Stark H (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature* 587:157–161
11. Frank J (2006) Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their

- native state, 2nd edn. Oxford University Press, Oxford/New York
12. Frank J (2006) Electron tomography methods for three-dimensional visualization of structures in the cell, 2nd edn. Springer, New York
 13. Wan W, Briggs JAG (2016) Chapter thirteen – cryo-electron tomography and subtomogram averaging. In: Crowther RA (ed) Methods in enzymology, vol 579. Academic Press, pp 329–367
 14. Lyumkis D (2019) Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem* 294:5181–5197
 15. Penczek P, Marko M, Buttle K, Frank J (1995) Double-tilt electron tomography. *Ultramicroscopy* 60:393–410
 16. Tegunov D, Xue L, Dienemann C, Cramer P, Mahamid J (2021) Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* 18:186–193
 17. Murata K, Zhang Q, Gerardo Galaz-Montoya J, Fu C, Coleman ML, Osburne MS, Schmid MF, Sullivan MB, Chisholm SW, Chiu W (2017) Visualizing adsorption of cyanophage P-SSP7 onto marine Prochlorococcus. *Sci Rep* 7:44176
 18. Liu X, Zhang Q, Murata K, Baker ML, Sullivan MB, Fu C, Dougherty MT, Schmid MF, Osburne MS, Chisholm SW, Chiu W (2010) Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus. *Nat Struct Mol Biol* 17:830–836
 19. Tanaka N, Fujita T, Takahashi Y, Yamasaki J, Murata K, Arai S (2020) Progress in environmental high-voltage transmission electron microscopy for nanomaterials. *Philos Trans A Math Phys Eng Sci* 378:20190602
 20. Murata K, Kaneko Y (2018) Visualization of DNA compaction in cyanobacteria by high-voltage cryo-electron tomography. *J Vis Exp* 137:e57197
 21. Murata K, Hagiwara S, Kimori Y, Kaneko Y (2016) Ultrastructure of compacted DNA in cyanobacteria by high-voltage cryo-electron tomography. *Sci Rep* 6:34934
 22. Okamoto K, Miyazaki N, Song C, Maia F, Reddy HKN, Abergel C, Claverie JM, Hajdu J, Svenda M, Murata K (2017) Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-tomography and energy-filtered electron cryo-microscopy. *Sci Rep* 7: 13291
 23. Chihara A, Burton-Smith RN, Kajimura N, Mitsuoka K, Okamoto K, Song C, Murata K (2022) A novel capsid protein network allows the characteristic inner membrane structure of *Marseilleviridae* giant viruses. *Sci Rep* 12: 21428
 24. Grant T, Rohou A, Grigorieff N (2018) *cistem*, user-friendly software for single-particle image processing. *Elife* 7:e35383
 25. Scheres SHW (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180:519–530
 26. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14:290–296
 27. Heymann JB, Belnap DM (2007) Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* 157:3–18
 28. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 157:38–46
 29. Moriya T, Saur M, Stabrin M, Merino F, Voicu H, Huang Z, Penczek PA, Raunser S, Gatsogiannis C (2017) High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *J Vis Exp* 123:e55448
 30. de la Rosa-Trevín JM, Quintana A, Del Cano L, Zaldívar A, Fochi I, Gutiérrez J, Gómez-Blanco J, Burguet-Castell J, Cuena-Alba J, Abrishami V, Vargas J, Otón J, Sharov G, Vilas JL, Navas J, Conesa P, Kazemi M, Marabini R, Sorzano CO, Carazo JM (2016) Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* 195:93–99
 31. Tegunov D, Cramer P (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* 16:1146–1152
 32. Grigorieff N (2016) Frealign: an exploratory tool for single-particle cryo-EM. *Methods Enzymol* 579:191–226
 33. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. *J Struct Biol* 116:17–24
 34. Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladadji M, Leith A (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116:190–199
 35. Grant T, Grigorieff N (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* 4:e06980
 36. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA (2017) MotionCor2: anisotropic correction of beam-induced

- motion for improved cryo-electron microscopy. *Nat Methods* 14:331–332
37. Zhang K (2016) Gctf: real-time CTF determination and correction. *J Struct Biol* 193:1–12
 38. Iudin A, Korir PK, Salavert-Torres J, Kleywegt GJ, Patwardhan A (2016) EMPIAR: a public archive for raw electron microscopy image data. *Nat Methods* 13:387–388
 39. Wu M, Lander GC, Herzik MA Jr (2020) Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200 keV. *J Struct Biol X* 4:100020
 40. Burton-Smith RN, Narayana Reddy HK, Svenda M, Abergel C, Okamoto K, Murata K (2021) The 4.4 Å structure of the giant Melbournevirus virion belonging to the *Marseillaviridae* family. *bioRxiv*:2021.2007.2014.452405
 41. Russo CJ, Passmore LA (2014) Electron microscopy: ultrastable gold substrates for electron cryomicroscopy. *Science* 346:1377–1380
 42. Carragher B, Kisseberth N, Kriegman D, Milligan RA, Potter CS, Pulokas J, Reilein A (2000) Leginon: an automated system for acquisition of images from vitreous ice specimens. *J Struct Biol* 132:33–45
 43. Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* 152:36–51
 44. Cheng A, Eng ET, Alink L, Rice WJ, Jordan KD, Kim LY, Potter CS, Carragher B (2018) High resolution single particle cryo-electron microscopy using beam-image shift. *J Struct Biol* 204:270–275
 45. Zivanov J, Nakane T, Scheres SHW (2020) Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* 7:253–267
 46. Kimanius D, Dong L, Sharov G, Nakane T, Scheres SHW (2021) New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* 478:4169–4185
 47. Li X, Mooney P, Zheng S, Booth CR, Braufeld MB, Gubbens S, Agard DA, Cheng Y (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* 10:584
 48. Thon F (1966) Notizen: Zur Defokussierungssabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Zeitschrift für Naturforschung A* 21:476–478
 49. Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, Hagel P, Sitsel O, Raisch T, Prumbaum D, Quentin D, Roderer D, Tacke S, Siebolds B, Schubert E, Shaikh TR, Lill P, Gatsogiannis C, Raunser S (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* 2:218
 50. Lawson CL, Baker ML, Best C, Bi C, Dougherty M, Feng P, van Ginkel G, Devkota B, Lagerstedt I, Ludtke SJ, Newman RH, Oldfield TJ, Rees I, Sahni G, Sala R, Velankar S, Warren J, Westbrook JD, Henrick K, Kleywegt GJ, Berman HM, Chiu W (2011) EMDDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res* 39: D456–D464
 51. Henderson R (2013) Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc Natl Acad Sci USA* 110: 18037–18041
 52. Scheres SH, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Methods* 9(9):853–854
 53. Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N (2012) Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol* 177:630–637
 54. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7:e42166
 55. Wolf M, DeRosier DJ, Grigorieff N (2006) Ewald sphere correction for single-particle electron microscopy. *Ultramicroscopy* 106: 376–382
 56. Russo CJ, Henderson R (2018) Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy* 187:26–33
 57. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
 58. Goddard TD, Huang CC, Ferrin TE (2007) Visualizing density maps with UCSF Chimera. *J Struct Biol* 157:281–287
 59. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 27:14–25
 60. Kucukelbir A, Sigworth FJ, Tagare HD (2014) Quantifying the local resolution of cryo-EM density maps. *Nat Methods* 11:63–65
 61. Vilas JL, Gómez-Blanco J, Conesa P, Melero R, Miguel de la Rosa-Trevín J, Otón J, Cuenca J, Marabini R, Carazo JM, Vargas J, Sorzano

- COS (2018) MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps. *Structure* 26:337–344 e334
62. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132
 63. Casañal A, Lohkamp B, Emsley P (2019) Current developments in coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci* 29: 1069–1078
 64. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221
 65. Croll TI (2018) ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* 74:519–530
 66. Zhu D, Wang X, Fang Q, Van Etten JL, Rossmann MG, Rao Z, Zhang X (2018) Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. *Nat Commun* 9:1552
 67. Fang Q, Zhu D, Agarkova I, Adhikari J, Klose T, Liu Y, Chen Z, Sun Y, Gross ML, Van Etten JL, Zhang X, Rossmann MG (2019) Near-atomic structure of a giant virus. *Nat Commun* 10:388
 68. Liu S, Luo Y, Wang Y, Li S, Zhao Z, Bi Y, Sun J, Peng R, Song H, Zhu D, Sun Y, Li S, Zhang L, Wang W, Sun Y, Qi J, Yan J, Shi Y, Zhang X, Wang P, Qiu HJ, Gao GF (2019) Cryo-EM structure of the African swine fever virus. *Cell Host Microbe* 26:836–843 e833
 69. Wang N, Zhao D, Wang J, Zhang Y, Wang M, Gao Y, Li F, Wang J, Bu Z, Rao Z, Wang X (2019) Architecture of African swine fever virus and implications for viral assembly. *Science* 366:640–644
 70. Pintilie G, Chiu W (2012) Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in cryo-EM density maps. *Biopolymers* 97: 742–760
 71. Bhella D (2019) Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophys Rev* 11:515–519
 72. Burton-Smith RN, Murata K (2021) Cryo-electron microscopy of the giant viruses. *Microscopy (Oxford)* 70:477–486
 73. Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SH, Henderson R (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135:24–35
 74. Fernandez-Leiro R, Scheres SHW (2017) A pipeline approach to single-particle processing in RELION. *Acta Crystallogr D Struct Biol* 73: 496–502
 75. Toussaint L, Bertrand L, Hue L, Crichton RR, Declercq JP (2007) High-resolution X-ray structures of human apoferritin H-chain mutants correlated with their activity and metal-binding sites. *J Mol Biol* 365:440–452



Chapter 12

Time-Resolved Small-Angle X-Ray Scattering of Protein Cage Assembly

Daisuke Sato, Takaaki Hikima, and Masamichi Ikeguchi

Abstract

Recent improvements in X-ray detectors and synchrotron light sources have made it possible to measure time-resolved small-angle X-ray scattering (TR-SAXS) at millisecond time resolution. As an example, in this chapter we describe the beamline setup, experimental scheme, and the points that should be noted in stopped-flow TR-SAXS experiments for investigating the ferritin assembly reaction.

Key words Exposure time, Ferritin, Radiation damage, SAXS, Stopped-flow

1 Introduction

Small-angle X-ray scattering (SAXS) is a versatile technique that provides low-resolution structural information of macromolecules or their complexes. Although the spatial resolution of a SAXS-derived structure is worse than that of an X-ray crystal structure, SAXS can be applied to kinetic studies to follow structural change, such as protein folding and assembly. The reaction initiated by stopped-flow mixing and followed by SAXS was reported as early as the 1980s [1–3]. However, the temporal resolution of time-resolved SAXS (TR-SAXS) has been limited by the readout time of detectors (an annular or linear position-sensitive proportional counter was employed). Although the temporal resolution has been improved by the introduction of charge-coupled device-based detectors, which have been used since the 2000s [4, 5], its readout time is still longer than 50 ms. To overcome this difficulty, a continuous flow setup has been used to follow rapid reactions [6–8]. In the continuous flow setup, data at different time points of a reaction are obtained by measuring SAXS at different distances from the mixing point. With this setup, the readout time of detectors does not limit the time resolution. While the continuous flow makes it possible to observe sub-millisecond reactions, it is difficult

to observe reactions slower than 100 ms. In many kinetic studies, therefore, the continuous flow experiment has been coupled with a stopped-flow experiment [6, 7]. The development of the photon-counting PILATUS detector has made it possible to acquire two-dimensional scattering images every few milliseconds [9]. In addition, high photon fluxes at synchrotron light sources lead to improved signal-to-noise ratios for data acquired within a short exposure time. As a result, TR-SAXS has become the best technique to study the assembly reactions of protein cages such as ferritin [10, 11] or virus capsids [12, 13]. It should be noted that the high photon fluxes also bring about damage to the protein sample. In this chapter, we describe the beamline setup, experimental scheme, and the points that should be noted in the stopped-flow TR-SAXS experiments in high-brilliance synchrotrons.

2 Materials

To obtain interpretable TR-SAXS data, it is desirable that the protein cage assembly is fully reversible (*see Note 1*). Previously, we observed the time-dependent change of SAXS accompanied by the assembly reaction of acid-dissociated *Escherichia coli* ferritin (EcFtnA) [10, 11]. At acidic pH, EcFtnA is known to dissociate to the dimer while maintaining the subunit tertiary structure and can fully reassociate to the 24-meric cage structure at neutral pH [10, 14]. The protein was overexpressed in *E. coli* and purified as described [14]. Here, we describe the procedures of sample preparation after protein purification.

2.1 Preparation of Solutions

1. Prepare a EcFtnA stock solution, with a protein concentration that is higher than 10 mg/mL in 20 mM phosphate buffer (pH 7.0).
2. Dilute the protein stock to a series of solutions of which protein concentrations are 2–10 mg/mL.
3. Dissociate EcFtnA 24-mer by mixing the protein solution with 0.5 M phosphoric acid in a ratio of 15:1, resulting in 50 mM sodium phosphate, pH 2.5 (dissociated EcFtnA solution).
4. Prepare a pH-up solution containing 50 mM Tris, 2 mM EDTA, and appropriate amount of NaOH.
5. Prepare the “assembly buffer,” which is 25 mM Tris, 25 mM phosphate, 1 mM EDTA (pH 8.0).
6. For acquiring the SAXS of the fully assembled EcFtnA under assembly conditions, prepare the EcFtnA solution in 25 mM Tris, 25 mM phosphate, 1 mM EDTA (pH 8.0).
7. Degas all solutions, except for the protein stock solution to prevent bubble formation on stopped-flow mixing.

2.2 Instruments (Beamline Setup)

Experiments were performed at the SPring-8 beamline 45XU (*see Note 2*) [15].

1. Two-dimensional scattering images were obtained using a photon-counting detector (PILATUS300k-w or PILATUS3X 2M; Dectris, Switzerland) with a readout time of 2.3 or 0.95 ms, respectively.
2. Set the wavelength of the X-ray and the sample-to-detector distance. In our previous experiments, these were 1.0 Å and 1.5 m, respectively. The sample-to-detector distance depends on the size of the protein cage of the assembly being measured and the size of the detectors.
3. Place a stopped-flow apparatus (Unisoku Co. Ltd., Osaka, Japan) equipped with a quartz capillary cell with a diameter of 2 mm located just above the mixer. Align the apparatus so that the X-ray beam passes through the capillary center.
4. The temperature of the stopped-flow apparatus is kept constant by water circulating from a thermostat bath.
5. Determine the accurate sample-to-detector distance and the beam center by measuring the scattering of silver behenate [16].
6. A Uniblitz XRS6 shutter system (Vincent Associates, NY, USA) is placed upstream from the sample and controlled as described below.

3 Methods

3.1 Experimental Procedures

1. Acquire the scattering from the buffer under standard reassembly conditions (25 mM Tris, 25 mM phosphate, 1 mM EDTA (pH 8.0)). To do this, fill the stopped-flow reservoirs with 50 mM phosphate buffer (pH 2.5) and the pH-up solution (*see above*).
2. Trigger the stopped-flow mixing. The stopped-flow apparatus we used mixes two reservoir solutions in a ratio of 1:1, to give standard reassembly conditions (25 mM Tris, 25 mM phosphate, and 1 mM EDTA, pH 8.0). After several stopped-flow mixings (*see Note 3*), acquire the scattering data during the acquisition period (10 ms as a standard) repeatedly (for example, 100 times).
3. Replace 50 mM phosphate buffer in a reservoir with the dissociated EcFtnA solution and trigger the stopped-flow mixing several times.
4. Trigger the stopped-flow mixing. Data acquisition is triggered by stopping the stopped-flow piston (*see Fig. 1*). The scattering

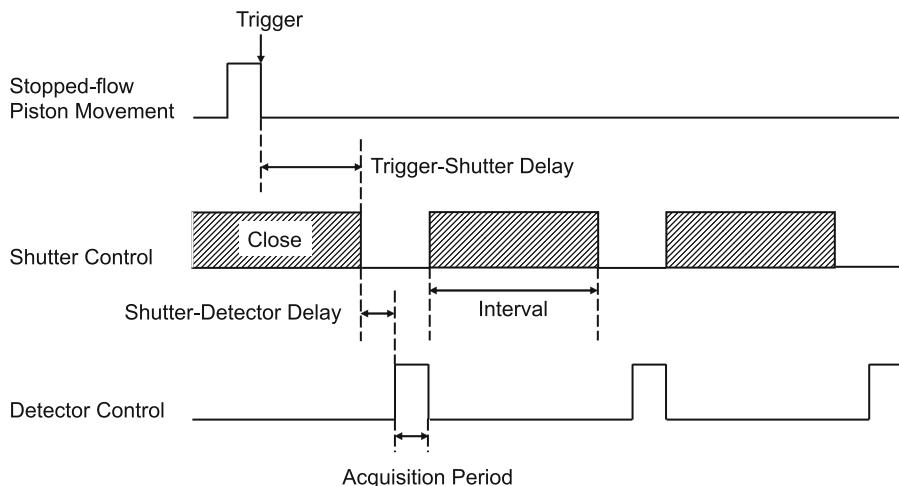


Fig. 1 Experimental sequence. Actions of a stopped-flow device, a shutter, and a detector are schematically shown

images during the acquisition period (10 ms as a standard) were corrected as time-point data. For each stopped-flow mixing, 10–30 points were acquired by setting various intervals. The data acquired at these intervals were combined to obtain a time course of the assembly reaction. When the interval was longer than 30 ms, the shutter was closed just after each data acquisition period and then reopened before the next data acquisition. The time delay between the shutter trigger and data acquisition (shutter-detector delay) was set at 10 ms to ensure the shutter was open. In some cases, the delay time was set between the mixing trigger and the shutter-open mode (trigger-shutter delay) to obtain data at a late reaction time without radiation damage to the sample (*see Note 4*).

5. To observe the scattering of fully reassembled EcFtnA, fill the two reservoirs with the reassembly buffer and wash the flow line of the stopped-flow apparatus.
6. Replace the reassembly buffer in one of reservoirs with the EcFtnA solution in the reassembly buffer and trigger the stopped-flow mixing several times.
7. Acquire scattering data during the acquisition period (10 ms as standard) repeatedly (for example, 100 times).

3.2 Data Processing

1. Because solution scattering is isotropic for all radial directions, two-dimensional images were azimuthally averaged using FIT2D [17], to obtain one-dimensional scattering profiles.
2. After subtraction of the scattering profile of the buffer solution, the apparent radius of gyration, R_{app} , and the forward

scattering intensity, $I(0)$, were calculated using the Guinier approximation [18]:

$$I(Q) = I(0) \exp\left(-\frac{R_{\text{app}}^2 Q^2}{3}\right), \quad (1)$$

where the scattering vector $Q = (4\pi/\lambda) \sin \theta$ (λ is the wavelength, and 2θ is the scattering angle), and $I(Q)$ is the scattering intensity at a given Q value. The $I(0)$ and R_{app} values were calculated from the intercept and slope of the Guinier plot in the Q range satisfying $QR_{\text{app}} \leq 1.3$. For the analysis of kinetic data, we used a wider Q range satisfying $QR_{\text{app}} \leq 1.8$ because the narrow Q range produced a large error in estimated $I(0)$ and R_{app} values. This analysis was performed automatically using procedures that were developed using Igor Pro (WaveMetrics, Inc., OR, USA).

3. In a system where a single particle is present in a solution, R_{app} is the radius of gyration (R_g) of that particle, and $I(0)$ is proportional to the mass concentration, the molecular weight, the beam intensity, and the square of the difference in the electron density between the particle and the solvent [19]. In a system where two or more particles are present in a solution, $I(0)$ and R_{app} are complex. $I(0)$ is an arithmetic average of the values obtained for the scattering bodies present in the solution. R_{app} is not a simple population average of a scattering body's R_g ; rather, it is biased toward the values of the larger scattering bodies. Assuming that all intermediate oligomers are spherical, the observed $I(0)$ and R_{app} are expressed by the following equations:

$$I(0) = \sum_i C_i I(0)_i, \quad (2)$$

and

$$R_{\text{app}} = \sqrt{\sum_i C_i I(0)_i R_{g,i}^2 / \sum_i C_i I(0)_i}, \quad (3)$$

where C_i , $I(0)_i$, and $R_{g,i}$ denote the mass concentration, the forward scattering intensity (per unit mass concentration), and the radius of gyration of i th species, respectively.

4 Notes

1. For obtaining interpretable TR-SAXS data, it is desirable that the protein cage assembly is fully reversible. Mammalian ferritins dissociate at acidic pH, and their subunit tertiary structures

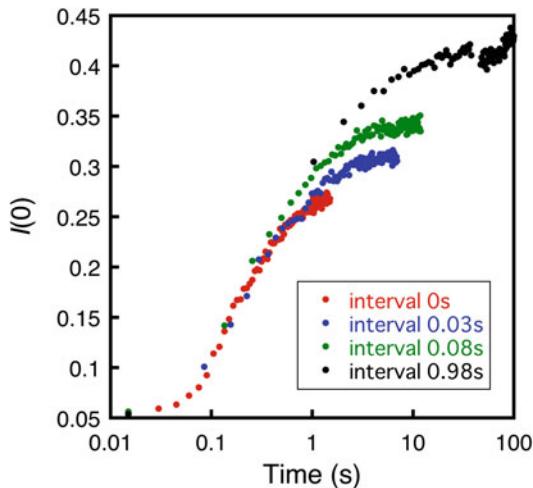


Fig. 2 $I(0)$ increments observed during EcFtnA assembly reactions at pH 8 and 25 °C. Protein concentration was 4.8 mg/mL. The same reaction was observed at different intervals

are concomitantly disrupted. When the acidified mammalian ferritin solution is neutralized, amorphous aggregates are formed during the refolding and reassembly reaction [20, 21]. The X-ray scattering from the larger amorphous aggregate predominates the observed scattering curve because it is stronger than the scattering from small, ordered structures. As a result, the interpretation of data becomes difficult.

2. Currently, SPring-8 beamline 45XU is not available for TR-SAXS. The experiments may be possible at other beamlines of similar or higher photon flux ($>1.0 \times 10^{12}$ photons/s).
3. Several shots (mixings) are necessary to replace the old solution in the route from the reservoir to the cell with the new solution.
4. Figure 2 shows a typical example of $I(0)$ increments observed after stopped-flow mixing with different intervals. In this experiment, the acquisition period and the trigger-shutter delay were set at 10 and 0 ms, respectively. The data acquisition was repeated 100 times. The $I(0)$ value increased with time, indicating the assembly of EcFtnA. However, the $I(0)$ values obtained at the same reaction time increased with increasing intervals (for example, see data around 1 s). Even at the same reaction time, the integrated radiation time was different when a different interval was used. In the case of zero interval, the upstream shutter was kept open since the stopped-flow trigger was on. At the reaction time of 1 s, the sample was exposed to X-rays for approximately 1 s. By contrast, in the case of the interval of 0.98 s, the exposure time was less than 40 ms at the

reaction time of 1.035 s. This difference in exposure time would cause different radiation damage to the sample. The damaged dimer or oligomer may lose the ability to assemble so that the $I(0)$ increment was depressed when the integrated exposure time was increased through repeated data acquisition. After this experiment, we reduced the number of data acquisition events to 10–30 times depending on the interval, and only data giving similar $I(0)$ values independent of the interval were adopted.

Acknowledgments

The synchrotron radiation SAXS experiments were performed at BL45XU of SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute, Hyogo, Japan (proposals 2011A1133, 2012A1217, 2012B1114, 2013B1392, 2015A1374, 2016B1217, 2017A1403, 2017B1308, 2018A1262, and 2018B1404). This research was supported in part by the Platform for Drug Discovery, Information, and Structural Life Science of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Fowler AG, Foote AM, Moody MF, Vachette P, Provencher SW, Gabriel A, Bordas J, Koch MHJ (1983) Stopped-flow solution scattering using synchrotron radiation: apparatus, data collection and data analysis. *J Biochem Biophys Methods* 7:317–329
2. Ueki T, Hiragi Y, Kataoka M, Inoko Y, Amemiya Y, Izumi Y, Tagawa H, Muroga Y (1985) Aggregation of bovine serum albumin upon cleavage of its disulfide bonds, studied by the time-resolved small-angle x-ray scattering technique with synchrotron radiation. *Biophys Chem* 23:115–124
3. Nagamura T, Kurita K, Tokikura E, Kihara H (1985) Stopped-flow X-ray scattering device with a slit-type mixer. *J Biochem Biophys Methods* 11:277–286
4. Schmölzer S, Gräßner D, Gradzielski M, Narayanan T (2002) Millisecond-range time-resolved small-angle X-ray scattering studies of micellar transformations. *Phys Rev Lett* 88:4
5. Arai M, Ito K, Inobe T, Nakao M, Maki K, Kamagata K, Kihara H, Amemiya Y, Kuwajima K (2002) Fast compaction of α -lactalbumin during folding studied by stopped-flow X-ray scattering. *J Mol Biol* 321:121–132
6. Segel DJ, Bachmann A, Hofrichter J, Hodgson KO, Doniach S, Kieffaber T (1999) Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J Mol Biol* 288:489–499
7. Akiyama S, Takahashi S, Kimura T, Ishimori K, Morishima I, Nishikawa Y, Fujisawa T (2002) Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc Natl Acad Sci U S A* 99:1329–1334
8. Kathuria SV, Guo L, Graceffa R, Barrea R, Nobrega RP, Matthews CR, Irving TC, Bilsel O (2011) Minireview: structural insights into early folding events using continuous-flow time-resolved small-angle X-ray scattering. *Biopolymers* 95:550–558
9. Kraft P, Bergamaschi A, Broennimann C, Dinapoli R, Eikenberry EF, Henrich B, Johnson I, Mozzanica A, Schleputz CM, Willmott PR, Schmitt B (2009) Performance of single-photon-counting PILATUS detector modules. *J Synchrotron Rad* 16:368–375
10. Sato D, Ohtomo H, Yamada Y, Hikima T, Kurobe A, Fujiwara K, Ikeguchi M (2016) Ferritin assembly revisited: a time-resolved small-

- angle X-ray scattering study. *Biochemistry* 55: 287–293
11. Sato D, Takebe S, Kurobe A, Ohtomo H, Fujiwara K, Ikeguchi M (2016) Electrostatic repulsion during ferritin assembly and its screening by ions. *Biochemistry* 55:482–488
 12. Asor R, Schlicksup CJ, Zhao Z, Zlotnick A, Raviv U (2020) Rapidly forming early intermediate structures dictate the pathway of capsid assembly. *J Am Chem Soc* 142:7868–7882
 13. Chevreuil M, Lecoq L, Wang S, Gargowitsch L, Nhiri N, Jacquet E, Zinn T, Fieulaine S, Bressanelli S, Tresset G (2020) Nonsymmetrical dynamics of the HBV capsid assembly and disassembly evidenced by their transient species. *J Phys Chem B* 124:9987–9995
 14. Ohtomo H, Ohtomo M, Sato D, Kurobe A, Sunato A, Matsumura Y, Khara H, Fujiwara K, Ikeguchi M (2015) A physicochemical and mutational analysis of intersubunit interactions of *Escherichia coli* ferritin A. *Biochemistry* 54: 6243–6251
 15. Fujisawa T, Inoue K, Oka T, Iwamoto H, Uruga T, Kumasaki T, Inoko Y, Yagi N, Yamamoto M, Ueki T (2000) Small-angle X-ray scattering station at the SPring-8 RIKEN beamline. *J Appl Crystallogr* 33: 797–800
 16. Huang TC, Toraya H, Blanton TN, Wu Y (1993) X-ray-powder diffraction analysis of silver behenate, a possible low-angle diffraction standard. *J Appl Crystallogr* 26:180–184
 17. Hammersley AP (2016) FIT2D: a multi-purpose data reduction, analysis and visualization program. *J Appl Crystallogr* 49:646–652
 18. Guinier A, Fournet G (1955) Small-angle scattering of X-rays. Wiley, New York
 19. Glatter O, Kratky O (eds) (1982) Small angle x-ray scattering. Academic Press, London/New York
 20. Kuwata T, Okada Y, Yamamoto T, Sato D, Fujiwara K, Fukumura T, Ikeguchi M (2019) Structure, function, folding, and aggregation of a neuroferritinopathy-related ferritin variant. *Biochemistry* 58:2318–2325
 21. Santambrogio P, Levi S, Arosio P, Palagi L, Vecchio G, Lawson DM, Yewdall SJ, Artymiuk PJ, Harrison PM, Jappelli R, Cesareni G (1992) Evidence that a salt bridge in the light chain contributes to the physical stability difference between heavy and light human ferritins. *J Biol Chem* 267:14077–14083



Chapter 13

Protein Cage-Stabilized Emulsions: Formulation and Characterization

Mridul Sarker, Samuel Watts, Stefan Salentinig, and Sierin Lim

Abstract

The formulation of Pickering emulsions using protein cages is gaining interest for applications in molecular delivery. Despite the growing interest, methods to investigate the at the liquid-liquid interface are limited. This chapter describes standard methods to formulate and protocols to characterize protein cage-stabilized emulsions. The characterization methods are dynamic light scattering (DLS), intrinsic fluorescence spectroscopy (TF), circular dichroism (CD), and small angle X-ray scattering (SAXS). Combining these methods allows understanding of the protein cage nanostructure at the oil/water interface.

Key words Emulsion, Formulation, DLS, Circular dichroism, Intrinsic fluorescence, SAXS

1 Introduction

Protein cages have found applications in medicine, particularly as molecular carriers where the molecules to be delivered are incorporated in or on the protein cages [1–3]. Recent work has expanded the use of protein cages as emulsion stabilizers for food and drug delivery applications [4, 5]. These particle-stabilized emulsions are referred to as Pickering emulsions [6–9]. The particles stabilize the liquid-liquid interface through their amphiphilic nature that is partly wetted by oil and water allowing them to adsorb at the interface. Particles preferentially wetted by water (contact angle < 90°) lead to oil-in-water emulsions [10]. An understanding of the protein cages’ structure, their potential conformational modification at the interface, as well as their wettability are crucial for informed designing of Pickering emulsions [11–14]. Amphiphilic proteins are surface-active as they contain a combination of hydrophilic and hydrophobic amino acids, allowing them to adsorb at oil-water interfaces which can lead to emulsion

Mridul Sarker and Samuel Watts contributed equally with all other contributors.

stabilization properties [15, 16]. The protein's adsorption to the interface will lead to modifications in the molecular interactions between the protein and its environment owing changes in the free energy landscape and potentially resulting in modifications in protein folding [17]. Since protein properties are related to its three-dimensional structure [18], conformation changes during adsorption can change their properties. Therefore, investigating the protein's structural transformations upon adsorption onto interfaces is critical for the design of advanced protein cage-stabilized emulsions.

Several techniques including dynamic light scattering (DLS), tryptophan fluorescence (TF), circular dichroism (CD), and small angle X-ray scattering (SAXS), have been used to study the molecular assembly and structure of protein cages in solution and at the oil-water interface [19–22]. In this context, we recently reported an investigation on conformational changes of recombinant protein with a cage-like structure at liquid-liquid interface and reviewed protein nanocage-stabilized emulsions for functional food materials [5, 19].

DLS measures the collective diffusion of the colloidal system and is mainly used for particle sizing in solution, covering the size range from micrometers to around 3 nm [23, 24]. The time-dependent fluctuations in scattered intensities resulting from the diffusion of the scattering centers (particles) are measured at a fixed scattering angle [25]. The Stokes-Einstein equation can then be used to relate the diffusion behaviour to the particle size. As particle interactions influence collective diffusion, caution must be applied to concentration effects. Further, the polydispersity of the particles and the impact of the particle shapes (the Stokes-Einstein equation assumes spherical geometry) must be considered [24, 26]. In colored samples, it is also crucial to check for the potential light absorption at the laser wavelength of the instrument that leads to diffusion from local heating with the laser and compromises the results.

TF is used to study the tertiary structure of the adsorbed protein. The fluorescence of tryptophan (Trp) is sensitive to the polarity of its local environment. As Trp's quantum yield depends on the solvent's polarity [27, 28]. In a lower polarity environment (e.g., within a folded protein in an aqueous environment), Trp has a high quantum yield hence a strong fluorescence signal. Whereas in a polar environment (e.g. on the surface of a folded protein in an aqueous environment) the fluorescence signal will be weaker [27, 28]. In addition to the modifications in intensity, the emission's maxima will depend on the polarity of the Trp's environment. A polar environment will lead to a red-shift when compared to the apolar environment [27]. These properties make TF a popular method to investigate protein folding into tertiary structures as well as protein denaturation [29, 30].

CD spectroscopy allows to probe rapidly the folding of proteins into their secondary structure as well as the binding properties of the protein [31]. CD measures the discrepancy in the absorbance of left- and right-hand circularly polarized light in function of its wavelength. The peptide bonds will contribute to such a signal in the far-UV region [32, 33]. The differences in orientation of the peptide bonds in the various protein secondary structures (alpha-helix, beta-sheets or random coil) will lead to distinct CD spectra with characteristic shape and magnitude [33]. Therefore, CD gives an insight on the secondary structures of proteins as well as their modification due to temperature, mutation, heat, denaturant, or binding interactions. Recently, CD allowed to characterize the conformational modifications of globular proteins when adsorbed to a liquid-liquid interface [34, 35].

SAXS is a non-destructive technique that provides information on the colloidal structure, including molecular weight, size, shape, and the interaction of proteins and protein cages in solution and on the oil/water interface [20, 21, 36–38]. It can further be used to study the conformation of proteins, complementing information from CD, TF, and Nuclear Magnetic Resonance (NMR) [39, 40]. SAXS provides information about the sample in the reciprocal (Fourier) space as an indirect method. The signal is a combination of intra-particle correlations (form factor) and inter-particle correlations (structure factor) [41]. Depending on the set-up (e.g., sample to detector distance, detector size, wavelength), SAXS mainly covers the size range between around 1 nm and hundreds of nanometers. The information is a statistical average over all structures in the sample. The electron density difference between the particles and the solvent defines the scattering intensity of the suspended particles and the minimum required particle concentration to achieve an acceptable signal-to-noise ratio. In general, SAXS is well suited to study particle concentrations from dilute solution systems with a typical lower limit of 0.01 mg/mL up to high concentrations, as well as powders. Combined with high-intensity synchrotron sources and appropriate microfluidics or stopped flow approaches, this technique can be used for *in situ* studies of modifications in the cage assembly [42, 43]. Inter-particle interactions must be considered in the analysis above a particular particle concentration threshold (around 10 mg/mL for non-charged particles) and additional interparticle interaction parameters can be extracted from the signal [41, 44]. These parameters include the effective volume fraction, the apparent interaction radius between particles, and their effective charge. The contribution of the structure factor to the scattering intensity and the accuracy of these parameters scale with the strength of interaction and particle concentration. Sophisticated data modeling approaches are necessary to extract the desired information from the experimental data [45, 46].

This chapter describes the protocol to formulate protein cage-stabilized Pickering emulsion using E2 protein cage and rosemary oil as an example, with the in-depth characterization of the protein cage structure at the liquid-liquid interface. E2 protein cage comprises 60 subunits and is derived from *Geobacillus stenorhizophilus* pyruvate dehydrogenase multienzyme complex [47]. It is the first protein cage used in Pickering emulsion system with in-depth characterization of its structure at the liquid-liquid interface [4, 19]. Other protein cages, such as ferritin, can be used for the formulation with careful optimization.

2 Materials

2.1 Pickering Emulsion Formulation

1. Oil (*see Note 1*).
2. Milli-Q water.
3. Beaker.
4. Protein cage (*see Note 2*).
5. Molecular weight cut-off (MWCO) filter (Amicon® Ultra-15 Centrifugal Filter Units, Merck).
6. Emulsifier (e.g., Ultrasonicator or mechanical homogenizer, or microfluidics) (*see Note 3*).

2.2 Dynamic Light Scattering (DLS)

1. DLS instrument.
2. Cuvette suited to the instrument's sample holder.
3. Milli-Q water or buffer.

2.3 Tryptophan Fluorescence (TF) Intensity

1. UV-Vis spectrophotometer (*see Note 4*).
2. Quartz cuvette.
3. Milli-Q water or an appropriate buffer that contains protein for baseline correction.

2.4 Circular Dichroism (CD)

1. Circular dichroism spectrophotometer to read far-UV spectra (e.g., 190–260 nm).
2. Quartz cell of 0.1 mm.
3. Sodium dodecyl sulfate (SDS).
4. Milli-Q water or buffer.

2.5 Small Angle X-Ray Scattering (SAXS)

1. SAXS camera aligned for the q -range of interest.
2. Glass or quartz capillaries with thin walls (<https://www.hilgenberg-gmbh.de/en/products/test-tubes/sample-tubes/>).
3. Milli-Q water.
4. Buffer for background measurement.

3 Methods

3.1 Emulsion Formulation

- To prepare emulsion, combine oil and water in a separate beaker. The ratio of oil and water is to be determined empirically (*see Note 5*).
- Measure the concentration of purified protein cage by Pierce™ BCA Protein Assay Kit (Thermo Scientific™) or Bradford protein assay (Bio-Rad).
- Calculate the mass of protein required to formulate emulsion by following the Protein Mass Fraction (PMF) equation [4].

$$= \frac{\text{Mass of Protein}}{\text{Mass of protein} + \text{Mass of water} + \text{Mass of Oil}}$$

$$= \frac{\text{Mass of Protein}}{\text{Mass of protein} + (\text{Density of water} \times \text{volume of water}) + (\text{Density of oil} \times \text{volume of Oil})}$$

- Concentrate the protein using a ultrafiltration device by centrifugation at $4815 \times g$ until the protein reaches the desired concentration (*see Notes 6 and 7*). Choose MWCO of the ultrafiltration device that is compatible with the protein. For protein cages, a typical MWCO of the ultrafiltration device is 100 kDa.
 - (Optional) This step is recommended if the protein is in a buffer. Replace the associated buffer with Milli-Q water by buffer exchange method using ultrafiltration [48] (*see Note 8*). This step can be merged with the **step 4** by discontinuous diafiltration method [49].
 - Repeat **step 2** to confirm the final concentration of protein.
 - Take a beaker of the expected total volume of emulsion and place the beaker on a magnetic stirrer to stir the mixture of oil and water slowly ($50\text{--}100 \times g$).
 - Prepare a coarse Pickering emulsion in a beaker by adding the concentrated protein cages to the mixture drop by drop while continuously stirring. For example, to prepare 3 mL of Pickering emulsion add 2.7 mL of concentrated protein into 0.3 mL of oil or add 1.0 mL of concentrated protein into 2 mL of oil-water mixture.
- Alternatively, for small volume, formulate a coarse emulsion by mixing concentrated protein solution with oil-water mixture by pipetting vigorously.
- Homogenize the coarse emulsion by ultrasonication using an ultrasonicator (500 W, 20 kHz) at 40% amplitude. Energy input was given through a stepped microtip 1/8 in. containing a piezoelectric crystal with a probe diameter of 3 mm



Fig. 1 Physical appearance of Pickering emulsion, stabilized by recombinant protein cages. (Adapted from Sarker et al. [19], © 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim)

(see Note 9). Alternatively, emulsification may be performed with a mechanical homogenizer, microfluidics, or membrane emulsification. Figure 1 shows a physical appearance of a homogenized and stable Pickering emulsion stabilized by protein cages.

3.2 Droplet Size Measurement by DLS

1. Sample considerations: For particle sizing, DLS works on the assumption of a dilute system [23]. Hence, the sample's concentration needs to be low enough to avoid multiple-scattering effects of light between particles and inter-particle interactions on the collective diffusion [50]. Therefore, it is good practice to measure DLS at multiple concentrations to determine the limit where the dependence of the particle size on concentration can be neglected. Moreover, the scattered intensities will scale to the 6th power of the particle's radius as well as to the contrast in refractive index between the particles and the buffer [51]. A good starting point to find the ideal concentration is between 1 and 10 mg/mL.
2. The operating software of the DLS instrument will allow inputting the droplet's refractive index, solvent viscosity and refractive index, and the temperature if there is no built-in thermometer. These values will be later used in the Stokes-Einstein equation to compute the droplet's hydrodynamic radius (R_h). The R_h of a given particle is the radius a hard spherical particle would have if it diffused at the same speed as that particle [52].

3. The DLS instrument will measure the fluctuations in scattering intensities. These fluctuations are linked to the motion of the particles in the scattering volume. The intensity autocorrelation function (ACF) is computed from them. This operation has been described in detail in other work [23, 24, 26]. The ACF is then further analyzed to obtain the collective diffusion coefficient and particle size using methods described in **step 5**.
4. The ACF contains essential information about the particles' measurement quality and size characteristics (e.g., modality of the distribution, polydispersity). The intercept of the ACF with the y-axis should be one from theory, representing full correlation at short correlation times. However, this intercept is usually slightly below 1 owing to experimental restrictions. As a rule of thumb, an intercept above 80% of that of a calibration standard can be considered a "good" measurement [26]. A too low intercept may indicate the lack of scattering events, the incoherence of the scattering, or multiple scattering events all of which lead to poor statistics [25]. Hence, if the intercept is lower than 80%, the experimental conditions (sample concentration, measurement time, etc.) need to be improved.
5. The ACF is fitted to obtain decay rate (Γ). The ACF of polydisperse samples is a sum of many differential exponential terms. There are two methods to extract the decay rate from the ACF: (i) The indirect Laplace transformation: several routines are available for this numerical analysis of the ACF, including the CONTIN package [53–55]. This method may be built into the instrument's software. Alternatively, it is freely available at <http://www.s-provencher.com/contin.shtml>. (ii) The cumulant method: The exponential function of a monodispersed system is a straight line in a semi-log plot and its slope is proportional to the mean decay constant ($\bar{\Gamma}$). Deviations from this linearity are a result of polydispersity. Instead, the ACF is fitted with the Tailor expansion of an exponential decay, where the first term is the average and the second gives the variance, the polydispersity index is defined as the ratio between the second and square of the first term [23, 26, 56].
6. Γ is related to the translational diffusion coefficient (D) by $D = \frac{\Gamma}{q^2}$ where $q = \frac{4\pi n}{\lambda} \sin(\frac{\theta}{2})$ with n the refractive index of the solvent, λ wavelength and θ the scattering angle. The R_h can then be calculated by the Stokes-Einstein equation.

3.3 Tertiary Structure Analysis by TF Intensity

1. In the operating software of the spectrophotometer, develop a method by setting up the excitation wavelength at 295 nm. Set the spectrophotometer to scan the emission fluorescence spectra between 310 and 450 nm with a 5 nm step size and a 60 nm/min scan speed (see **Note 10**).

2. Prepare a blank sample with Milli-Q water or the buffer in a quartz cuvette and scan the sample for baseline subtraction following the method set up in **step 1**.
3. Centrifuge the emulsions at $7378 \times g$ for 5 min to separate the dispersed phase from the continuous phase. Remove the continuous phase as it contains all the unabsorbed protein cages, use the dispersed phase to prepare samples to measure tryptophan fluorescence.
4. Dilute the dispersed phase of the emulsion with a continuous phase of emulsion (e.g., Milli-Q water) up to 100 times.
5. Transfer the freshly diluted emulsion samples into a quartz cuvette and scan the sample following the method described in **step 1**.
6. Prepare a protein solution in buffer with the exact same purified protein at a concentration similar to the average protein concentration in the emulsion (*see Note 11*).
7. Transfer the freshly prepared protein samples into a quartz cuvette and scan the sample following the method described in **step 1**.
8. Plot the fluorescence intensity over the range of the emission wavelengths to compare the fluorescence intensity and shift of emission spectra, e.g., red-shift or blue-shift (*see Fig. 2a*) to analyze the conformational integrity of protein cage at the liquid-liquid interface. Figure 2b shows no shift of the E2 protein cage at the oil-water interface, confirming the conformational integrity of the cage structure at the interface.

3.4 Secondary Structure Analysis by Far-UV (CD)

1. For baseline correction of the emulsion sample, formulate an emulsion of the same droplet size using a mass fraction of sodium dodecyl sulfate (SDS) following the method described in Subheading 3.1. It recommended using the same oil: water ratio to formulate SDS-stabilized emulsion that was used to formulate the protein cage-stabilized emulsion sample.
2. Centrifuge the protein cage-stabilized emulsions at $7378 \times g$ for 5 min to separate the dispersed and continuous phases. Remove the continuous phase, containing all the unabsorbed protein cages, and use the dispersed phase to prepare samples for measurement in Far-UV CD.
3. Dilute the dispersed phase of the emulsion with a continuous phase of emulsion (e.g., Milli-Q water) up to at least 100 times.
4. Set the CD instrument to record the ellipticity and the photomultiplier tube (PMT) voltage. Scan the protein cage-stabilized emulsion from 190 to 260 nm with 1 nm step size with 1 s averaging time and 3 scans for each sample. Generate an average spectrum from 3 scans (*see Note 12*).

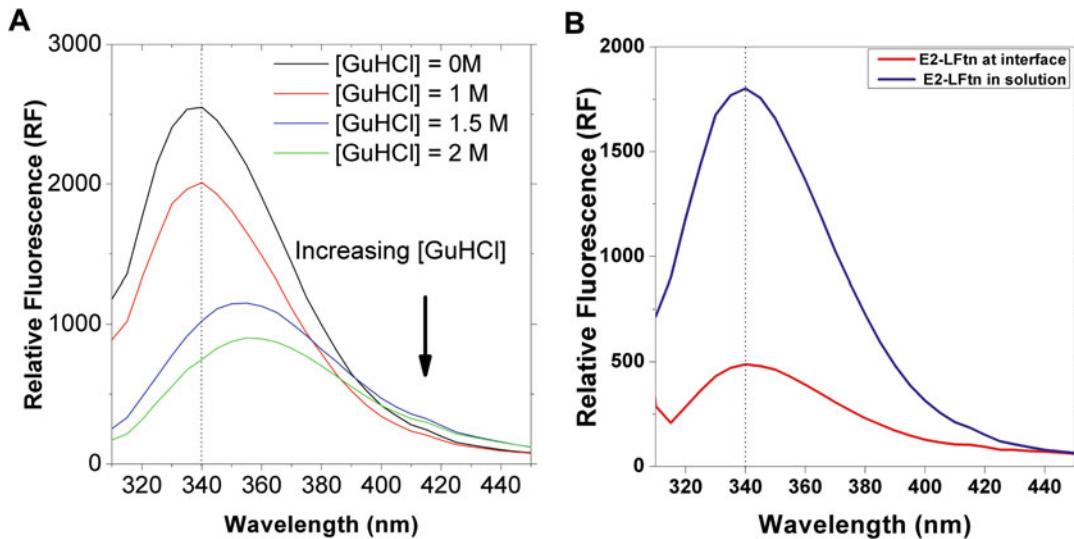


Fig. 2 (a) Shift of emission spectra of tryptophan fluorescence at different concentrations of GuHCl shows the red-shift of Tryptophan Fluorescence Spectra with the change of molecular structure of a protein cage, (b) Comparison of a tryptophan emission spectrum of protein cages before and after adsorption at the oil-water interface, no changes of the peak position of the spectrums shows the conformational integrity of protein cage at the liquid-liquid interface. (Adapted from Sarker et al. [19], © 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim)

5. Scan the SDS-stabilized emulsion from 190 to 260 nm with 1 nm step size with 1 s averaging time and 3 scans for each sample. Generate an average spectrum from 3 scans to obtain a baseline spectrum.
6. Subtract the baseline spectrum generated at **step 5** from sample spectrum generated at **step 4**.
7. Scan freshly purified protein following similar method as **step 4** and subtract the baseline by scanning Milli-Q water (*see Note 13*).
8. After baseline subtraction, smoothen the Spectra using 5 points Savitzky-Golay smoothing filter. Check the PMT voltage of the corresponding spectrum and repeat the scan if applicable (*see Note 14*).
9. Convert the spectra to molar ellipticity (theta) using the following equation:

$$[\theta]_{\text{molar ellipticity}} = \frac{100 \times [\theta]_{\text{deg}}}{c \times l}$$

where $[\theta]_{\text{deg}}$ is the circular dichroism in degree, c is the protein concentration in molar, l is the path length cell in cm.

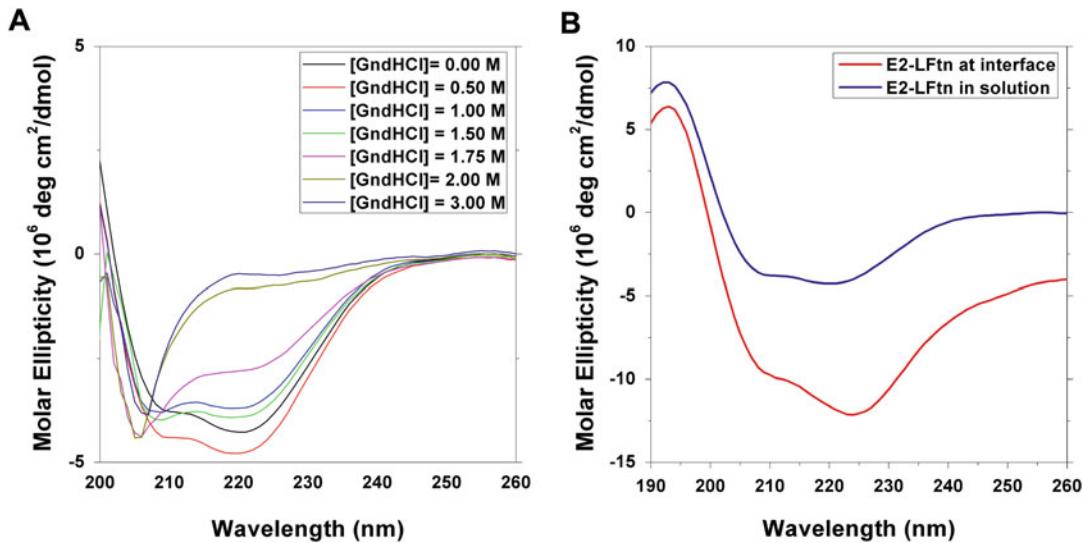


Fig. 3 (a) Far-UV CD spectra of E2 at different concentrations of GuHCl, shows that GuHCl-induced denaturation of E2 is associated with the reduction of both negative (at 208 and 222 nm) and positive (at 190 nm) ellipticity. Reduction of α -helicity and the increase of β -sheet content at high concentrations of GuHCl can be postulated from the decrease in ellipticity. (b) A comparison between the far-UV CD spectrum of E2 in a solvent and E2 after adsorption at the oil-water interface. Both spectra show a positive ellipticity at 193 nm and a negative ellipticity at 210 nm. However, the second negative peak has shifted from 221 nm for E2 in a solvent to 224 nm for E2 adsorbed at the oil-water interface. The relative molar ellipticity in each peak position has increased, which could have resulted from the polarity of a hydrophobic surface on which the cages are adsorbed. The shift of the negative peak of E2 after adsorption reflected in the CD spectra shows a significant increase in α -helix and a decrease in β -strand proportions. (Adapted from Sarker et al. [19], © 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim)

10. Plot the CD spectra obtained at **steps 6** and **7** over the range of far-UV wavelengths to compare the peak position corresponding to the specific secondary structure of protein. Figure 3a shows guanidine hydrochloride (GuHCl) induced shift of CD spectrum of E2 protein cage and Fig. 3b shows a comparison between the far-UV CD spectrum of E2 in a solvent and E2 after adsorption at the oil-water interface.
11. Use any deconvolution software (e.g., K2D2) to estimate the secondary structure content of the protein (*see Note 15*).

3.5 Protein Cage in Bulk and at the Interface Structure Determination using SAXS

1. Set-up and sample considerations: To obtain high quality data, the SAXS system must be aligned to cover the dimensions of interest in the sample. When using laboratory SAXS cameras, this is mostly done by adjusting the appropriate sample-to-detector distances.
2. For the most common protein cages with globular structure and diameter under 30 nm, the maximum dimension probed by a standard SAXS set-up with about 1 m sample detector

distance and the use of a standard copper anode ($\lambda_{\text{Cu}} = 1.5406 \text{ \AA}$) as a laboratory X-ray source, should allow collecting major sample parameters, including overall size, and morphology. At high-intensity synchrotron sources, the wavelength of the X-rays can also be adjusted within seconds.

Similar to DLS, TF and CD, sample purity is highly important as SAXS is a statistical method providing an average over all particles in the scattering volume. Hence, coexisting particles can make appropriate data analysis difficult or impossible. When studying protein cages, it is also important to consider the electron density of the cargo. For example, Ferritin has an iron core with an electron density well above the protein shell; hence the signal will be dominated by the scattering of the iron core [22]. The particle concentration should be selected for an appropriate signal-to-noise ratio. A too-high concentration leads to interparticle interactions (structure factor scattering), which need to be considered when evaluating the scattering curves.

3. For high-quality measurements, use the same clean, thin-wall quartz capillary for all measurements (empty capillary, water, buffer, sample). If necessary, clean with Hellmanex solution (<https://www.hellma.com/laborbedarf/kuevetten/hellmanex/>).
4. Measure the empty capillary and water under the same conditions. Integrate the 2-dimensional (2D) detector read-out to the 1-dimensional (1D) scattering curve using device software or programs such as Fit2D [57]. It is important to exclude areas with dead pixels and pixels with unusually high intensity, compared to neighbors, arising, for instance, from cosmic rays. Ideally, measure triplicates and compare for masking/average. The 1D scattering curves represent the scattering intensity ($I(q)$) at each scattering vector magnitude (q) measured.
5. Subtract the scattering curve of the empty capillary from that of water to determine the water scattering ($I(0)$). This is necessary for the absolute scale calibration of the SAXS data [58].
6. Measure buffer and sample, ideally at different concentrations. After integration from a 2D pattern to 1D curve, as described in step 4, subtract the buffer's scattering curve from the sample's scattering curve (see Fig. 4).
7. Parameters such as the radius of gyration (R_g) can be extracted directly from this background-subtracted scattering curve. The R_g is calculated with the Guinier analysis using the Guinier equation to fit the low- q region of the curve.

$$I(q) = I(0) e^{-\frac{1}{3} R_g^2 q^2}$$

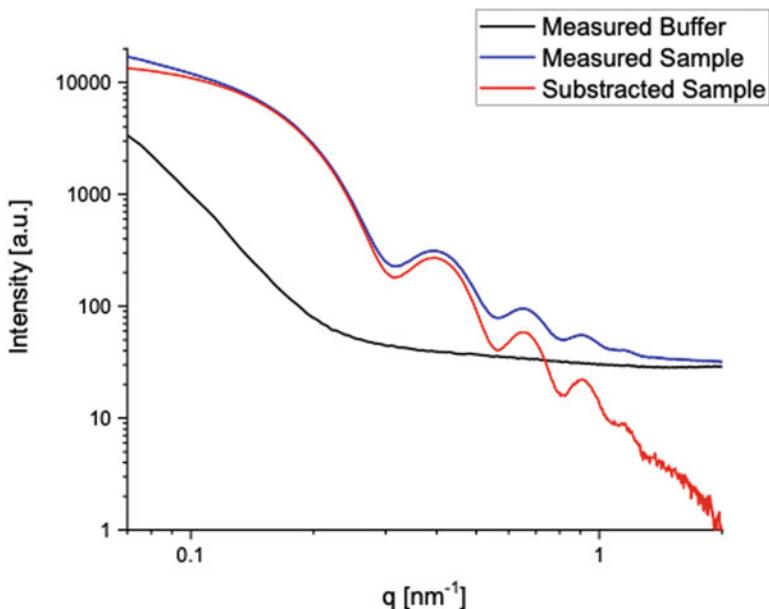


Fig. 4 SAXS scattering curves. The scattering intensity is represented against the magnitude of the scattering vector q . The scattering intensity of the buffer is shown in the black line, the measured scattering intensity of sample (blue line), and the subtracted sample curve (red line)

In a so-called Guinier plot ($\ln(I(q))$ versus q^2) the low- q region appears linear (power-law dependence of $I(q)$ with q). Practically, q -values $< 1.3/R_g$ are considered as the limit for the fitting [59]. This region is used to calculate the R_g from the slope of the fit, which is $-\frac{1}{3}R_g^2$, without any prior assumptions on shape and structure [59]. From the extrapolation of the Guinier equation to $q = 0$, $I(0)$ can also be determined, which can be used to calculate the molecular weight, given that the data is on an absolute scale. The molecular weight of the molecules/particles can then be determined from the forward scattering using water as the standard to calibrate to the absolute intensity [58]. However, one has to be very careful with this analysis: Interparticle interactions, polydispersity, and inhomogeneities in the electron density within the particle (i.e., core-shell type) lead to wrong R_g and $I(0)$ values. For instance, attractive interactions between particles result in an increased upturn of the scattering curve at low- q ; repulsive interparticle interactions induce a decrease in the forward scattering [20, 46, 60–62].

8. The protein conformation, or polymer folding, can be extracted from the $I(q)$ with the help of a Kratky plot $I(q)q^2$ versus q [40, 59]. In this plot, a solid particle with a smooth interface (folded protein) shows a hump. Following Gaussian

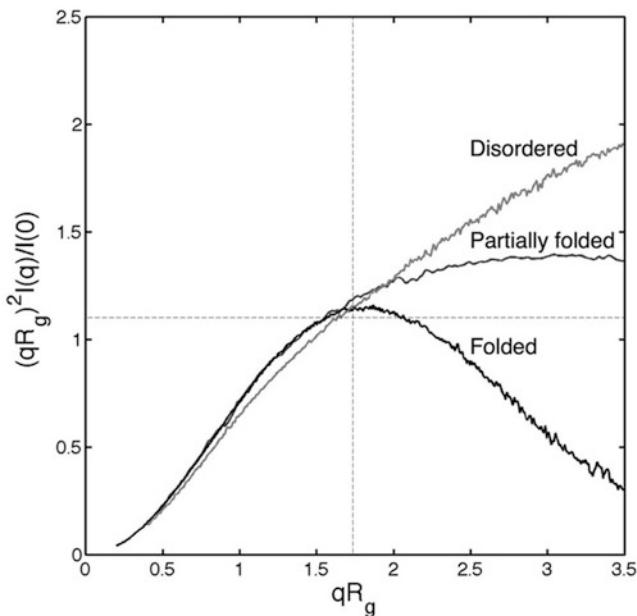


Fig. 5 Dimensionless Kratky plot from folded, partially folded, and disordered proteins from Burger et al. [78]. Dotted lines are drawn at $qR_g = \sqrt{3}$ and $(qR_g)^2 I(q)/I(0) = 1.104$. Folded proteins have a local maximum where the two lines intersect. (Adapted from Burger et al. [78] used under Creative Commons CC-BY license)

chain behavior, partially folded proteins show a plateau at higher q values. Unfolded, disordered proteins show an increase at higher q . The multiplication of q with R_g leads to a dimensionless Kratky plot where information about the size is removed, and only information on the folding is kept. An example of a Kratky plot for different polymer conformations is shown in Fig. 5.

9. Further sophisticated fitting methods exist and allow the investigation of particle size, shape, and morphology. However, they are out of the scope of this chapter. The reader is referred, for instance, to the work of Glatter for model independent fitting methodology [61, 63, 64], and to that of Pedersen for the model dependent fitting methodology [46]. We further point out that some software packages may be useful, including SAS-View, SASfit, Irena, pySAXS, ATSAS, AXES, SCATTER, and BornAgain [65–71]. CRYSTOL in the ATSAS package may be particularly interesting as it allows to use of protein database files as an initial guess for the model [72]. A complete list of software packages for fitting and data reduction can be found at <http://smallangle.org/content/software>. Understanding the models and fitting process is key to obtaining meaningful results here. One can produce reasonable fits to a scattering

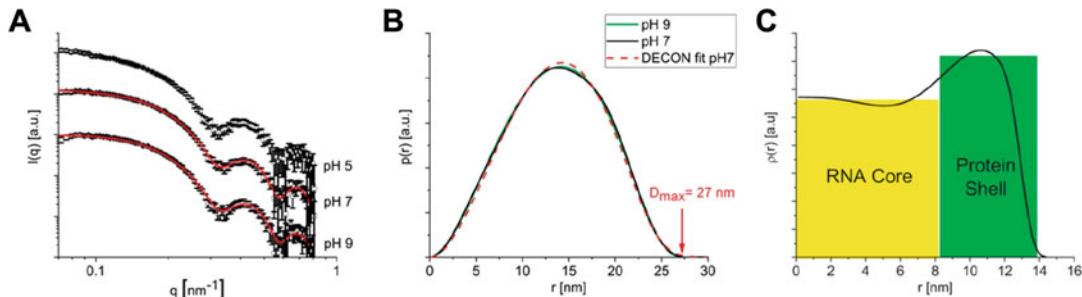


Fig. 6 (a) Experimental SAXS data (open squares) and corresponding fits (red curves) calculated using the generalized indirect Fourier transformation method for MS2. (b) Pair distance distribution calculated from data shown in A (solid black line for pH = 7.0 and green for pH = 9.0). The fit calculated from the deconvolution for pH = 7.0 with a convolution square root operation is shown as a red dashed line. (c) Radial excess electron density of the virus at pH = 7.0, calculated by the deconvolution of part B. (Reprinted (adapted) with permission from Watts et al. [20]. Copyright 2020 American Chemical Society)

curve using various models. The more parameters, the more one is lost in the variable space. Hence, it is essential to gather all possible information on size, polydispersity, etc., ahead of the SAXS fitting from complementary methods such as DLS, and electron microscopy to limit some variables. Further, ensuring that the obtained results have physical meaning is essential.

For example, Fig. 6 shows the model free analysis of scattering data from an MS2 protein cage at different pH values [20]. At pH 5, aggregation is indicated by the low- q upturn of the scattering data, and pH 7 and 9 show a slight downturn from repulsive interactions. The pair distance distribution function is then obtained by transforming the scattering data into real space, using the indirect Fourier transformation method taking the structure factor actions into account (*see* Fig. 6b). It shows the typical shape for core-shell spherical particles with a maximum dimension of 27 nm. Further information on the excess electron density distribution within the protein cage, relative to that of the buffer, can be obtained by deconvolution of this $p(r)$ function using a convolution square-root operation (*see* Fig. 6c). This function demonstrates the dimension of the core and shell, with their relative excess electron densities.

The scattering of cages on the surface of larger particles is also difficult to analyze. The convolution of scattering from the smaller self-assemblies on the surface and the larger substrate particle surface leads to deviation from a simple linear combination of the two signals, requiring more sophisticated models [60, 73, 74]. Figure 7 demonstrates the experimental scattering curves for protein cages with nanocellulose fibers at different pH values [60]. A linear combination of the scattering function from the individual particles (protein cage and

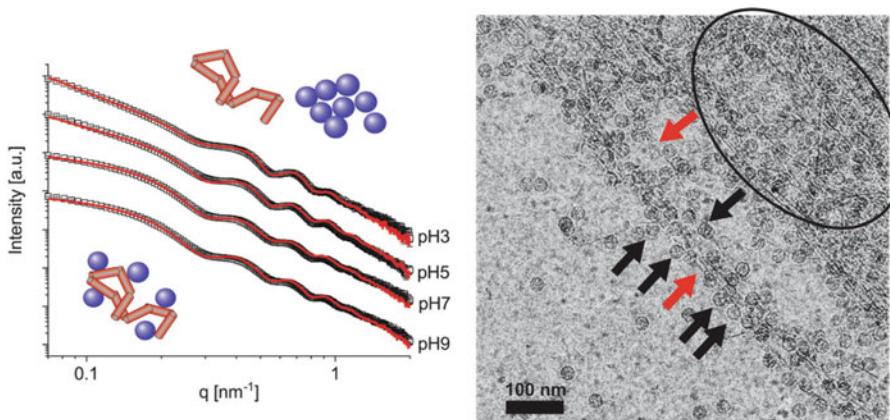


Fig. 7 pH-responsive colloidal interactions among Qbeta protein cages of around 29 nm and nanocellulose fibers. The corresponding fits (red curves) were calculated with a linear combination of the experimental scattering curves from the pure cage and the pure fiber at the corresponding pH value. In the case of interactions, the convolution of waves from the fibers and cages sitting on their surface leads to a significant deviation from the linear combination of the individual scattering signals. This method can reproduce the scattering in the absence of interactions. (Adapted from Watts et al. [79] used under Creative Commons CC-BY license)

nanocellulose) was used to calculate the best fit for the data. Hence, the extent of interaction can be estimated from the mean deviation of the fit. At high pH values, the protein cages interact with the fibers and are bound to their surface, which is represented by a bad fit (high mean deviation). On the contrary, the interaction is absent at low pH, resulting in a good fit.

4 Notes

1. Stable emulsions are best formulated with emulsifiers or a combination of emulsifiers having hydrophile-lipophile balance (HLB) values close to that required for the oil phase [75]. The HLB value and the Refractive Index of rosemary oil are reported to be 15 and 1.468, respectively. Therefore, stabilizing an oil/water (o/w) emulsion with minimum droplet size using rosemary oil requires emulsifiers or a combination of emulsifiers with an HLB value close to 15. Studying the emulsion stability using different mineral oils or triglycerides is recommended. For example, medium-chain triglyceride results in formulating the most stable emulsion while using an E2 protein cage as an emulsifier.
2. It is recommended to use a pure protein cage preparation for formulating the Pickering emulsion. The purity of protein

cages has been found to affect the stability of the emulsion and characterization processes such as CD. For example, impure protein may cause unusual background noise towards shorter wavelengths which is difficult to subtract by background cancellation. Purity of protein 95–99%, potentially achieved by chromatographic separation technique, results in stable emulsion (i.e., no creaming) for up to 3 years. Refer to [76] for the purification protocol of the E2 protein cage.

3. Emulsion formulation by a mechanical homogenizer (blade type) requires careful selection of the dispersing element. Optimizing the energy input is recommended based on the operating volume and oil:water ratio to formulate a stable and homogeneous emulsion with minimum droplet size.
4. Trp excitation and emission wavelengths are 295 nm and near 350 nm respectively. Therefore, the spectrophotometer must be able to excite at 295 nm and read the emission intensity over a wide range of wavelengths (at least 300–450 nm).
5. For preliminary empirical determination of the oil and protein ratio, it is recommended to use v/v ratio of oil and water, e.g., oil:water = 5:95 or 10:90, the higher percentage of oil (e.g., oil:water 40:60) leads to gel formation which is not ideal for the study following the current protocol.
6. Measure the initial volume, V_0 , and final volume V_F at the end of centrifugation to calculate the theoretical final concentration. Take note of the protein loss in the filter surface during the concentration process. Depending on the membrane material or surface area, the loss could be up to 20%.
7. Proteins are susceptible to aggregate at high concentrations; therefore, it is advisable to study the stability of the protein at high concentrations at different time points and in different storage conditions.
8. Soluble proteins are typically prepared in buffers such as Tric-HCl, PBS, etc. In the formulation of emulsion, its recommended to exchange the buffers with Milli-q water following the protocol described for buffer exchange by ultrafiltration method [48]. Therefore, the oil:water ratio remains unchanged in the formulation of emulsion.
9. The coarse emulsion is unstable at ambient conditions. It is advisable to ultrasonicate soon after the crude emulsion formulation. Coarse emulsion often separates into the oil and water phases; as a result, the subsequent homogenization step often fails to homogenize the oil phase floating on the top of the liquid surface of the beaker. In case of such an occurrence, the coarse emulsion must be mixed repetitively until homogenization by ultrasonication.

10. The excitation wavelength of 295 nm was chosen over 280 nm, the tryptophan and tyrosine's excitation wavelength, to selectively excite the Trp residue.
11. For different concentrations of proteins, the fluorescence intensity differs in numerical value as the number of Trp residue differs. It is recommended to prepare fresh protein sample of comparable concentration to achieve similar numerical value of the fluorescent intensity.
12. Load 30 μL of samples in the quartz cuvette cell and fill the space with no trace of air bubble.
13. For baseline correction of fresh protein, use Milli-Q water or the dissolving buffer that contains the protein. For Far-UV CD sample preparation, it is recommended to avoid buffers containing chlorine ions. The chlorine ions generate uncontrolled spectra at wavelengths below 180 nm. Chlorine ions can be removed from the preparation by buffer exchanging by centrifugation or chromatographic method.
14. A current is induced when light hits the photomultiplier of the CD machine. Most CD machines maintain constant current by raising the voltage as the amount of light decreases. As it scans to lower wavelengths, the absorbance will increase and the PMT voltage will rise. The signal-to-noise ratio will greatly diminish once the PMT voltage exceeds 500 V, and the data often becomes very noisy and unreliable. Ensure the PMT voltage is less than 0.5 kV over the wavelength range to ensure the transmitted light level to the detector is adequate for accurate measurement [31].
15. K2D2 is an artificial intelligence program used to find correlations in data using a self-organizing map (SOM) algorithm, a type of neural network [77]. It estimates the percentages of alpha helix and beta strand of a protein from its far-UV CD spectrum where a working wavelength range is 190–240 nm [31, 77].

Acknowledgments

The work was supported in parts by the Singapore Ministry of Education through NTUitive Gap Fund (NGF-2018-04-022), the Swiss National Science Foundation through projects no. 200021_192051; P500PN_206638, and the NCCR Bioinspired Materials.

References

1. Bhaskar S, Lim S (2017) Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Mater* 9(4):e371–e371. <https://doi.org/10.1038/am.2016.128>
2. Neek M, Kim TI, Wang S-W (2019) Protein-based nanoparticles in cancer vaccine development. *Nanomed Nanotechnol Biol Med* 15(1): 164–174. <https://doi.org/10.1016/j.nano.2018.09.004>
3. Steinmetz NF, Lim S, Sainsbury F (2020) Protein cages and virus-like particles: from fundamental insight to biomimetic therapeutics. *Biomater Sci* 8(10):2771–2777. <https://doi.org/10.1039/D0BM00159G>
4. Sarker M, Tomczak N, Lim S (2017) Protein nanocage as a pH-switchable Pickering emulsifier. *ACS Appl Mater Interfaces* 9(12): 11193–11201. <https://doi.org/10.1021/acsmami.6b14349>
5. Lim S, Salentinig S (2021) Protein nanocage-stabilized Pickering emulsions. *Curr Opin Colloid Interface Sci* 56:101485. <https://doi.org/10.1016/j.cocis.2021.101485>
6. Ramsden W, Gotch F (1904) Separation of solids in the surface-layers of solutions and ‘suspensions’ (observations on surface-membranes, bubbles, emulsions, and mechanical coagulation).—Preliminary account. *Proc R Soc Lond* 72(477–486):156–164. <https://doi.org/10.1098/rspa.1903.0034>
7. Pickering SU (1907) CXCVI.—Emulsions. *J Chem Soc Trans* 91(0):2001–2021. <https://doi.org/10.1039/CT9079102001>
8. Leal-Calderon F, Schmitt V (2008) Solid-stabilized emulsions. *Curr Opin Colloid Interface Sci* 13(4):217–227. <https://doi.org/10.1016/j.cocis.2007.09.005>
9. Glatter O, Salentinig S (2020) Inverting structures: from micelles via emulsions to internally self-assembled water and oil continuous nano-carriers. *Curr Opin Colloid Interface Sci* 49: 82–93. <https://doi.org/10.1016/j.cocis.2020.05.003>
10. Glatter O (2018) The inverse scattering problem. In: Scattering methods and their application in colloid and interface science, pp 33–74. <https://doi.org/10.1016/b978-0-12-813580-8.00003-1>
11. Madivala B, Fransaer J, Vermant J (2009) Self-assembly and rheology of ellipsoidal particles at interfaces. *Langmuir* 25(5):2718–2728. <https://doi.org/10.1021/la803554u>
12. Kalashnikova I, Bizot H, Cathala B, Capron I (2011) New Pickering emulsions stabilized by bacterial cellulose nanocrystals. *Langmuir* 27(12):7471–7479. <https://doi.org/10.1021/la200971f>
13. Dugyala VR, Daware SV, Basavaraj MG (2013) Shape anisotropic colloids: synthesis, packing behavior, evaporation driven assembly, and their application in emulsion stabilization. *Soft Matter* 9(29):6711–6725. <https://doi.org/10.1039/C3SM50404B>
14. Wu J, Ma G-H (2016) Recent studies of Pickering emulsions: particles make the difference. *Small* 12(34):4633–4648. <https://doi.org/10.1002/smll.201600877>
15. Dickinson E (2013) Stabilising emulsion-based colloidal structures with mixed food ingredients. *J Sci Food Agric* 93(4):710–721. <https://doi.org/10.1002/jsfa.6013>
16. Dickinson E, Matsumura Y (1994) Proteins at liquid interfaces: role of the molten globule state. *Colloids Surf B: Biointerfaces* 3(1–2): 1–17. [https://doi.org/10.1016/0927-7765\(93\)01116-9](https://doi.org/10.1016/0927-7765(93)01116-9)
17. Yano YF (2012) Kinetics of protein unfolding at interfaces. *J Phys Condens Matter* 24(50): 503101. <https://doi.org/10.1088/0953-8984/24/50/503101>
18. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321–331. <https://doi.org/10.1006/jmbi.1999.3110>
19. Sarker M, Lee H, Gonçalves RA, Lam YM, Su H, Lim S (2020) Supramolecular protein assembly retains its structural integrity at liquid–liquid Interface. *Adv Mater Interfaces* 7(4):1901674. <https://doi.org/10.1002/admi.201901674>
20. Watts S, Julian TR, Maniura-Weber K, Graule T, Salentinig S (2020) Colloidal transformations in MS2 virus particles: driven by pH, influenced by natural organic matter. *ACS Nano* 14(2):1879–1887. <https://doi.org/10.1021/acsnano.9b08112>
21. Watts S, Maniura-Weber K, Siqueira G, Salentinig S (2021) Virus pH-dependent interactions with cationically modified cellulose and their application in water filtration. *Small* 17(30):e2100307. <https://doi.org/10.1002/smll.202100307>
22. Kim M, Rho Y, Jin KS, Ahn B, Jung S, Kim H, Ree M (2011) pH-dependent structures of ferritin and apoferritin in solution: disassembly and reassembly. *Biomacromolecules* 12(5): 1629–1640. <https://doi.org/10.1021/bm200026v>

23. Berne BJ, Pecora R (2000) Dynamic light scattering. Dover Publications, Mineola
24. Bhattacharjee S (2016) DLS and zeta potential – what they are and what they are not? *J Control Release* 235:337–351. <https://doi.org/10.1016/j.jconrel.2016.06.017>
25. Schätzl K (1987) Correlation techniques in dynamic light scattering. *Appl Phys B* 42(4): 193–213. <https://doi.org/10.1007/BF00693937>
26. Hassan PA, Rana S, Verma G (2015) Making sense of Brownian motion: colloid characterization by dynamic light scattering. *Langmuir* 31(1):3–12. <https://doi.org/10.1021/la501789z>
27. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Bioophys J* 80(5):2093–2109. [https://doi.org/10.1016/s0006-3495\(01\)76183-8](https://doi.org/10.1016/s0006-3495(01)76183-8)
28. Rampon V, Genot C, Riaublanc A, Anton M, Axelos MA, McClements DJ (2003) Front-face fluorescence spectroscopy study of globular proteins in emulsions: displacement of BSA by a nonionic surfactant. *J Agric Food Chem* 51(9):2482–2489. <https://doi.org/10.1021/jf026168g>
29. Day L, Zhai J, Xu M, Jones NC, Hoffmann SV, Wooster TJ (2014) Conformational changes of globular proteins adsorbed at oil-in-water emulsion interfaces examined by Synchrotron Radiation Circular Dichroism. *Food Hydrocoll* 34:78–87. <https://doi.org/10.1016/j.foodhyd.2012.12.015>
30. Wong BT, Zhai J, Hoffmann SV, Aguilar M-I, Augustin M, Wooster TJ, Day L (2012) Conformational changes to deamidated wheat gliadins and β -casein upon adsorption to oil–water emulsion interfaces. *Food Hydrocoll* 27(1): 91–101. <https://doi.org/10.1016/j.foodhyd.2011.08.012>
31. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* 1(6):2876–2890. <https://doi.org/10.1038/nprot.2006.202>
32. Bannister WH, Bannister JV (1974) Circular dichroism and protein structure. *Int J Biochem* 5(9):673–677. [https://doi.org/10.1016/0020-711X\(74\)90052-4](https://doi.org/10.1016/0020-711X(74)90052-4)
33. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta (BBA) - Proteins Proteomics* 1751(2):119–139. <https://doi.org/10.1016/j.bbapap.2005.06.005>
34. Husband FA, Garrood MJ, Mackie AR, Burnett GR, Wilde PJ (2001) Adsorbed protein secondary and tertiary structures by circular dichroism and infrared spectroscopy with refractive index matched emulsions. *J Agric Food Chem* 49(2):859–866. <https://doi.org/10.1021/jf000688z>
35. Zhai J, Miles AJ, Pattenden LK, Lee TH, Augustin MA, Wallace BA, Aguilar MI, Wooster TJ (2010) Changes in beta-lactoglobulin conformation at the oil/water interface of emulsions studied by synchrotron radiation circular dichroism spectroscopy. *Biomacromolecules* 11(8):2136–2142. <https://doi.org/10.1021/bm100510j>
36. Mertens HD, Svergun DI (2010) Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol* 172(1):128–141. <https://doi.org/10.1016/j.jsb.2010.06.012>
37. Faccio G, Salentinig S (2017) Enzyme-triggered dissociation of a FRET-based protein biosensor monitored by synchrotron SAXS. *Biophys J* 113(8):1731–1737. <https://doi.org/10.1016/j.bpj.2017.08.044>
38. Watts S, Tran B, Salentinig S (2022) (Anti)viral material design guided by scattering methods. *CHIMIA* 76(10):846. <https://doi.org/10.2533/chimia.2022.846>
39. Bernardo P, Blackledge M (2009) A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J* 97(10):2839–2845. <https://doi.org/10.1016/j.bpj.2009.08.044>
40. Kikhney AG, Svergun DI (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett* 589(19 Pt A):2570–2577. <https://doi.org/10.1016/j.febslet.2015.08.027>
41. Glatter O (2018) Concentration effects, interactions. In: Scattering methods and their application in colloid and interface science, pp 75–95. <https://doi.org/10.1016/b978-0-12-813580-8.00004-3>
42. Sato D, Ohtomo H, Yamada Y, Hikima T, Kurobe A, Fujiwara K, Ikeguchi M (2016) Ferritin assembly revisited: a time-resolved small-angle X-ray scattering study. *Biochemistry* 55(2):287–293. <https://doi.org/10.1021/acs.biochem.5b01152>
43. Zhang L, Bailey JB, Subramanian RH, Groisman A, Tezcan FA (2018) Hyperexpandable, self-healing macromolecular crystals with integrated polymer networks. *Nature* 557(7703):86–91. <https://doi.org/10.1038/s41586-018-0057-7>
44. Glatter O (2018) Concentrated, interacting systems. In: Scattering methods and their application in colloid and interface science, pp

- 299–339. <https://doi.org/10.1016/b978-0-12-813580-8.00013-4>
45. Glatter O (2018) Numerical methods. In: Scattering methods and their application in colloid and interface science, pp 137–174. <https://doi.org/10.1016/b978-0-12-813580-8.00008-0>
46. Pedersen JS (1997) Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv Colloid Interface* 70:171–210. [https://doi.org/10.1016/s0001-8686\(97\)00312-6](https://doi.org/10.1016/s0001-8686(97)00312-6)
47. Milne JLS, Wu X, Borgnia MJ, Lengyel JS, Brooks BR, Shi D, Perham RN, Subramaniam S (2006) Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy *. *J Biol Chem* 281(7):4364–4370. <https://doi.org/10.1074/jbc.M504363200>
48. Phillips AT (1995) Desalting, concentration, and buffer exchange by dialysis and ultrafiltration. *Curr Protoc Protein Sci* 2(1): 4.4.1–4.4.13. <https://doi.org/10.1002/0471140864.ps0404s02>
49. Foley G (2016) Discontinuous diafiltration by sequential dilution (sequential dilution diafiltration). In: Drioli E, Giorno L (eds) Encyclopedia of membranes. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 570–570. https://doi.org/10.1007/978-3-662-44324-8_887
50. LSInstruments (2022) Experimental guidelines - DLS sample preparation. <https://lsinstruments.ch/en/theory/dynamic-light-scattering-dls/experimental-guidelines-dls-sample-preparation>
51. Vysotskii VV, Uryupina OY, Gusel'nikova AV, Roldugin VI (2009) On the feasibility of determining nanoparticle concentration by the dynamic light scattering method. *Colloid J* 71(6):739. <https://doi.org/10.1134/S1061933X09060027>
52. Russel WB, Saville DA, Schowalter WR (1992) Colloidal dispersions, paperback edn. Cambridge University Press, Cambridge
53. Provencher SW (1982) A constrained regularization method for inverting data represented by linear algebraic or integral equations. *Comput Phys Commun* 27(3):213–227. [https://doi.org/10.1016/0010-4655\(82\)90173-4](https://doi.org/10.1016/0010-4655(82)90173-4)
54. Provencher SW, Štěpánek P (1996) Global analysis of dynamic light scattering autocorrelation functions. Part Part Syst Charact 13(5): 291–294. <https://doi.org/10.1002/ppsc.19960130507>
55. Provencher SW (1982) CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput Phys Commun* 27(3): 229–242. [https://doi.org/10.1016/0010-4655\(82\)90174-6](https://doi.org/10.1016/0010-4655(82)90174-6)
56. Mailer AG, Clegg PS, Pusey PN (2015) Particle sizing by dynamic light scattering: non-linear cumulant analysis. *J Phys Condens Matter* 27(14):145102. <https://doi.org/10.1088/0953-8984/27/14/145102>
57. Hammersley AP (2016) FIT2D: a multi-purpose data reduction, analysis and visualization program. *J Appl Crystallogr* 49(2): 646–652. <https://doi.org/10.1107/s1600576716000455>
58. Orthaber D, Bergmann A, Glatter O (2000) SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J Appl Crystallogr* 33(2):218–225. <https://doi.org/10.1107/s00218899015216>
59. Guinier A, Fournet G (1955) Small-angle scattering of X-rays. John Wiley, New York
60. Pedersen JS (2001) Structure factors effects in small-angle scattering from block copolymer micelles and star polymers. *J Chem Phys* 114(6):2839–2846. <https://doi.org/10.1063/1.1339221>
61. Bergmann A, Fritz G, Glatter O (2000) Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing (BSSA). *J Appl Crystallogr* 33(5):1212–1216. <https://doi.org/10.1107/S0021889800008372>
62. Hansen J, Pedersen JN, Pedersen JS, Egelhaaf SU, Platten F (2022) Universal effective interactions of globular proteins close to liquid–liquid phase separation: corresponding-states behavior reflected in the structure factor. *J Chem Phys* 156(24):244903. <https://doi.org/10.1063/5.0088601>
63. Glatter O (1977) Data evaluation in small angle scattering: calculation of the radial electron density distribution by means of indirect Fourier transformation. *Acta Physica Austriaca* 47(1–2):83–102
64. Glatter O, Hainisch B (1984) Improvements in real-space deconvolution of small-angle scattering data. *J Appl Crystallogr* 17(6): 435–441. <https://doi.org/10.1107/S0021889884011894>
65. Bressler I, Kohlbrecher J, Thunemann AF (2015) SASfit: a tool for small-angle scattering data analysis using a library of analytical expressions. *J Appl Crystallogr* 48(Pt 5):1587–1598.

- <https://doi.org/10.1107/S1600576715016544>
66. Ilavsky J, Jemian PR (2009) Irena: tool suite for modeling and analysis of small-angle scattering. *J Appl Crystallogr* 42(2):347–353. <https://doi.org/10.1107/S0021889809002222>
 67. SASView (2021) SASView. <https://www.sasview.org/publications/>. Accessed 27 Aug 2021
 68. Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV, Molodenskiy DS, Panjkovich A, Mertens HDT, Gruzinov A, Borges C, Jeffries CM, Svergun DI, Franke D (2021) ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr* 54(1):343–355. <https://doi.org/10.1107/S1600576720013412>
 69. Grishaev A, Guo L, Irving T, Bax A (2010) Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *J Am Chem Soc* 132(44):15484–15486. <https://doi.org/10.1021/ja106173n>
 70. Förster S, Apostol L, Bras W (2010) Scatter: software for the analysis of nano- and meso-scale small-angle scattering. *J Appl Crystallogr* 43(3):639–646. <https://doi.org/10.1107/S0021889810008289>
 71. Pospelov G, Van Herck W, Burle J, Carmona Loaiza JM, Durniak C, Fisher JM, Ganeva M, Yurov D, Wuttke J (2020) BornAgain: software for simulating and fitting grazing-incidence small-angle scattering. *J Appl Crystallogr* 53(1):262–276. <https://doi.org/10.1107/S1600576719016789>
 72. Svergun D, Barberato C, Koch MHJ (1995) CRYSTOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Appl Crystallogr* 28(6):768–773. <https://doi.org/10.1107/S0021889895007047>
 73. Rao J, Zhang H, Gaan S, Salentinig S (2016) Self-assembly of polystyrene-b-poly(2-vinylpyridine) micelles: from solutions to silica particles surfaces. *Macromolecules* 49(16):5978–5984. <https://doi.org/10.1021/acs.macromol.6b01074>
 74. Larson-Smith K, Jackson A, Pozzo DC (2010) Small angle scattering model for Pickering emulsions and raspberry particles. *J Colloid Interface Sci* 343(1):36–41. <https://doi.org/10.1016/j.jcis.2009.11.033>
 75. Aulton M (1995) Pharmaceuticals: the science of dosage from design. John Wiley & Sons, Inc. New York
 76. Dalmau M, Lim S, Chen HC, Ruiz C, Wang SW (2008) Thermostability and molecular encapsulation within an engineered caged protein scaffold. *Biotechnol Bioeng* 101(4):654–664. <https://doi.org/10.1002/bit.21988>
 77. Perez-Iratxeta C, Andrade-Navarro MA (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct Biol* 8:25. <https://doi.org/10.1186/1472-6807-8-25>
 78. Burger VM, Arenas DJ, Stultz CM (2016) A structure-free method for quantifying conformational flexibility in proteins. *Sci Rep* 6(1):29040. <https://doi.org/10.1038/srep29040>
 79. Watts S, Maniura-Weber K, Siqueira G, Salentinig S (2021) Virus pH-dependent interactions with cationically modified cellulose and their application in water filtration. *Small* 17(30):2100307. <https://doi.org/10.1002/smll.202100307>



Chapter 14

A Method to Investigate the Mechanism of Charge Transport Across Bio-Molecular Junctions with Ferritin

Senthil Kumar Karuppannan and Christian A. Nijhuis

Abstract

To investigate the mechanisms of charge transport (CT) across biomolecular tunnel junctions, it is required to make electrical contacts by a non-invasive method that leaves the biomolecules unaltered. Although different methods to form biomolecular junctions are available, here we describe the EGaIn-method because it allows us to readily form electrical contacts to monolayers of biomolecules in ordinary laboratory settings and to probe CT as a function of voltage, temperature, or magnetic field. This method relies on a non-Newtonian liquid-metal alloy of Ga and In with a few nm thin layer of GaO_x floating on its surface giving this material non-Newtonian properties allowing it to be shaped in to cone-shaped tips or stabilized in microchannels. These EGaIn structures form stable contacts to monolayers making it possible to investigate CT mechanisms across biomolecules in great detail.

Key words EGaIn, Charge transport, Tunnel junction, Biomolecular junction, Self-assembled monolayer, Ferritin

1 Introduction

A good understanding of the mechanisms of charge transport across biomolecules is important to realize biomolecular electronic devices [1–4]. Although the mechanisms of charge transfer across biomolecules are widely studied in wet electrochemical environments [5, 6], in practical settings it is important how such molecules behave in the solid state. Therefore, we developed a method that makes it possible to study the electrical behavior of biomolecules incorporated in tunnel junctions of the form electrode—biomolecule—electrode to measure charge transport rates (i.e., current densities J) as a function of the applied voltage V . Such studies give detailed insights of peculiar phenomena such as long-range tunnelling where CT is essentially independent of the temperature T over incredibly long tunnelling distances exceeding tens of nm [7, 8]. Since biomolecules are soft, it is important to ensure

that the biomolecules are not damaged during fabrication of the top-electrodes. In addition, possible contaminations related to the fabrication of the top-electrode or roughening of the electrode surface have to be eliminated to ensure high quality junctions that do not suffer from defects. This Chapter describes the fabrication of non-invasive EGaIn top-contacts, and ultra-smooth and clean bottom electrodes that do not require additional cleaning steps (which risk unwanted roughening or residual contamination from photoresist) [9, 10]. These pre-patterned ultra-flat and clean bottom-electrode surfaces free of residual polymer or photoresist support high quality monolayers of biomolecules, which in turn, can be contacted with the EGaIn electrodes to complete the junctions.

Fabrication of molecular junctions involves first the deposition of a monolayer onto a bottom-electrode. Self-assembly has been widely used because it readily form dense, well-organized monolayers of molecules. However, the bottom electrode is mostly obtained from direct metal deposition methods which yield surface with relatively high root-mean-squared (rms) surface roughness resulting in defective monolayers [11]. Therefore, it is imperative to use procedures that yield ultra-smooth bottom electrodes with rms surface roughness well-below 1 nm. In this context, template-stripping (TS) is an important method for the fabrication of ultra-smooth bottom electrodes that can be free from photoresist contamination and can be stored under ordinary laboratory conditions with risking deposition of adventitious materials [10, 12]. A clean and ultra-flat substrate (Si/SiO₂, or mica) serve as a template (in the case of Ni template stripping, surface functionalized Si/SiO₂ was used) onto which the metal film is deposited [9]. A composite of curable adhesive and cleaned glass help to strip the thin film away from the templates. The advantage of this template stripping method gives flat surfaces that are scalable and is compatible with a variety of metals including Ag, Au, Pt, Pd, Cu, and Ni, and, since the surface of interest in is contact with the template, the surfaces can be stored for prolonged periods of time without risking the adsorption of contamination [13].

The second step is the deposition of a top electrode onto the molecular monolayer. Direct deposition methods of metals usually require ultra-high vacuum conditions and high temperatures at which biomolecules are not stable or the metal damages the monolayers during deposition resulting in, e.g., pinholes or shorts. Here we demonstrate that EGaIn stabilized in the micro-channels of polydimethylsiloxane (PDMS) can be used as a non-destructive technique to from quality electrical contact [12, 14]. This technique is a variation of the EGaIn-based based on cone-shaped EGaIn tips suspended from a syringe (which in turn is mounted on a micromanipulator) demonstrated in 2008 by Chiechi et al. [15] which lacks mechanical stability prohibiting temperature-dependent charge transport measurements down to 150–200 K.

This technique is based on EGaIn stabilized in a micro-scale through-hole in PDMS which can be directly placed onto the SAM, it can be removed from the SAMs once the measurements are completed and re-used to form a new junction. We also describe even more stable junctions based on defect and photoresist free micropores obtained by template-stripping which allows measurements over a large range of temperature (8.5–340 K) [10].

2 Materials

Ultrapure gold (Au) and nickel (Ni) with a purity of 99.999% and silicon wafers (100, p-type, $500 \pm 25 \mu\text{m}$) with one side polished from Superconductor Materials, Inc. (USA) and University Wafers (USA) were used for bottom-electrode fabrication. Indium tin oxide (ITO) cover slide purchased sigma-Aldrich. The 6-mercaptophexanoic acid, polydimethylsiloxane (PDMS, Sylgard 184, Dow Corning), 1H,1H,2H,2H-perfluoro-octyl-triethoxysilane (FOTS), and eutectic gallium indium 75.5% Ga and 24.5% In by weight were purchased from Sigma Aldrich. 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) was purchased from TCI chemicals. Optical adhesive (OA, Norland, No. 61) was purchased from Edmund optics Inc. All solutions were AR grade or distilled before use.

2.1 Equipment

1. Electrical measurements, Keithley 6430 source meter (*see Note 1*).
2. Electrical measurements, LabVIEW 2010 software.
3. Temperature and magnetic field dependent $J(V)$, probe station, Lakeshore CRX-VF.
4. Thermal evaporator, ShenYang KeYi (*see Note 2*).
5. Sputtering system, AJA ATC-2200, Ultra-high vacuum sputtering system.
6. Thermal oven.
7. Spin coater, SPIN150i spin coater.
8. Mask aligner, MA6 SUSS MicroTec.
9. Oxygen plasma cleaner, VITA mini Reactive ion-etching (RIE), Femto Science Inc.
10. Multimeter to measure the electrical connectivity.

3 Methods

3.1 Fabrication of Template Stripped Au Bottom-Electrodes

1. Silicon (Si) wafer with native oxide (SiO_2) cleaned with acetone and isopropanol and then gently dried in a flow of N_2 before loading it into the deposition chamber (see Fig. 1a).
2. Wait until the deposition chamber reaches the base pressure of $<10^{-6}$ mbar, and then evaporate 150 nm of Au at a rate of 0.5 Å/s.
3. Vacuum anneal Si/Au wafer for 2 h at 200 °C after deposition using in-built heating coil inside the sample holder with heating rates of 15.0 °C/min for 200 °C for 30 min followed by overnight cooling under vacuum to yield the metal coated surfaces (see Fig. 1b). The importance of annealing has been detailed in Ref. [16].

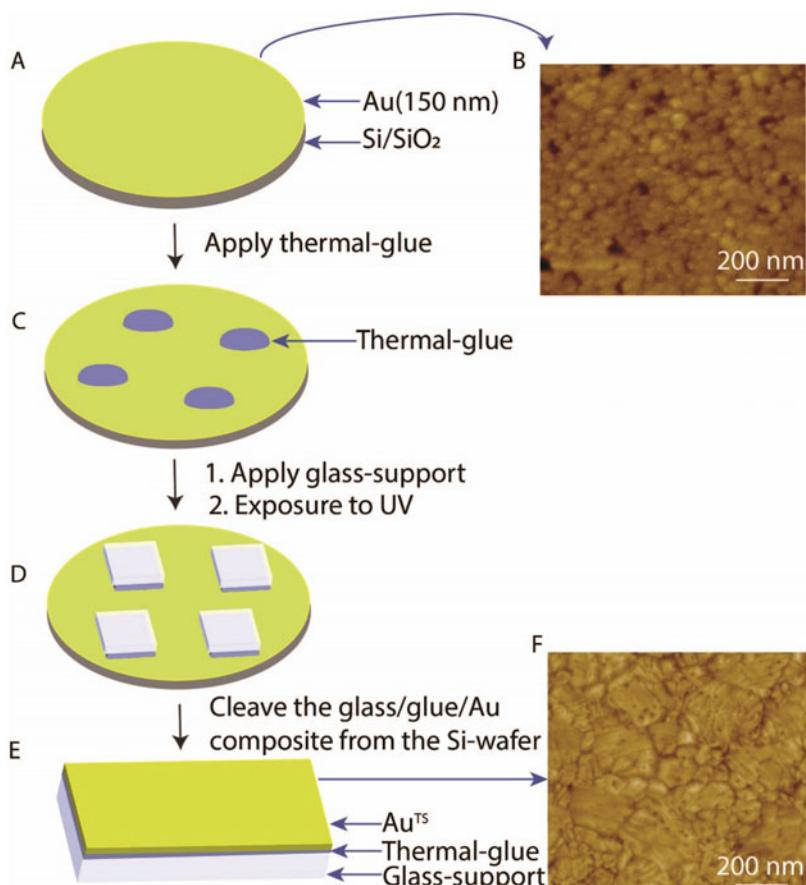


Fig. 1 Schematic illustration of the fabrication process of template-stripped Au (Au^{TS}). (a) Au (150 nm) deposited on a clean Si/SiO_2 surface, (b) AFM image of as deposited Au/Si wafer, (c) thermal-glue applied on the Au metal surface, (d) Glass supports were positioned and curing of the adhesive, (e) the metal film around the glass support was cut out using a razor blade, and (f) AFM image of template stripped Au surface

4. Prepare glass slides for template stripping, clean glass slides with piranha solution (30% H₂O₂: concentrated H₂SO₄ = 1:3) for 15 min, then rinse thoroughly with deionized water and dry under stream of N₂.
5. Use a thermal adhesive (Epotek 353 ND; *see Note 3*) to glue the slides onto the Au surface followed by curing at 80 °C in an oven for 2 h (*see Fig. 3c*).
6. Store the wafer in dry cabinet.
7. Template strip the glass/AO/Au stack from the Si/SiO₂ template just before use to yield ultra-flat and clear Au^{TS} surfaces (*see Fig. 1d*).

3.2 Template--Stripped Nickel (NITS)

1. Si/SiO₂ wafer cleaned with acetone and isopropanol and then gently dry with a flow of N₂ gas.
2. Place the Si/SiO₂ wafer into the vacuum desiccator for 30 min at 1.0 mbar in the presence of small sample vial which contains a few drops of FOTS inside the desiccator to produce a passivation layer before do the Ni deposition (*see Note 4*).
3. Load the FOTS functionalized Si/SiO₂ wafer into the sputter systems and evaporate 150 nm of Ni at a rate of 0.78 Å/s by using an ultra-high vacuum sputtering system (AJA).
4. Prepare glass slides for template stripping process, clean glass slides with piranha solution (30% H₂O₂: concentrated H₂SO₄ = 1:3) for 15 min, then rinse thoroughly with deionized water and finally dry under a stream of N₂ (*see Note 5*).
5. Use an epoxy glue (Norland No. 61) which was cured under UV-light irradiation for 2 h following previously reported procedures [9].
6. To minimize exposure of template stripped Ni film surface with air and spontaneous oxidation of Ni^{TS} film, template stripping was carried out in a N₂ flow hood and then the Ni^{TS} surface was immediately immersed in the corresponding thiol solution to minimize oxidation of Ni^{TS} surface before SAM formation.

3.3 Self-Assembly of Protein Monolayer on the Bottom Electrode

1. Prepare 3 mM of linker molecule (6-merhaptohexanoic acid) solution in 5 mL ethanol.
2. Strip off the glass/glue/Au stack from the Si/SiO₂ template (*see Fig. 1e*), immerse the freshly prepared Au^{TS} substrate directly into the 3 mM ethanolic solution of the 6-merhaptohexanoic acid under N₂ atmosphere (*see Note 6*) to minimize contamination from air (*see Fig. 3a*).
3. In the case of Ni^{TS} (*see Fig. 2*), strip off the glass/OA glue/Ni from the Si/SiO₂-FOTS template in a N₂ flow hood to yield oxide-free Ni^{TS} surfaces. The freshly prepared Ni^{TS} surfaces

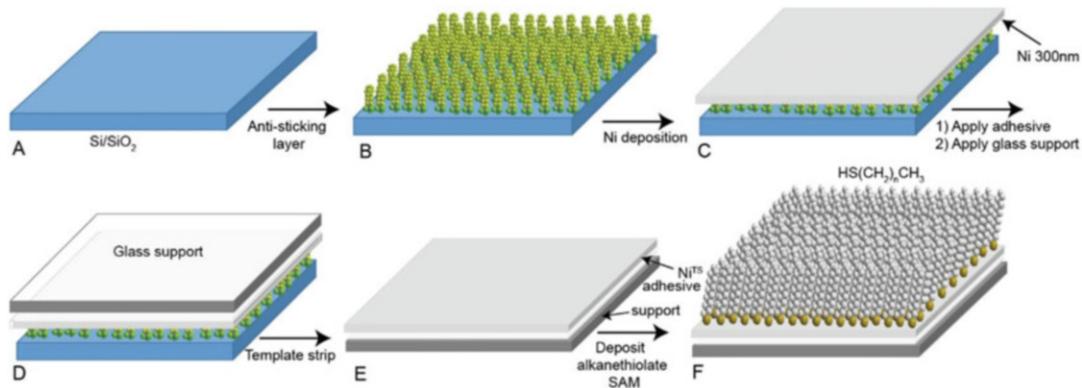


Fig. 2 Schematic illustration of the fabrication process of template-stripped Ni (Ni^{TS}) coated with a SAM. **(a)** On a clean Si/SiO_2 surface **(b)** FOTS was deposited followed by **(c)** deposition of 300 nm Ni. **(d)** Glass supports were glued on the Ni surface using photocurable optical adhesive. **(e)** After curing of the adhesive, the metal film around the glass support was cut out using a razor blade. Next, the nickel film, with its glass support, was lifted off to expose an ultra-flat Ni film (inside of a N_2 flow hood). **(f)** The Ni^{TS} surface was then immediately transferred into a solution of n-alkanethiolates. (Reproduced from Ref. [9] with permission from the Royal Society of Chemistry)

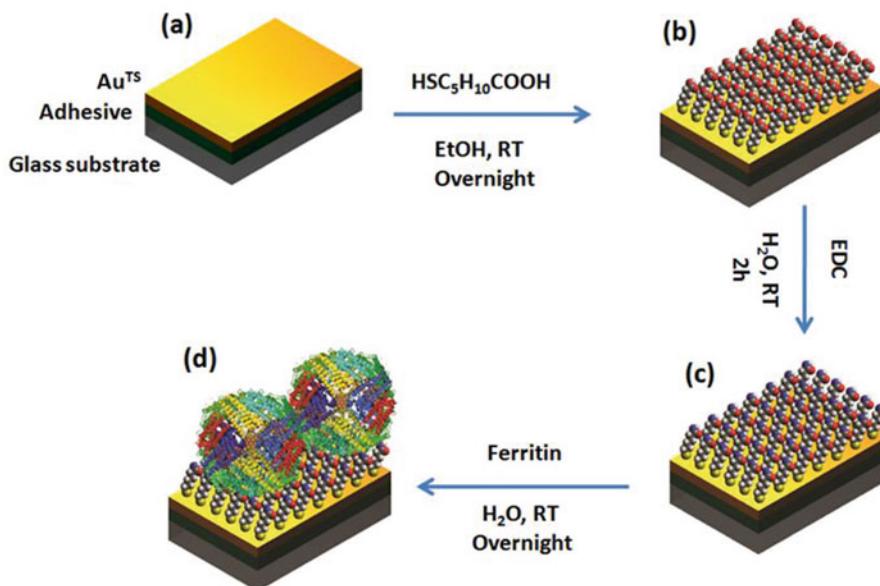


Fig. 3 Schematic description of the fabrication of the ferritin monolayer **(a)** the fabricated template stripped gold substrate (Au^{TS}), **(b)** the Au^{TS} surface was immediately used to prepare SAMs by transferring (through air) into a solution of 6-mercaptophexonic acid, **(c)** monolayer of SAMs on Au^{TS} transferred into 1 mM solution of EDC in water for activation of $-\text{COOH}$ group before **(d)** transferred into the ferritin solution. (Reproduced from Ref. [7] with permission from the John Wiley and Sons)

were immersed immediately (within a few seconds) into ethanolic solutions of 6-merhaptohexanoic acid to form SAMs overnight under a N₂ atmosphere.

4. Then the SAM modified substrates were rinsed with ethanol and blown to dryness in a stream of N₂ gas (*see Fig. 3b*).
5. Next, they were immersed in 1-ethyl-3-(3-imethylaminopropyl) carbodiimide hydrochloride (EDC) (50 mg/mL) solution in EtOH for 2 h (to activate the surface carboxylic acid group; *see Fig. 3c*).
6. The surface activated samples were immersed into the phosphate buffer solution containing ferritin or apo-ferritin (1 mg/mL) overnight in air (*see Note 7*).
7. Finally, the samples were rinsed with deionized water, which removed non-specifically bound ferritin to leave a monolayer of ferritin which were blown to dryness with N₂ gas (*see Fig. 3d*).

3.4 Fabrication of EGaIn Electrodes Stabilized in a Through-Hole

1. The EGaIn stabilized in a through-hole in PDMS was fabricated using a previously reported method [14] which we used to also fabricate arrays of junctions [17].
2. Fabricate the two molds by a standard two-step photolithography process.
3. Oxygen plasma clean Si/SiO₂ wafer using mini RIE at 100W DC power at the oxygen flow of 50 sccm for 1 min.
4. Spin coat SU-8 2010 negative photoresist with 22,360 × g for 70 s on oxygen-plasma cleaned Si/SiO₂ wafer to make mold-1 for a vacuum channel (10 μm × 10 μm × 1 cm) and postbake at 115 °C for 1 min.
5. Expose UV-light on a spin-coated wafer using a mask aligner (SUSS MA-6) for 60 s and then develop a pattern using SU-8 developer for 2 min followed by immersion in DI water for 1 min followed by drying in a stream of N₂ gas and by a postbake at 150 °C for 5 min to harden mold-1.
6. Spin coat SU-8 2035 negative resist with 5590 × g for 70 s to make the connected pillar (with a height of 60 μm and a diameter of 50 μm) and follow the same exposure, development, and postbake as mentioned above.
7. Prepare the mold-2 for the EGaIn injection top-channel with a dimension of 1.0 cm × 200 μm × 100 μm separately using a similar photolithography process using SU-8 2035 negative resist as described above.
8. Place the fabricated mold 1&2 into the vacuum desiccator for 30 min at 1.0 mbar in the presence of small sample vial which contains a few drops of FOTS inside the desiccator to produce passivation layer before spin coating the PDMS (*see Note 8*).

9. Spin coat a 20 μm layer of PDMS onto the mold-1 to cover only the line (*see Fig. 4b*) but not the pillar followed by partially curing in the oven at 80 °C for 20 s (*see Note 9*).
10. Spin coat a 80 μm layer of PDMS onto the mold-2 and follow a similar curing procedure but in this case cured for 45 min.
11. Next, the top PDMS-injection channel was peeled off from the mold-2 and aligned over pillar mold-1 (*see Fig. 4c*).
12. After the alignment, backfill the thin layer of PDMS by pouring carefully more PDMS on the wafer (*see Fig. 4d*).
13. Cure the whole device using standard curing procedure (2 h at 80 °C) and then the PDMS with the through-hole connected to two channels was carefully peeled off from the mold-1 and placed on an ITO substrate for injecting EGaIn into the injection channel (*see Fig. 4e*).
14. The EGaIn is injected into the injection top channel using a 10 mL syringe.
15. Fill the EGaIn in the through-hole by applying vacuum into the vacuum channel.
16. Before use, check the electrical connectivity by measuring the resistance between ITO and EGaIn present at the injection channel using a multimeter to confirm that the EGaIn filled the through-hole and formed good contact with ITO.
17. Finally, remove this top-electrode from the ITO support and place it onto the SAM gently to form the junction (*see Note 10*).

3.5 Fabrication of Micropore-Based Junction

We reported a micropore-based junction to improve the stability of the EGaIn junctions enabling mechanistic studies of charge transport at low $T (> 8.5 \text{ K})$ [10].

1. Deposit patterned Au strips (150 nm thick) along with contact pads on a Si wafer with its native SiO_2 layer by shadow mask deposition (*see Fig. 5a* and **Note 11**).
2. After deposition of Au (*see Fig. 5b*), functionalize the wafer with FOTS to reduce the interaction of the thermal adhesive with the Si/ SiO_2 template to facilitate template-stripping as previously mentioned above (*see Fig. 5c*).
3. Remove the glass/glue/Au stack from the template (*see Fig. 5d*) followed by deposition of 35 nm Al_2O_3 on the entire surface by atomic layer deposition at 90 °C (350 cycles, pulse delay 60 s between Al precursor and water vapour, flow rate 20 sccm) (*see Fig. 5e*).
4. Prepare micropore with a diameter of 10 μm in the Al_2O_3 surface using standard e-beam lithography (*see Fig. 5g* and **Note 12**).

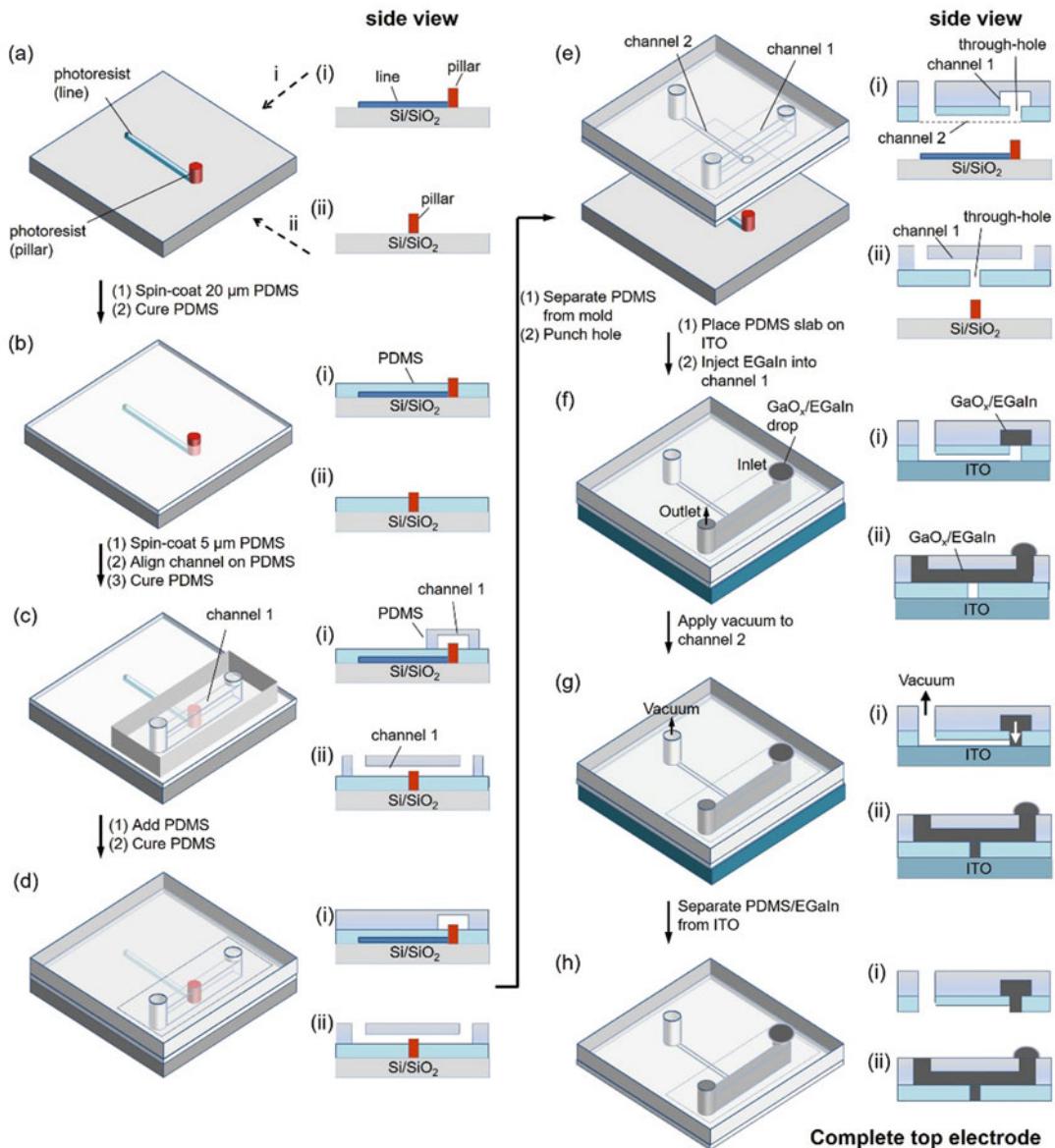


Fig. 4 Fabrication of the top-electrode. (a) The mold-1 consists of a line and a pillar on a Si/SiO₂ wafer with a layer of FOTS (FOTS is not indicated for clarity). (b) A layer of PDMS (20 μm) was spin-coated on the mold-1 to fully cover the photoresist line, but not the pillar, and cured. (c) A thin layer of PDMS (5 μm) was spin-coated on the first layer of PDMS and channel 1 in PDMS was aligned over the pillar perpendicularly with respect to the line of the mold-1. The thin layer of PDMS was cured. (d) More uncured PDMS was added to stabilize the thin layer of PDMS and cured. (e) The microfluidic device was peeled off from the mold-2 and a hole was punched at the end of the small channel. (f) We placed the microfluidic device on an ITO substrate and injected GaO_x/EGaIn into the PDMS channel. (g) The through-hole was filled with GaO_x/EGaIn by applying vacuum to channel 2. (h) Separation of the microfluidic device from the ITO yielded a complete top-electrode. (Reproduced from Ref. [14] with permission from the John Wiley and Sons)

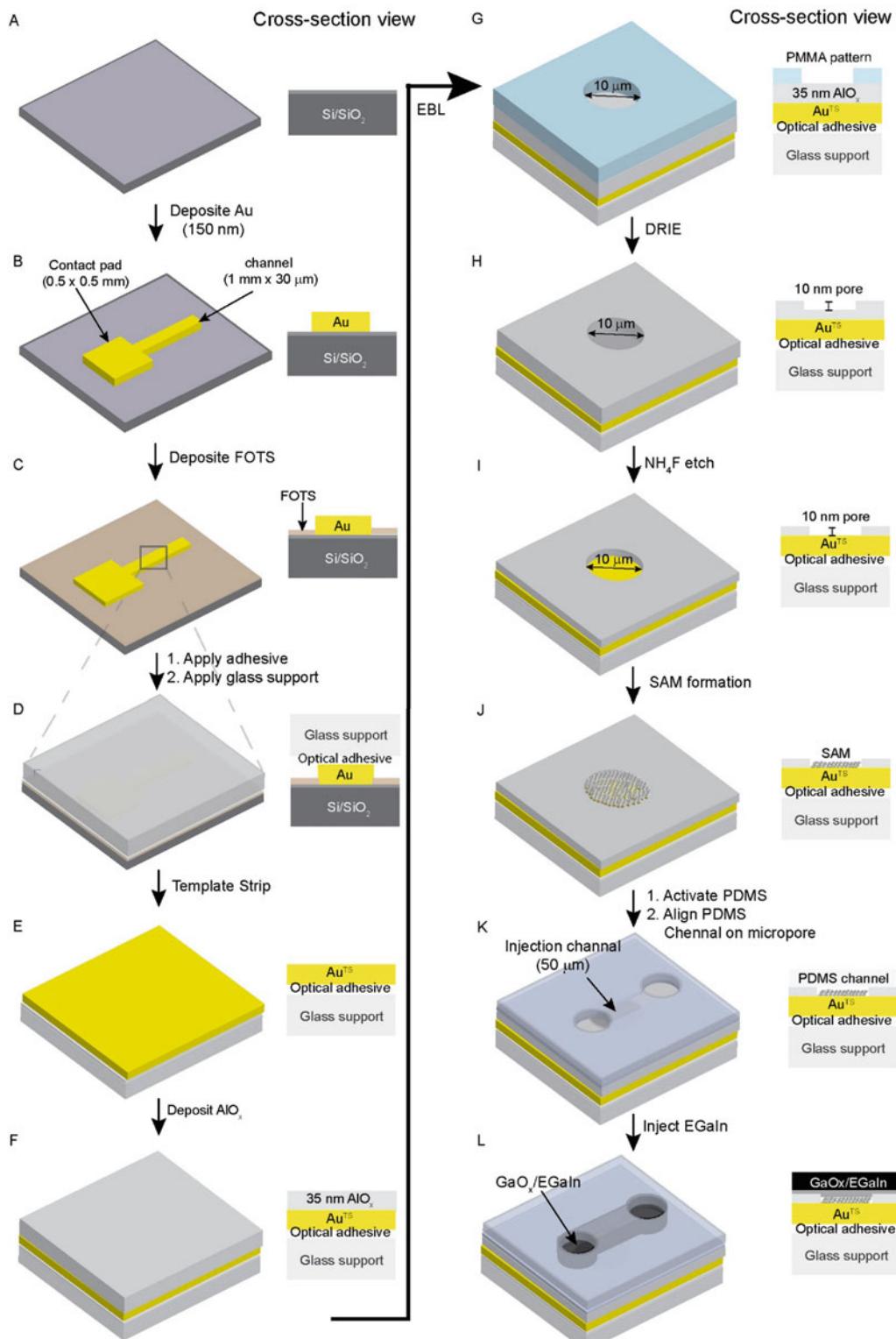


Fig. 5 Schematic illustration of the fabrication processes of the molecular junctions. **(a)** Si wafer with its native SiO_2 layer, **(b)** deposition of the bottom electrode with contact pad (150 nm Au deposited by thermal

5. After e-beam lithography patterning, use deep-reactive ion etching to etch 10 nm Al_2O_3 (*see Note 13* for optimized etching recipe) (*see Fig. 5h*).
6. Etch 25 nm of Al_2O_3 using NH_4F (40%) solution for 2 min and 45 s (*see Fig. 5i*).
7. The SAMs were formed inside the micropore by the same method as described above (*see Fig. 5j*).
8. Align the top contact of EGaIn stabilized in PDMS channels of 80 μm wide, 40 μm high, and 2 cm long perpendicularly over the Au electrodes.
9. Use a stereoscope to align the PDMS channels over the micro-pores (*see Fig. 5k*).
10. A drop of EGaIn was placed on one side of the microchannel and then the microchannels were filled with EGaIn using a syringe pump connected with a small tube to reduce the pressor and to fill the channel with EGaIn at a rate of 0.4–5.0 mm/s to complete the junctions (*see Note 14*).

3.6 Electrical Characterization

3.6.1 Data Collection of DC Measurements

3.6.2 Temperature-Dependent Measurements

The data collection and analysis of the $J(V)$ curves followed previously reported procedure [18–20]. We formed junctions on four to five different samples and determined the $J(V)$ characteristics of five to six junctions on each sample. On each junction, first 5 scans were recorded to stabilize the junction and then 20 traces were recorded. All data were used for data analysis. We defined a short when the value of J reached 10^2 A/cm^2 (the compliance value of J of our instrument) during the scanning (including the first five scans).

Measurements of $J(V)$ as a function of temperature T (K) was carried out in a probe station (Lakeshore CRX-VF) at a pressure of 1.0×10^{-6} mTorr.

1. Gently place the whole device inside the probe station chamber, check the electrical connectivity using Keithley 6430 source meter before closing the chamber.

Fig. 5 (continued) evaporation), (c) deposition of FOTS on Si/SiO₂ to reduce the interaction between optical adhesive (OA) and the Si/SiO₂ substrate, (d) deposition of the adhesive and glass support followed by curing of the adhesive, and (e) template stripping to yield the glass/glue/Au composite layer. On the Au^{TS} substrate, (f) 35 nm AlO_x was deposited by atomic layer deposition, followed by (g) spin-coating of PMMA and patterning of holes ($\varnothing = 10 \mu\text{m}$) in PMMA with electron-beam lithography, (h) etching of 10 nm of AlO_x with deep reactive ion etching and removal of PMMA with O₂ plasma. Next, (i) 25 nm AlO_x was etched with a 40% NH₄F solution followed by (j) formation of the SAM. Finally, (k) a microchannel in PDMS activated with O₂ plasma was aligned over the micropores into which (l) EGaIn was injected to complete the junction. (Reproduced from Ref. [10] with permission from the John Wiley and Sons)

2. Make sure that all the probes are in the up position (remove the probes from the contact position) to avoid junction failure due to vibration or static electricity.
3. Gently close the chamber and then first switch on the rotary pump for only 30 min, to slowly bring the chamber pressure 1.0×10^{-1} mTorr. After reaching a chamber pressure of 0.1 mbar, check the electrical connectivity of the junction.
4. Switch on the turbo pump to reduce the chamber pressure to 1.0×10^{-6} mTorr.
5. Switch on the cryogenic compressor to cooldown the sample stage and set the desired sample stage temperature and measure $J(V)$.
6. After measuring the $J(V)$ curves at the desired temperature, keep the probe in the up position before changing to the next temperature (*see Note 15*).

4 Notes

1. Several other source meters and software can be used, however, we encountered that with the Keithley source-meter auto-scaling can lead to junction failure. To avoid the auto-scaling effect we used the Keithley 6430 and our LabVIEW program with a user-defined current limit.
2. Thermal or electron beam evaporation can be used for metal deposition, we preferred to use thermal evaporation with a heater at the back of the sample holder (Shen Yang Ke Yi, China) for post-deposition annealing process, which is important to minimize small grains on template stripped surface.
3. We can use thermal adhesive or epoxy glue for the template-stripping process based on the solvent and metal film used to make a self-assembled monolayer. The optical adhesive is not stable with THF and hexane solvents, but the thermal adhesive is more stable against all solvents but risks cracking in the case of Ni and Pt metal thin film.
4. The FOTS (anti-stacking layer) functionalization of Si/SiO₂ substrate is very important before the Ni film deposition because this anti-stacking layer effectively reduces the Ni-template interaction enabling us to obtain Ni^{TS} in high yields.
5. A N₂ stream passed through the syringe filter with 1.0 micron pore size to remove particle contaminants from the N₂ gas.
6. To maintain the N₂ atmosphere inside the SAM solution, we purged the ethanol SAM solution with a stream of N₂ gas for 5 min to minimize oxygen in the ethanol solution.

7. Important to maintain high relative humidity environment overnight otherwise the ferritin solution risks to dry on the metal film surface. To maintain high humidity environment, place folded wet Kimwipes around the sample inside the plastic petri dish and covered with parafilm.
8. The mold should be treated with FOTS to reduce the interaction between the PDMS and the mold to ensure the defect-free separation of the PDMS from the mold.
9. Before curing PDMS, we have removed the bubbles using a vacuum desiccator. Place the mold with PDMS inside the vacuum desiccator for 20 min at 1.0 mbar and then slowly release the vacuum.
10. After gently placing the top-electrode on the SAM, apply vacuum for a few seconds if it did not make an electrical connection with SAMs. Important to note to avoid physical force on the channels that after the top electrode is placed on the SAM.
11. Before the metal mask was placed on the silicon wafer, both the metal mask and silicon wafer were cleaned with acetone and IPA followed by drying in a stream of N₂ gas. Next, Au was deposited using thermal evaporator as described before. After Au deposition, the metal mask was cleaned with acetone in an ultrasonication bath for 5 min to ensure the metal shadow mask can be used multiple times.
12. We can use different methods to make a 10 μm micropores in the Al₂O₃ surface, but we preferred to use e-beam lithography because the developer does not etch the Al₂O₃ surface whereas the photolithography developer can etch the Al₂O₃ complicating controlled fabrication of well-defined Al₂O₃ thicknesses.
13. The e-beam patterned substrate with a 10 μm diameter of the micropore on Al₂O₃ surface were loaded into the Oxford Plasma Pro Cobra 100 Deep Reactive Ion Etching (DRIE) tool to etch the 10 nm Al₂O₃, PMMA A4 resist used as a mask layer for DRIE (etching rate for PMMA is 60 nm/min). We used following parameter for C₄F₈ (17 sccm), SF₆ (70 sccm), and O₂ (5 sccm) gases were used for the Al₂O₃ etching process. Pressure in DRIE chamber was maintained at 10 mTorr, and 10 nm Al₂O₃ was etched at 650 W ICP power for 2 min and this process repeated 3 times. Important to note that we should not continuously etch for 6 min because longer etching time make e-beam resist harden and leave contamination on the surface.
14. We should not apply more pressure while tube placed on the microchannel (gently place the tube on the microchannel), if we applied more pressor which leads to short circuit.
15. Change the step size from 10 K to 5 K when the substrate approaches 240 K until 200 K which helps to avoid a shorts.

Acknowledgements

This work support via the National Quantum Fabless foundry, IMRE/A*star, Singapore funded via the QEP 2.0 program under award NRF2021-QEP2-03-P07.

References

1. Gray HB, Winkler JR (2005) Long-range electron transfer. *Proc Natl Acad Sci* 102:3534–3539. <https://doi.org/10.1073/pnas.0408029102>
2. Beratan DN, Liu C, Migliore A, Polizzi NF, Skourtis SS, Zhang P, Zhang Y (2015) Charge transfer in dynamical biosystems, or the treachery of (static) images. *Acc Chem Res* 48:474–481
3. Kim H, Segal D (2017) Controlling charge transport mechanisms in molecular junctions: distilling thermally induced hopping from coherent-resonant conduction. *J Chem Phys* 146:164702. <https://doi.org/10.1063/1.4981022>
4. Karuppannan SK, Pasula RR, Herring TS, Ding J, Chi X, Del Barco E, Roche S, Yu X, Yakovlev N, Lim S, Nijhuis CA (2021) Room-temperature tunnel magnetoresistance across biomolecular tunnel junctions based on ferritin. *J Phys Mater* 4:035003. <https://doi.org/10.1088/2515-7639/abfa79>
5. Bostick CD, Mukhopadhyay S, Pecht I, Sheves M, Cahen D, Lederman D (2018) Protein bioelectronics: a review of what we do and do not know. *Rep Prog Phys* 81:026601. <https://doi.org/10.1088/1361-6633/aa85f2>
6. Akram MS, Rehman JU, Hall EAH (2014) Engineered proteins for bioelectrochemistry. *Annu Rev Anal Chem* 7:257–274. <https://doi.org/10.1146/annurev-anchem-071213-020143>
7. Kumar KS, Pasula RR, Lim S, Nijhuis CA (2016) Long-range tunneling processes across ferritin-based junctions. *Adv Mater* 28:1824–1830. <https://doi.org/10.1002/adma.201504402>
8. Gupta NK, Pasula RR, Karuppannan SK, Ziyu Z, Tadich A, Cowie B, Qi D-C, Bencok P, Lim S, Nijhuis CA (2021) Switching of the mechanism of charge transport induced by phase transitions in tunnel junctions with large biomolecular cages. *J Mater Chem C* 9: 10768–10776. <https://doi.org/10.1039/D0TC05773H>
9. Senthil Kumar K, Jiang L, Nijhuis CA (2017) Fabrication of ultra-smooth and oxide-free molecule-ferromagnetic metal interfaces for applications in molecular electronics under ordinary laboratory conditions. *RSC Adv* 7: 14544–14551. <https://doi.org/10.1039/C6RA27280K>
10. Karuppannan SK, Hongting H, Troadec C, Vilan A, Nijhuis CA (2019) Ultrasmooth and photoresist-free micropore-based EGaIn molecular junctions: fabrication and how roughness determines voltage response. *Adv Funct Mater* 29:1904452. <https://doi.org/10.1002/adfm.201904452>
11. Yuan L, Jiang L, Zhang B, Nijhuis CA (2014) Dependency of the tunneling decay coefficient in molecular tunneling junctions on the topography of the bottom electrodes. *Angew Chem Int Ed* 53:3377–3381. <https://doi.org/10.1002/anie.201309506>
12. Vogel N, Zieleniecki J, Köper I (2012) As flat as it gets: ultrasmooth surfaces from template-stripping procedures. *Nanoscale* 4:3820. <https://doi.org/10.1039/c2nr30434a>
13. Weiss EA, Kaufman GK, Kriebel JK, Li Z, Schalek R, Whitesides GM (2007) Si/SiO₂-templated formation of ultraflat metal surfaces on glass, polymer, and solder supports: their use as substrates for self-assembled monolayers. *Langmuir* 23:9686–9694. <https://doi.org/10.1021/la701919r>
14. Wan A, Jiang L, Sangeeth CSS, Nijhuis CA (2014) Reversible soft top-contacts to yield molecular junctions with precise and reproducible electrical characteristics. *Adv Funct Mater* 24:4442–4456. <https://doi.org/10.1002/adfm.201304237>
15. Dickey MD, Chiechi RC, Larsen RJ, Weiss EA, Weitz DA, Whitesides GM (2008) Eutectic gallium-indium (EGaIn): a liquid metal alloy for the formation of stable structures in micro-channels at room temperature. *Adv Funct Mater* 18:1097–1104. <https://doi.org/10.1002/adfm.200701216>
16. Jiang L, Wang T, Nijhuis CA (2015) Fabrication of ultra-flat silver surfaces with sub-micrometer scale grains. *Thin Solid Films* 593:26–39. <https://doi.org/10.1016/j.tsf.2015.09.023>

17. Wan A, Suchand Sangeeth CS, Wang L, Yuan L, Jiang L, Nijhuis CA (2015) Arrays of high quality SAM-based junctions and their application in molecular diode based logic. *Nanoscale* 7:19547–19556. <https://doi.org/10.1039/C5NR05533D>
18. Reus WF, Nijhuis CA, Barber JR, Thuo MM, Tricard S, Whitesides GM (2012) Statistical tools for analyzing measurements of charge transport. *J Phys Chem C* 116:6714–6733. <https://doi.org/10.1021/jp210445y>
19. Jiang L, Sangeeth CSS, Nijhuis CA (2015) The origin of the odd–even effect in the tunneling rates across EGaIn junctions with self-assembled monolayers (SAMs) of n-alkanethiolates. *J Am Chem Soc* 137: 10659–10667. <https://doi.org/10.1021/jacs.5b05761>
20. Nijhuis CA, Reus WF, Barber JR, Dickey MD, Whitesides GM (2010) Charge transport and rectification in arrays of SAM-based tunneling junctions. *Nano Lett* 10:3611–3619. <https://doi.org/10.1021/nl101918m>



Chapter 15

Potato Virus X Inactivation and Characterization

Yifeng Ma and Nicole F. Steinmetz

Abstract

The nucleoprotein components of plant viruses self-assemble into monodisperse, nanoscale structures with a high degree of symmetry and polyvalency. Of particular interest are the filamentous plant viruses which provide uniform high aspect ratio nanostructures—such structures remain challenging to obtain using purely synthetic approaches. Potato virus X (PVX) has drawn interest by the materials science community because of its filamentous structure measuring 515×13 nm; and both genetic engineering and chemical conjugation methods have been reported to impart new functionalities and develop PVX-based nanomaterials for applications in the health and materials sector. Toward environmentally safe materials—i.e., materials that are not infectious toward crops, such as potato, we reported methods to inactivate PVX. In this chapter, we describe the three methods to inactivate PVX and render it non-infectious toward plants, while maintaining structure and function.

Key words Virus inactivation, Potato virus X (PVX), Viral nanoparticles (VNPs), UV irradiation, β -propiolactone, Formalin

1 Introduction

Potato virus X (PVX) is a 515 nm long and 13 nm wide filamentous plant virus; the nucleoprotein consists of a ssRNA and 1270 identical copies of a coat proteins [1]. Its high aspect ratio structure allows for large payload delivery [2]. Various approaches have been demonstrated that allow the repurposing of PVX for applications including drug delivery [3], biomedical imaging [3], sensing [4], and biocatalysis [5]. Infectious PVX is regulated by the United States Department of Agriculture (USDA) under the Plant Protection and Quarantine program to protect the environment and crops. PVX infects 62 species from 27 plant families, including economic plants as potato, tomato, tobacco, and turnip [6]. To enable the transition from research and development into translational or commercial applications, methods needed to be devised to render PVX-based materials non-infectious to plants. We developed three methods to inactivate PVX: specifically, we applied UV

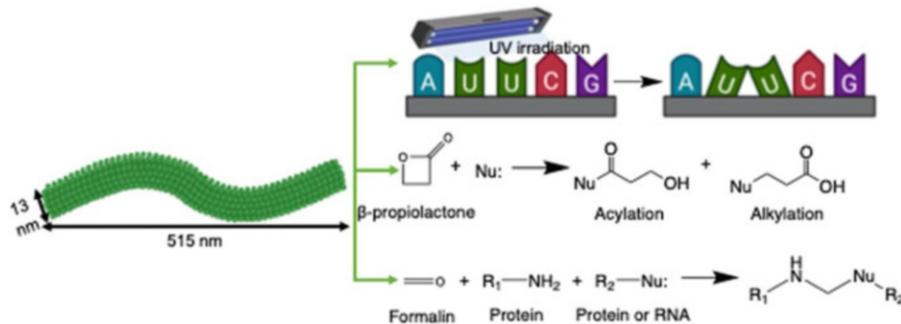


Fig. 1 Schematic of the three PVX inactivation methods: UV irradiation; β -propiolactone; and formalin. (Created with BioRender.com)

irradiation as well as chemical treatment methods (β -propiolactone (β PL) and formalin). These methods have already been reported for inactivation of human pathogens used for vaccine formulations [7, 8]. UV irradiation inhibits RNA translation by dimerizing adjacent uracil [9]. β -propiolactone (β PL) induces acylation and alkylation to the RNA and the protein of the virus [10]. Formalin crosslinks coat proteins and RNA [11]. The mechanisms are outlined in Fig. 1.

Here, we describe these three inactivation methods using PVX as the testbed. However, we note that these methods are also applicable to other plant viruses, as we have demonstrated inactivation of cowpea mosaic virus [12] and tobacco mild green mosaic virus [13] as well. In addition to detailing the inactivation methods, we also report the methods to characterize the structural integrity of the inactivated PVX using dynamic light scattering (DLS), size exclusion chromatography (SEC), and transmission electron microscopy (TEM). The protein and RNA integrity is analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), agarose gel, and ultraviolet-visible spectroscopy (UV-Vis) spectroscopy. Lastly, we describe assays to verify that the inactivated viruses are indeed rendered non-infectious using an *in planta* infection assay and reverse transcription polymerase chain reaction (RT-PCR). Lastly, we report methods for chemical addressability of the inactivated virus—these methods could be used for post-inactivation functionalization of PVX.

2 Materials

All the chemicals can be obtained from Fisher Scientific and Millipore Sigma unless otherwise stated.

2.1 PVX Inactivation

1. PVX was produced as reported previously [14] (*see Note 1*).
2. 0.1 M potassium phosphate (KP) buffer (pH 7.0): Prepare 1 L buffer with 9.343 g of potassium phosphate dibasic and 6.309 g of potassium phosphate monobasic. Bring the pH to 7.0 with sodium hydroxide and hydrochloric acid (*see Note 2*).
3. β -propiolactone (β PL) working solution: 50 mg/mL in DMSO (*see Note 3*).
4. Formalin: 10% (v/v) formaldehyde (buffered, pH 7.0).
5. 230 V, 50 Hz UVP crosslinker (Analytik Jena).
6. Incubated/Refrigerated Stackable Shakers.
7. 30% (w/v) sucrose solution: 50 mL of solution was prepared by mixing 15 g of sucrose in 50 mL of 0.1 M KP buffer (*see Note 4*).
8. 1.5 mL microfuge tube polypropylene (Beckman Coulter).
9. Centrifuges and rotors
 - (i) Optima MAX-TL tabletop ultracentrifuge (Beckman Coulter).
 - (ii) TLA-55 Fixed-Angle Rotor (Beckman Coulter).
 - (iii) Benchtop centrifuge.

2.2 Fluorophore Conjugation

1. Dimethyl sulfoxide (DMSO).
2. Oregon Green 488 (O488)-maleimide: 5 mg/mL in DMSO.
3. O488 carboxylic acid, succinimidyl ester, 5-isomer (O488-NHS): 5 mg/mL in DMSO.

2.3 Virus Particle Characterization

1. NanoDrop™ 2000 1-position Spectrophotometer (Thermo Fisher Scientific).
2. Dry heat block.
3. Mini gel tank.
4. PowerEase® 500 power supply (Invitrogen).
5. NuPAGE™ Bis-Tris mini protein gels (4–12%).
6. MOPS running buffer: 1 L of buffer is prepared by add 50 mL of NuPAGE™ MOPS SDS running buffer (20 \times) into 950 mL DI water.
7. NuPAGE™ LDS sample buffer (4 \times).
8. NuPAGE™ MOPS SDS running buffer.
9. SeeBlue™ Plus2 pre-stained protein standard.
10. Coomassie brilliant blue.
11. Distaining solution: prepare 500 mL with 50 mL acetic acid, 250 mL methanol, and 200 mL DI water.

12. Coomassie stock: prepare 1 L of stock with 2.5 g Coomassie brilliant blue, 100 mL acetic acid, and 450 mL methanol, 450 mL DI water.
13. Coomassie working solution: prepare 500 mL with 13 mL Coomassie stock and 487 mL DI water.
14. Gelred™ working solution: Prepare 50 mL by adding 5 µL of Gelred™ Nucleic Acid Gel Stain (10,000× in water) in 49.995 µL DI water.
15. FluorChem R system (ProteinSimple).
16. Zetasizer Nano ZSP (Malvern Panalytical).
17. Disposable micro cuvette.
18. AKTA Purifier 10 FPLC system (GE Lifesciences).
19. Superose® 6 Increase 10/300 GL (Cytiva).
20. Parafilm M laboratory film.
21. Formvar/Carbon film 10/1 nm on square 400 mesh Cu grid.
22. 2% (w/v) uranyl acetate.
23. Tecnai™ G2 Spirit BioTWIN (FEI company).
24. UltraPure™ DNase/RNase-Free distilled water.
25. 10% (w/v) Sodium dodecyl sulfate (SDS): 50 mL was prepared by dissolving 5 g of SDS in 50 mL of UltraPure™ water (*see Note 5*).
26. 200 proof Ethanol.
27. Ultrapure™ water.
28. Hotplate shaker.
29. UltraPure™ Phenol:chloroform:isoamyl alcohol (PCI) (25: 24:1, v/v).
30. Quick-RNA™ miniprep plus kit (Zymo Research).
31. 1× TAE buffer: Dilute 20 mL TAE buffer (50×) stock with 980 mL DI water.
32. 1.2% (w/v) agarose gel: prepare 100 mL gel by mixing 1.2 g agarose into 1× TAE—dissolve the powder using a microwave (*see Note 6*).
33. 1 kb DNA ladder.
34. Gel Loading Dye, Purple (6×).

2.4 In Planta Inoculation

1. *Nicotiana benthamiana* seeds (harvested from noninfected plants).
2. Carborundum.
3. Cheesecloth.
4. Pro Mix BX soil (Greenhouse Megastore).

5. Plant incubating chamber (Conviron).
6. Spray bottle.

2.5 Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

1. Liquid nitrogen.
2. Precision balance.
3. SuperScript IV One-Step RT-PCR System kit.
4. Primers working solution: 1 mL of 10 μ M aliquot of the forward (5'-CAC CGA AAT CAG CAT GAG CG-3') and reverse (5'-GCC GTG TAC ATC ACA TTC GC-3') primers based on the data sheet offered by the manufacturer and Ultra-Pure™ water.
5. Master stock for 10 sample: 170 μ L of UltraPure™ water, 250 μ L of 2 \times Platinum™ SuperFi™ RT-PCR Master Mix, 25 μ L of each primer, and 5 μ L of SuperScript™ IV RT Mix.
6. 0.2 mL PCR 8-strip Tubes.
7. Thermal cycler.
8. 1.8% (w/v) agarose gel: prepare 100 mL gel by mixing 1.8 g agarose into 1 \times TAE—dissolve the powder using a microwave (*see Note 6*).
9. 100 bp DNA ladder.

3 Methods

3.1 PVX Inactivation

Native and inactivated PVX solutions are prepared with 0.1 M KP buffer (pH 7.0) at 10 mg/mL unless noted otherwise. Concentrations of PVX solution are determined using UV-Vis (*see Subheading 3.3.1*).

3.1.1 UV Inactivation

1. Add 1 mL of 1 mg/mL PVX solution in a 35*10 mm petri dish.
2. Apply 0.5 J/cm² UV irradiation with the UVP crosslinker.
3. Collect the sample and store at 4 °C.

3.1.2 β PL Inactivation

1. Cool 0.1 mL of PVX stock solution on ice.
2. Add 1.44 μ L of β PL working solution (*see Note 7*).
3. Bring the total volume of reaction to 1 mL by adding 899 μ L of cold 0.1 M KP buffer (pH 7.0) on ice.
4. The reaction is performed at 4 °C for 24 h.
5. Inactivate β PL by incubating the reaction mixture at 37 °C on a shaker.

6. Purify the inactivated PVX by ultracentrifugation for 1 h at 50,000 g at 4 °C over a 30% (w/v) sucrose cushion (*see Note 8*).
7. Resuspend the pellet in 200 µL of 0.1 M KP buffer (pH 7.0) overnight (*see Note 9*).
8. Remove any aggregates by centrifugation at 10,000 g for 5 min using a bench top centrifuge.
9. Collect the supernatant and store at 4 °C.

3.1.3 Formalin Inactivation

1. Add 7.5 µL of formalin working solution to 0.1 mL of PVX stock solution.
2. Bring the volume to 1 mL by adding 892 µL of 0.1 M KP buffer (pH 7.0).
3. Incubate the reaction at 37 °C on shaker for 5 days.
4. Purify the inactivated PVX by ultracentrifugation for 1 h at 50,000 g at 4 °C over a 30% (w/v) sucrose cushion (*see Note 8*).
5. Resuspend the pellet in 200 µL of 0.1 M KP buffer (pH 7.0) overnight (*see Note 9*).
6. Remove any aggregates by centrifugation at 10,000 g for 5 min using a bench top centrifuge.
7. Collect the supernatant and store at 4 °C.

3.2 Oregon Green 488 (O488) Conjugation to PVX

3.2.1 Thiol-Maleimide Chemistry to Conjugate O488 to PVX's Cysteines

1. Add 100 µL of native or inactivated PVX (10 mg/mL).
2. Add 3.707, 7.414, 18.54, 37.07 µL of O488-maleimide. The final concentration of O488 in the reaction has 1, 2, 5, 10 dye-to-coat protein ratios.
3. Add 0.1 M KP buffer (pH 7.0) to bring the total volume to 900 µL.
4. Add 100 µL of DMSO.
5. Incubate the reaction at room temperature overnight on a shaker.
6. Purify PVX-O488 by ultracentrifugation for 1 h at 50,000 g at 4 °C over a 30% (w/v) sucrose cushion (*see Note 8*).
7. Resuspend PVX-O488 (pellet) in 200 µL of 0.1 M KP buffer (pH 7.0) overnight (*see Note 9*).

3.2.2 N-Hydroxysuccinimide (NHS)-Amine Chemistry to Conjugate O488 to PVX's Lysines

1. Add 100 µL of native or inactivated PVX (10 mg/mL).
2. Add 4.075, 8.150, 20.4, 40.8 µL of O488-NHS. The final concentration of O488 in the reaction has 1, 2, 5, 10 dye-to-coat protein ratios.
3. Add 0.1 M KP buffer (pH 7.0) to bring the total volume to 900 µL.

4. Add 100 μ L of DMSO.
5. Incubate the reaction at room temperature overnight on a shaker.
6. Purify PVX-O488 by ultracentrifugation for 1 h at 50,000 g at 4 °C over a 30% (w/v) sucrose cushion (*see Note 8*).
7. Resuspend PVX-O488 (pellet) in 200 μ L of 0.1 M KP buffer (pH 7.0) overnight (*see Note 9*).

3.3 Particle Characterization

3.3.1 UV-Vis Spectroscopy

1. Measure the absorbance of PVX (native, inactivated, or O488-conjugated PVX) in 0.1 M KP buffer (pH 7.0) at 260, 280, and 488 nm (use 488 if conjugated with O488) or record the spectrum of 200–700 nm.
2. Determine the concentration of PVX and O488 using the extinction coefficient of PVX (2.97 mL/mg cm) at 260 nm and 76,000/M cm for O488 at 488 nm (*see Note 10*).

3.3.2 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE)

1. Dilute PVX to 1 mg/mL in 0.1 M KP buffer (pH 7.0).
2. Add 4 μ L of NuPAGE™ LDS sample buffer into 10 μ L sample.
3. Heat the sample to 100 °C for 5 min on a heat block, then keep samples on ice.
4. Load a NuPAGE™ 12% Bis-Tris protein gel into the mini gel tank (*see Note 11*) and lock it with the cassette clamp.
5. Fill the front of the gel chamber (*see Note 12*) and at least half of the back of the gel chamber with 1 \times MOPS running buffer.
6. Remove the comb and wash the pockets with buffer, then load 10 μ L pre-stained protein standard into the first well and the samples into the remainder of the wells.
7. Run the gel at 200 V/120 mA for 40 min.
8. Crack the cassette open with a spatula, carefully remove the gel and immerse it in Gelred™ working solution (RNA staining).
9. Stain the gel in Gelred™ working solution for 1 h at room temperature.
10. Image with a FluorChem R system with the “Ethidium Bro-mide” setting to capture nucleic acid staining.
11. Image with “MultiFluor Blue” setting using a FluorChem R system to capture O488 labeling.
12. Then incubate the gel in destaining solution for 1 h at room temperature (*see Note 13*).
13. Incubate the destained gel in Coomassie working solution for 1 h at room temperature (protein staining).
14. Image with a FluorChem R system with the “Coomassie Blue” setting (*see Fig. 2a*).

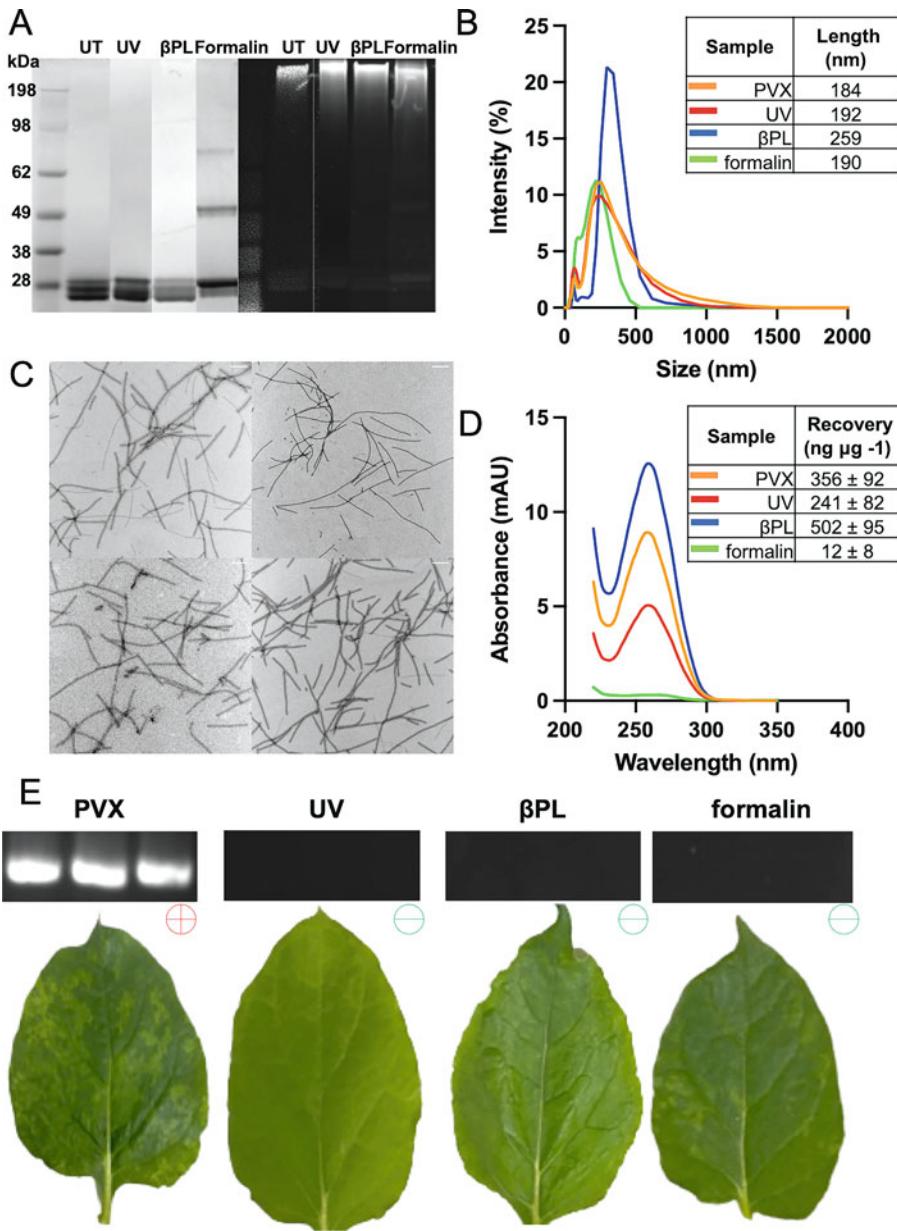


Fig. 2 Characterizations of native and inactivated PVX. **(a)** Protein (left) and RNA (right)-stained SDS-PAGE gel; **(b)** Dynamic light scattering; **(c)** Transmission electron microscopy of negative-stained native (upper left); UV inactivated (upper right); β PL inactivated (lower left); and formalin inactivated (lower right) PVX; the scale bar indicates 200 nm; **(d)** UV-vis spectroscopy; inset shows the amount of RNA extracted from 500 μg of native and inactivated PVX; **(e)** Bottom: *N. benthamiana* inoculated with native (positive control) and inactivated PVX. Visual inspection of the leaves: (–) indicates negative and (+) indicates positive for symptoms. Top: Corresponding RNA-stained agarose gel ran analyzing the RT-PCR amplicons from RNA extracted from the leaves ($n = 3$). (Adapted with permission from Ma et al. [15]. Copyright 2021 American Chemical Society)

3.3.3 Dynamic Light Scattering (DLS)

1. Prepare 250 µg/mL PVX sample in 0.1 M KP buffer (pH 7.0).
2. Load 200 µL of the sample in a cuvette.
3. Measure size information with 173° backscatter at 25 °C using a Zetasizer (*see Fig. 2b*).

3.3.4 Size Exclusion Chromatography (SEC)

1. Prepare 250 µL of 0.5 mg/mL PVX sample in 0.1 M KP buffer (pH 7.0).
2. Inject 200 µL of the sample into the AKTA Purifier 10 FPLC system with a Superose® 6 Increase 10/300 GL column (*see Note 14*).
3. Record the absorbance at 260, 280, and 488 nm (the latter for O488-conjugated PVX).

3.3.5 Transmission Electron Microscopy (TEM)

1. Prepare 0.1 mg/mL PVX sample and DI water.
2. Add 10 µL of sample on the parafilm (*see Note 15*).
3. Load the sample onto the grid by dipping the grid onto the sample drop (*see Note 16*).
4. Incubate for 2 min at room temperature (all steps are performed at room temperature).
5. Tab the tweezers to remove excess liquid.
6. Wash the grid two times using 10 µL of DI water and 45 s incubations.
7. Stain the grid using 10 µL of 2% (w/v) uranyl acetate.
8. Tab the tweezers and blot on filter paper to remove any excess liquid. Then air dry the grid prior to imaging.
9. Use a Tecnai™ G2 Spirit BioTWIN at 80 kV and 30,000× magnification to capture images of PVX (*see Fig. 2c*).

3.3.6 PVX RNA Extraction

1. Sterilize all equipment including the tube holder, the mortar and the pestle in 10% (v/v) bleach overnight; sterilize the tubes and spatulas using an autoclave; and sterilize the pipettes with 4 cycles of UV irradiation at 1 J/cm using the UV crosslinker. Wipe the bench with 10% (v/v) bleach (*see Note 17*).
2. Prepare 500 µL of 1 mg/mL PVX sample in 0.1 M KP buffer (pH 7.0) in a sterilized 2 mL tube.
3. Add 125 µL 10% (w/v) SDS to the sample and heat at 60 °C on a hotplate shaker (*see Note 18*).
4. Add 1.25 mL PCI to the sample (*see Note 19*).
5. Vortex the sample for 20 s and centrifuge at 13,000 g for 10 min at room temperature using a bench top centrifuge.
6. Extract and transfer the upper aqueous phase to a new sterilized tube (*see Note 20*).

7. Repeat steps 3–5 for two more times (see Note 21).
8. Collect the upper aqueous phase and add 600 μL ethanol to the sample.
9. Transfer the ethanol/supernatant mixture to a 2 mL tube loaded with a Zymo-Spin™ IIICG Column.
10. Centrifuge at 13,000 g for 10 min at room temperature using a bench top centrifuge.
11. Add 400 μL RNA prep buffer into the filter, and centrifuge at 13,000 g for 1 min at room temperature using a bench top centrifuge.
12. Add 700 μL RNA wash buffer (see Note 22) into the filter, and centrifuge at 13,000 g for 1 min at room temperature using a bench top centrifuge.
13. Add 400 μL RNA wash buffer into the filter, and centrifuge at 13,000 g for 1 min at room temperature using a bench top centrifuge.
14. Transfer the spin column to a new sterilized 2 mL tube, add 30 μL Ultrapure water, and incubate for 7 min at room temperature.
15. Centrifuge at 13,000 g for 1 min 30 s at room temperature using a bench top centrifuge.
16. Discard the spin column and obtain the RNA (see Note 23).
17. Measure RNA concentration using the NanoDrop™ Spectrophotometer (see Fig. 2d).

3.3.7 RNA Agarose Gel Electrophoresis

1. Heat 1.2% (w/v) agarose gel using a microwave, and ensure the gel is melted into a liquid.
2. Mix 25 mL of melted agarose gel with 2.5 μL Gelred™.
3. Cast the agarose gel in the mold and the comb (see Note 24).
4. After the gel is set, remove the comb.
5. Take 1 μg RNA from the extracted RNA stock (see Note 25) and mix with 1/6 volume of loading dye.
6. Load 2.5 μL 1 kb DNA ladder to the first well and the samples mixed with loading dye to the rest wells.
7. Run the gel at 110 V for 30 min in 1× TAE buffer and take image with a FluorChem R system with the “Ethidium Bromide” setting.

3.4 PVX Plan Infectivity

3.4.1 Plant Inoculation

1. Grow and maintain *N. benthamiana* for 6–8 weeks as previously reported [16].
2. Prepare 0.1 mg/mL of PVX (native or inactivated) in 0.1 M KP buffer (pH 7.0).

3. Dust four primary leaves with carborundum (*see Note 26*).
4. Add 20 µL of sample (*see Note 27*) on each carborundum-dusted leaf and gently rub the solution onto the leaf.
5. Keep the plants in the dark for 1–2 h after mechanical inoculation.
6. Rinse the carborundum off with a spray water bottle and turn the light back on to a normal 16-to-8 h photoperiod.
7. Harvest the secondary leaves after 3 weeks after inoculation (*see Note 28*), when mottling symptoms are visible for the positive controls (*see Fig. 2e*).

3.4.2 RNA Extraction from Leaves

1. Flash-freeze leaf in liquid nitrogen.
2. Grind the frozen leaf using a sterilized mortar and the pestle.
3. Add the homogenized leaf powder into a sterilized tube and weigh to determine the mass of leaf material (*see Note 29*).
4. Add 1 mL ultrapure water per gram of leaf, vortex for 1 min, centrifuge at 13,000 g for 10 min at room temperature using a bench top centrifuge.
5. Collect 500 µL supernatant and transfer to a sterilized 2 mL tube.
6. Extract the viral RNA as described in Subheading 3.3.6.

3.4.3 RT-PCR

1. Add 3 µL of extracted RNA to 47 µL of master stock into a sterilized PCR strip tube.
2. Perform RT-PCR reaction with a program thermal cycler (*see Note 30*).
3. Store the PCR product at 4 °C.

3.4.4 DNA Agarose Gel Electrophoresis

1. Heat 1.8% (w/v) agarose gel using a microwave, and ensure the gel is melted into a liquid.
2. Mix 25 mL of melted agarose gel with 2.5 µL Gelred™.
3. Cast the agarose gel in the mold and the comb (*see Note 24*).
4. After the gel is set, remove the comb.
5. Take 2 µL PCR product and mix with 2 µL of loading dye.
6. Load 2.5 µL 100 bp DNA ladder to the first well and the samples mixed with loading dye to the rest wells.
7. Run the gel at 110 V for 30 min in 1× TAE buffer and take image with a FluorChem R system with the “Ethidium Bro-mide” setting (*see Fig. 2e*).

4 Notes

1. PVX (10 mg/mL) is kept in 0.1 M KP (pH 7.0) buffer at 4 °C.
2. Buffers are kept at room temperature; before use buffers are visually inspected for contaminates (visible as floaters and agglomerates).
3. The final βPL concentration is 1 mM. Store the solution on ice.
4. Sucrose can be mixed using a shaker at 37 °C to accelerate dissolving.
5. SDS can be mixed using a shaker at 60 °C to accelerate dissolving.
6. RNA extracted from PVX is analyzed on a 1.2% (w/v) agarose gel; while the PCR product is analyzed on a 1.8% (w/v) agarose gel.
7. βPL has short half-life at room temperature, perform all steps at room temperature quickly and keep solution on ice. Perform reaction in a chemical safety hood with proper personal protection equipment (PPE), it is carcinogenic [10].
8. The sucrose cushion is prepared by adding 200 µL of 30% (w/v) sucrose slowly to the bottom of the tube. The liquid interface should be visible at lower part of the tube.
9. Mix the solution by pipetting it up and down.
10. Concentration is calculated according to Beer-Lambert law in Eq. 1. Optical path length is 1 mm in Nanodrop (check manufacturer manual if a different equipment is used). Extinction coefficients of the fluorophores are provided by manufacturer. Extinction coefficient of PVX is reported by literature [13].

$$\text{Concentration} = \frac{\text{Absorbance}}{\text{Extinction coefficient} \times \text{optical path length}} \quad (1)$$

11. Remember to peel off the tape at the bottom of the gel.
12. Make sure the MOPS level is higher than that of the wells.
13. Always image nucleic acid staining and fluorescent staining prior to the destaining and Coomassie staining step. Destain the gel even when only imaging for protein (Coomassie blue), so the staining is not oversaturated.
14. More sample is prepared to offset the loss of the volume in the needle.
15. Stick the parafilm on the bench with water.

16. Use reverse tweezers to hold the grid to avoid folding and puncturing. Load the sample onto the grid coated with carbon film—the film coating can be checked optically.
17. Change the gloves frequently and especially after touching a non-sterilized surface or material.
18. Heat and SDS treatment with denature the protein to release the RNA.
19. PCI is a hazardous material. Exposure to it may cause irritation to many organs and carcinogenicity. Perform reaction in a chemical safety hood with proper PPE [17].
20. Use a 200 μL pipette tip to extract the aqueous phase; place the pipette against the wall of the tube and slowly pipette up while avoiding the white precipitant at the interface (denatured protein and salts) and the organic solvent phase in the bottom.
21. Add 1 mL PCI at the last repeat to account for the volume losses in each extraction step.
22. RNA wash buffer working solution is prepared by adding 96 mL of ethanol to the 24 mL RNA wash buffer concentrate from the kit.
23. Extracted RNA is stored at -20°C and used within 3 days.
24. Make sure the surface is flat with a bubble level.
25. The molar extinction coefficient of RNA is 40 $\text{mL}/\mu\text{g cm}$ at 260 nm absorbance.
26. Wear proper PPE (including mask) to avoid touching and inhaling carborundum, which may cause skin and respiratory tract irritation [18].
27. 20 μL of 0.1 M KP buffer (pH 7.0) is used as negative control.
28. Harvested leaves are stored at -80°C .
29. The tube is weighted before the experiment. The weight of the leaf content is calculated by the weight of the tube after the homogenized leaf powder is added and subtracting by that of the empty tube.
30. The PCR cycle is set as following: (1) Reverse transcription at 50°C for 10 min; (2) RT inactivation at 98°C for 2 min; (3) 35 cycles of amplification (10 s at 98°C , 10 s at 55°C , and 30 s at 72°C); (4) Final extension at 72°C for 5 min; (5) Storage at 12°C .

Acknowledgements

This work was supported in part a grant from the National Institute of Food and Agriculture NIFA-2020-67021-31255 and through the UC San Diego Materials Research Science and Engineering Center (UCSD MRSEC), DMR-2011924.

Conflict of Interest Statement Dr. Steinmetz is a co-founder of, has equity in, and has a financial interest with Mosaic ImmunoEngineering Inc. Dr. Steinmetz serves as Director, Board Member, and Acting Chief Scientific Officer, and paid consultant to Mosaic. The other author declares no potential COI.

References

- Steinmetz NF, Mertens ME, Taurog RE, Johnson JE, Commandeur U, Fischer R, Manchester M (2010) Potato virus X as a novel platform for potential biomedical applications. *Nano Lett* 10:305–312. <https://doi.org/10.1021/nl9035753>
- Le DHT, Hu H, Commandeur U, Steinmetz NF (2017) Chemical addressability of potato virus X for its applications in bio/nanotechnology. *J Struct Biol* 200:360–368. <https://doi.org/10.1016/j.jsb.2017.06.006>
- Shukla S, Roe AJ, Liu R, Veliz FA, Commandeur U, Wald DN, Steinmetz NF (2020) Affinity of plant viral nanoparticle potato virus X (PVX) towards malignant B cells enables cancer drug delivery. *Biomater Sci* 8:3935–3943. <https://doi.org/10.1039/D0BM00683A>
- Uhde-Holzem K, McBurney M, Tiu BDB, Advincula RC, Fischer R, Commandeur U, Steinmetz NF (2016) Production of immunoabsorbent nanoparticles by displaying single-domain protein a on potato virus X. *Macromol Biosci* 16:231–241. <https://doi.org/10.1002/mabi.201500280>
- Carette N, Engelkamp H, Akpa E, Pierre SJ, Cameron NR, Christianen PCM, Maan JC, Thies JC, Weberskirch R, Rowan AE, Nolte RJM, Michon T, Van Hest JCM (2007) A virus-based biocatalyst. *Nat Nanotechnol* 2: 226–229. <https://doi.org/10.1038/nnano.2007.76>
- Potato virus X (PVX) | BCH-ORGA-SCBD-114685 | Organism | Biosafety clearing-house. <https://bch.cbd.int/en/database/record?documentID=114685>. Accessed 14 Jul 2022
- Bonafous P, Nicolaï M-C, Taveau J-C, Chevalier M, Barrière F, Medina J, Le Bihan O, Adam O, Ronzon F, Lambert O (2014) Treatment of influenza virus with Beta-propiolactone alters viral membrane fusion. *Biochim Biophys Acta (BBA) - Biomembranes* 1838:355–363. <https://doi.org/10.1016/j.bbamem.2013.09.021>
- U.S. Food and Drug Administration (2019) Common ingredients in U.S. licensed vaccines. <https://www.fda.gov/vaccines-blood-biologics/safety-availability-biologics/common-ingredients-us-licensed-vaccines>. Accessed 18 Jul 2022
- Rae C, Koudelka KJ, Destito G, Estrada MN, Gonzalez MJ, Manchester M (2008) Chemical addressability of ultraviolet-inactivated viral nanoparticles (VNPs). *PLoS One* 3:e3315. <https://doi.org/10.1371/journal.pone.0003315>
- Spaninger E, Bren U (2020) Carcinogenesis of β-propiolactone: a computational study. *Chem Res Toxicol* 33:769–781. <https://doi.org/10.1021/acs.chemrestox.9b00389>
- Hoffman EA, Frey BL, Smith LM, Auble DT (2015) Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem* 290:26404–26411. <https://doi.org/10.1074/jbc.R115.651679>
- Chariou PL, Beiss V, Ma Y, Steinmetz NF (2021) In situ vaccine application of inactivated CPMV nanoparticles for cancer immunotherapy. *Mater Adv* 2:1644–1656. <https://doi.org/10.1039/D0MA00752H>
- Chariou PL, Ma Y, Hensley M, Rosskopf EN, Hong JC, Charudattan R, Steinmetz NF (2021) Inactivated plant viruses as an agro-chemical delivery platform. *ACS Agric Sci Technol*. <https://doi.org/10.1021/acsagscitech.1c00083>
- Lee KL, Uhde-Holzem K, Fischer R, Commandeur U, Steinmetz NF (2014) Genetic engineering and chemical conjugation of potato virus X. In: Lin B, Ratna B (eds) *Virus hybrids as nanomaterials: methods and protocols*. Humana Press, Totowa, NJ, pp 3–21. https://doi.org/10.1007/978-1-62703-751-8_1
- Ma Y, Commandeur U, Steinmetz NF (2021) Three alternative treatment protocols for the efficient inactivation of potato virus X. *ACS Appl Bio Mater* 4:8309–8315. <https://doi.org/10.1021/acsabm.1c00838>
- Shukla S, Dickmeis C, Fischer R, Commandeur U, Steinmetz NF (2018) In planta production of fluorescent filamentous

- plant virus-based nanoparticles. In: Wege C, Lomonosoff GP (eds) Virus-derived nanoparticles for advanced technologies: methods and protocols. Springer, New York, pp 61–84. https://doi.org/10.1007/978-1-4939-7808-3_5
17. UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v). <https://www.thermofisher.com/order/catalog/product/15593031>. Accessed 18 Jul 2022
18. Carborundum (Powder), Fisher Chemical | Fisher Scientific. <https://www.fishersci.com/shop/products/carborundum-powder-fisher-chemical/C192500?searchHijack=true&searchTerm=carborundum-powder-fisher-chemical&searchType=SearchType=rapid&matchedCatNo=C192500>. Accessed 18 Jul 2022



Chapter 16

Molecular Dynamics Simulation of Protein Cages

Chenlin Lu, Xue Peng, and Diannan Lu

Abstract

Molecular dynamics (MD) simulations enable the description of the physical movement of the system over time based on classical mechanics at various scales depending on the models. Protein cages are a particular group of different-size proteins with hollow, spherical structures and are widely found in nature, which have vast applications in numerous fields. The MD simulation of cage proteins is particularly important as a powerful tool to unveil their structures and dynamics for various properties, assembly behavior, and molecular transport mechanisms. Here, we describe how to conduct MD simulations for cage proteins, especially technical details, and analyze some of the properties of interest using GROMACS/NAMD packages.

Key words Molecular dynamics (MD) simulations, Protein cages, GROMACS, NAMD

1 Introduction

Protein Cages are a special group of different-size proteins that have hollow, spherical structures and are widely found in nature [1]. Maize chlorotic mottle viruses (30 nm diameter) [2], lumazine synthases (15 nm diameter) [3], ferritins (12 nm diameter) [4], and DNA binding proteins in starved cells (9 nm diameter) [5] are some examples of cage proteins that have received great attention due to their unique structural features, robustness, and uniform size. These proteins are used in various applications, including drug delivery [6, 7], biominerilization [8, 9], nanoparticle synthesis [10–12], and catalysis [13].

Molecular dynamics (MD) simulation is a computer simulation technology that analyzes the physical movements of molecules and atoms based on their potential energy, as defined by classical mechanics. The potential parameters comprise bonded forces associated with chemical bonds, bond angles, and bond dihedral angles, as well as the non-bonded forces associated with van der Waals and electrostatic interactions, which are typically derived from the empirical fits of high-precision experimental results and quantum

chemical calculations. In 1977, the first MD simulation was conducted by M. Karplus et al. on bovine trypsin [14]. Since then, the technology has developed greatly, and today, it is widely applied in the study of protein folding [15], large-scale protein conformation changes [16], molecular transmembrane transport mechanisms [17], enzyme-substrate recognition [18], and small molecule diffusion [19]. Moreover, due to the rapid development of high-performance computing [20, 21] and the emergence of special-purpose hardware [22, 23] larger systems (up to billion-atom biomolecular systems) [24] can be analyzed by MD, and simulation times have been shortened to the millisecond scale [25]. Currently, MD simulation is recognized as an indispensable technology that can be used to supplement, interpret, and even predict experimental results.

The MD simulation of cage proteins is particularly important as a powerful tool to unveil their structures and dynamics for various properties, assembly behavior, and molecular transport mechanisms. Using this technique, S. Lim et al. demonstrated the importance of interactions between trimers during the self-assembly of the E2 protein cage [26]. Similarly, R. D. Coalson et al. used MD simulations to investigate the metal-binding sites and the mechanism of iron transport to ferroxidase sites in human H-ferritin [27]. The simulations were also used to study the mechanism of iron(II) release from human H-ferritin under different pH conditions [28], as well as the evolution of virus-host interactions [29]. Previously, we had used MD simulations to unveil the mechanism of Au(III) transport into apo-ferritin and to analyze the effect of its mechanism on the structural transition of ferritin [30].

In this chapter, we describe how to conduct MD simulation for cage proteins, especially technical details, and analyze some of the properties of interest. MD simulations can be performed in packages such as GROMACS [31], AMBER [32], NAMD [33], CHARMM [34], and LAMMPS [35]. Here, we use the GROMACS and NAMD packages for demonstration, both of which are fast, free, popular, user-friendly, and suitable for bio-molecules. All input files used for simulation are available at https://github.com/thuresearch/MD_Protein-Cages.

2 MD Simulation Workflow

As shown in Fig. 1, the workflow of MD simulations for cage proteins is typically divided into four main parts: (1) system preparation, (2) equilibration, (3) production simulation, and (4) analysis.

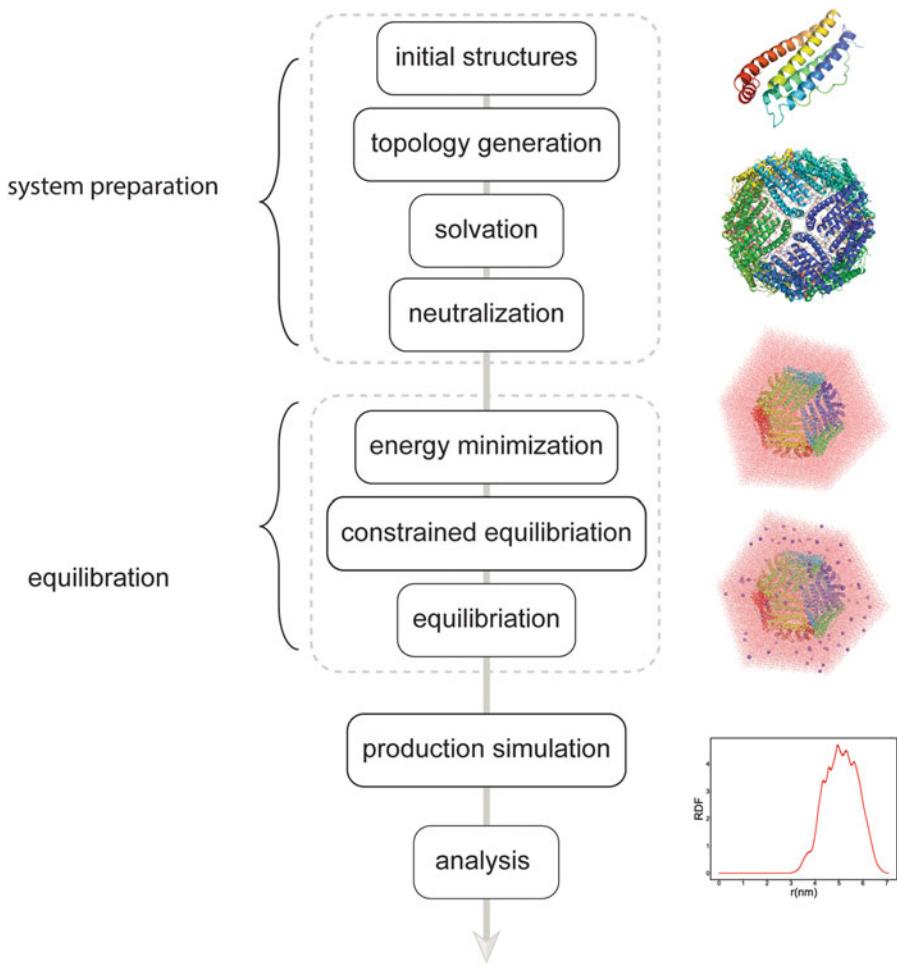


Fig. 1 A typical MD simulation workflow for cage proteins

2.1 System Preparation

2.1.1 Initial Structure

To conduct an MD simulation of a particular protein, its initial structure must be provided. The structures of proteins are usually available in the PDB database, and they are determined based on single-crystal diffraction, nuclear magnetic resonance, cryogenic electron microscopy, and neutron scattering analyses, among others. When multiple structures are provided in the PDB library for the same protein, the one with higher resolution and lower R-value, clash score, Ramachandran outliers, and sidechain outliers is preferred. However, if the protein of interest does not have a structure in the database, its initial structure could be obtained by homology modeling or ab initio modeling. The homology modeling of a single template usually requires a sequence similarity greater than 60%, and at least 30% homology is required for multiple templates. The protein structures that do not meet these conditions may be constructed by ab initio modeling. Recent advances in structure prediction using deep learning models, especially

AlphaFold [36] and RoseTTAFold [37], have facilitated the determination of reliable initial structures based on the protein sequences.

The monomers and their so-called biological assemblies or biological units are needed to elucidate the structures of caged proteins. Therefore, before conducting the simulation, the biological assemblies of monomeric structures must be reconstructed. Using the jsPISA web server [38] or its desktop software PISA, the stable assembly structure can be automatically detected based on the symmetry crystal space group. This structure may also be constructed manually through the symmetry operation (REMARK 300/350) in the PDB file.

2.1.2 Protonation States

MD simulations may be carried out under neutral, acidic, or basic conditions. Empirical pKa predictions such as RPOPKA [39] and H++ [40] were used to determine the protonation states of polar/ionizable residues, including Lys/Arg/His and Glu/Asp, that may be positively or negatively charged.

2.1.3 Topology Generation

In MD simulations, topology refers to the description of the system in which atoms are connected to one another via chemical bonding. Different topologies contribute differently to potential functions. By combining the force field parameters, the numbers and types of residues/atoms in the system, as well as the types of bonds between them, could be determined. The protonation states estimated above are included at this stage.

2.1.4 Solvation

In most cases, proteins perform their physiological functions in aqueous solutions. Therefore, a reasonable description of water molecules must be included in MD simulations of proteins. Generally speaking, solvent models are either explicit or implicit. The explicit models, such as TIP3P, SPC, SPC/E, OPC3, and TIP4P [41], assume that the solvent molecules are directly added to the simulation system. Among these models, TIP3P, SPC, and SPC/E are most commonly used to simulate biomolecules. As for the implicit models [42], they treat the solvent as a continuous medium with a specific dielectric constant, and the most commonly used ones are the Generalized Born (GB) model and the Poisson Boltzmann (PB) model [42].

2.1.5 Neutralization

To maintain the electrical neutrality of the system during MD simulation, counter ions must be added. Considering that the protein comprises charged residues, its net charge in the aqueous solution is not zero, and so, it must interact with oppositely charged ions in the solution, called counter ions, to achieve electric neutrality. Na^+ and Cl^- are the cation and anion most commonly available in biological systems, respectively. Notably, the maintenance of system electric neutrality is a requirement of Ewald, PME, and other algorithms.

2.2 Equilibration

2.2.1 Energy Minimization

Before running the production simulation, the system must be fully equilibrated via a three-stage process: (1) energy minimization, (2) constrained pre-equilibration, and (3) equilibration without pos constraint.

2.2.2 Constrained Pre-equilibration

To avoid simulation crashing under the effect of large forces acting on the atoms, steric clashes and inappropriate geometries must be eliminated, and low-resolution experimental structures must be refined by minimizing the energy of the system. Energy minimization algorithms include the steepest descent, conjugate gradient, Newton, Velocity Verlet, and Leap-frog methods, among others.

2.2.3 Equilibration Without Position Constraint

To keep the structure of the protein unchanged during simulation, the solvent molecules and ions around the protein must be equilibrated in the constrained pre-equilibration stage. If the non-constrained simulation is started directly, the system may collapse. The constraint of position is achieved by adding a harmonic potential to each heavy atom in the protein. The temperature and pressure of the system are raised to the desired values using the NVT and NPT ensembles, respectively.

2.3 Production Simulation

In the second stage of equilibration, the position constraint imposed on the protein is removed, and the simulation is performed using the NPT ensemble. When the simulation is finished, the convergence of various thermodynamic parameters, such as temperature, pressure, energy, volume, density, and root mean square deviation (RMSD), is checked, and if any of these parameters do not converge, additional simulations are performed.

2.4 Analysis of MD Trajectories

2.4.1 RMSD

In the production simulation, the equilibrated system is used as input, and the parameters are the same as those implemented in equilibration simulation without pos constraint, except for simulation time. To ensure that good quality results are obtained, the convergence of various thermodynamic parameters must be checked, as in the case of equilibration simulation.

In this section, we show how the preliminary analysis of production simulation trajectories may be used to obtain some information of interest concerning cage proteins.

RMSD is defined by the following equation,

$$\text{RMSD}(t) = \left[\frac{1}{M} \sum_{i=1}^N m_i \| \mathbf{r}_i(t) - \mathbf{r}_0 \| ^2 \right]^{\frac{1}{2}}$$

where $M = \sum_i^N m_i$, $\mathbf{r}_i(t)$ is the position of atom i at time t . It characterizes the deviation of the protein from the reference structure (usually the crystal structure) during the simulation process.

2.4.2 RMSF

RMSF is another important indicator that is given by the following equation,

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{t_j=1}^T (\mathbf{r}_i(t_j) - \bar{\mathbf{r}}_i)^2}$$

where T is the duration of the simulation (time steps), and $\mathbf{r}_i(t_j)$ is the coordinate of atom i at time t_j . This indicator characterizes the deviation of each residue from the average structure during the simulation, similar to the B-factor in crystallography, and it quantitatively measures the flexibility of each residue.

2.4.3 Radius of Gyration

The radius of gyration is defined as follows,

$$R_g = \left(\frac{\sum_i \|\mathbf{r}_i\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}$$

where m_i and r_i are the mass and position of atom i with respect to the center of mass of the molecule, respectively. R_g is the root mean square distance of the protein from its center of mass. This parameter describes the size or compactness of the protein during the simulation process. In the case of cage proteins, R_g is equivalent to the radius of the cage structure.

2.4.4 Radial Distribution Function

In statistical mechanics, the radial distribution function (RDF) refers to the probability distribution of particles in space (how far away from a given particle), given the coordinates of a certain particle. This function describes the thickness of the protein shell in the cage structure.

2.4.5 Custom Physical Properties

In addition to the commonly calculated properties such as RMSD and RMSF, other custom physical properties may be analyzed, depending on the nature of the scientific problem in question. These custom properties include changes in the number of water molecules surrounding some residues during simulation, the number of hydrogen bonds between two residues, and changes in the size of cage proteins during simulation. Generally, these properties may be estimated using the MDAnalysis [43, 44] and MDTraj [45] analysis tools in the Python library, among others, and using the VMD visualization/analysis software that is commonly used to visualize/analyze MD simulation trajectories [46]. The latter software provides a Tcl console interface that can be applied in the analysis of various custom physical properties.

3 MD Simulation with GROMACS

3.1 Preparation of the System

3.1.1 Initial Structure

The recombinant L-chain apo-rHLFr protein identified in the horse liver (apo-rHLFr, PDB: 1DAT) was chosen as the system, and its initial structure was obtained by X-ray single-crystal diffraction (2.05 Å resolution). Using the jsPISA webserver, three different stable quaternary structures of the protein were detected: (1) 24-mer cage, (2) tetramer, and (3) dimer. Figure 2 shows the cage structure of the apo-rHLFr protein that is formed by the self-assembly of 24 identical monomers. The highly symmetrical structure has an inner diameter of 8 nm and an outer diameter of 12 nm, with twofold, threefold, and fourfold axis-symmetric interfaces.

3.1.2 Protonation States

The 24-mer structure constructed by the jsPISA webserver was used as an input file to determine the protonation states of the system by PROPKA. First, the pKa values were predicted (Table 1); then, the protonation states of all polar residues were determined at

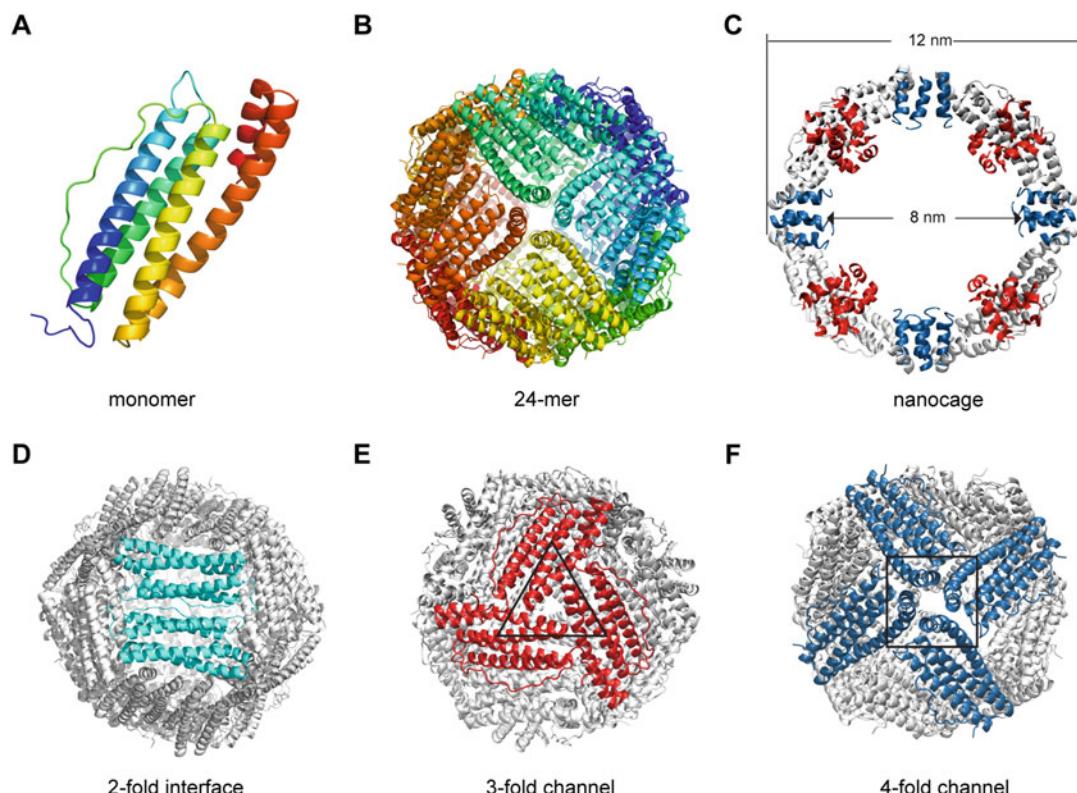


Fig. 2 (a) The monomer, (b) 24-mer, (c) nanocage, (d) twofold interface, (e) threefold channel, and (f) fourfold channel of apo-rHLFr

Table 1
pKa predictions by PROPKA

Residue type	Resid	pKa	Model-pKa
ASP	38	2.9	3.8
ASP	40	4.45	3.8
ASP	41	2.2	3.8
ASP	80	3.48	3.8
ASP	87	3.92	3.8
ASP	94	2.68	3.8
ASP	112	4.38	3.8
ASP	122	4.29	3.8
ASP	127	6.87	3.8
ASP	135	1.47	3.8
ASP	146	6.37	3.8
ASP	174	3.68	3.8
GLU	11	4.16	4.5
GLU	13	3.91	4.5
GLU	45	5.28	4.5
GLU	53	4.65	4.5
GLU	56	4.26	4.5
GLU	57	6.92	4.5
GLU	60	5.72	4.5
GLU	63	5.18	4.5
GLU	88	4.37	4.5
GLU	103	3.38	4.5
GLU	130	10.84	4.5
GLU	136	5.23	4.5
GLU	137	9.16	4.5
GLU	163	4.95	4.5
GLU	167	4.75	4.5
HIS	49	6.19	6.5
HIS	114	4.25	6.5
HIS	124	5.45	6.5
HIS	132	5.06	6.5

(continued)

Table 1
(continued)

Residue type	Resid	pKa	Model-pKa
HIS	147	6.8	6.5
HIS	173	5.42	6.5
LYS	58	14.7	10.5
LYS	67	10.57	10.5
LYS	83	10.66	10.5
LYS	97	10.05	10.5
LYS	104	9.67	10.5
LYS	139	10.78	10.5
LYS	142	12	10.5
LYS	143	11.78	10.5
LYS	172	9.56	10.5
ARG	5	12.84	12.5
ARG	18	12.04	12.5
ARG	25	11.74	12.5
ARG	39	13.66	12.5
ARG	52	13.49	12.5
ARG	59	11.9	12.5
ARG	64	12.11	12.5
ARG	72	12.13	12.5
ARG	75	11.94	12.5
ARG	153	11.05	12.5
ARG	168	14.17	12.5

the pH of interest (pH 7). The protonation states of His173 and Glu130 are HISE/HISD and GLUP, respectively. The predicted pKa values of Glu130 and Glu137 (greater than 7) are significantly different from the model values, and so, they must be corrected later.

3.1.3 Topology Generation

The topology file was generated using the gmx2pdb command in GROMACS. To run this command, the PDB input file of the 24-mer structure must be provided, and the force field and water model must be specified. To use force fields other than the built-in

options, the correct format of the force field file must be placed in the working directory. Considering that different force fields are developed using different water models, it is recommended to read the original force field paper to obtain the best matching model. Here, we made use of the Charmm36 force field and the TIP3P water model, and we applied the following command.

```
gmx pdb2gmx -f 1da_xt_24mer.pdb -o 1dat_24mer.gro -water tip3p
-glu
```

When prompted, choose “1” to use the manually added Charmm36 force field. As shown below, the “-glu” option was also added to interactively correct the protonation states of Glu residues based on the previous pKa prediction. Option “1” was chosen to select the protonated states of Glu130 and Glu137.

Which GLUTAMIC ACID type do you want for residue 130
 0. Not protonated (charge -1) (GLU)
 1. Protonated (charge 0) (GLUP)

Which GLUTAMIC ACID type do you want for residue 137
 0. Not protonated (charge -1) (GLU)
 1. Protonated (charge 0) (GLUP)

3.1.4 Solvation

Considering that most proteins perform their physiological functions in aqueous solution, the behavior of apo-rHLFr was simulated in an aqueous solution. For this purpose, a box was defined and solvated using the following command,

```
gmx editconf -f 1dat_24mer.gro -o box.gro -c -d 1.2
```

where -d sets the distance between the box boundary and the system atom. This distance must be at least greater than half of the non-bonding interaction cutoff. Otherwise, the protein will interact with the neighboring mirror-image protein and produce false phenomena due to the existence of periodic boundary conditions. Based on the output shown below, the volume of the box is 2998.20 nm³.

Note that major changes are planned in future for editconf, to improve usability and utility. Read 66528 atoms
 Volume: 6199.08 nm³, corresponds to roughly 2789500 electrons
 No velocities found

```

system size : 12.018 12.018 12.018 (nm)
center : 0.000 -0.000 -0.000 (nm)
box vectors : 18.370 18.370 18.370 (nm)
box angles : 90.00 90.00 90.00 (degrees)
box volume : 6199.08 (nm^3)
shift : 7.209 7.209 7.209 (nm)
new center : 7.209 7.209 7.209 (nm)
new box vectors : 14.418 14.418 14.418 (nm)
new box angles : 90.00 90.00 90.00 (degrees)
new box volume : 2997.20 (nm^3)

```

The solvation command is as follows,

```
gmx solvate -cp box.gro -cs spc216.gro -p topol.top -o sol.gro
```

where spc216.gro is a pre-equilibrated three-point water model that can be used for the SPC/E and TIP3P solvent topologies. Based on the output file shown below, 78,163 water molecules were added to the system.

```

Generating solvent configuration
Will generate new solvent configuration of 8x8x8
boxes
Solvent box contains 320160 atoms in 106720 residues
Removed 25098 solvent atoms due to solvent-solvent
overlap
Removed 60573 solvent atoms due to solute-solvent
overlap
Sorting configuration
Found 1 molecule type:
    SOL ( 3 atoms): 78163 residues
Generated solvent containing 234489 atoms in 78163
residues
Writing generated configuration to sol.gro

Output configuration contains 301017 atoms in 82339
residues
Volume : 2997.2 (nm^3)
Density : 1061.95 (g/l)
Number of solvent molecules: 78163

```

3.1.5 Neutralization

The following command was used for neutralization in GROMACS, and -maxwarn was used to skip the warning of a system with net charge.

```
gmx grompp -f ionsmdp -c sol.gro -p topol.top -o ions.tpr
-maxwarn 1
```

The content of the mdp file is shown below.

integrator	= steep	; Algorithm (steep = steepest descent minimization)
emtol	= 1000.0	; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep	= 0.01	; Energy step size
nsteps	= 50000	; Maximum number of (minimization) steps to perform
; Parameters describing how to find the neighbors of each atom and how to calculate the interactions		
nstlist	= 1	; Frequency to update the neighbor list and long range forces
cutoff-scheme	= Verlet	
ns_type	= grid	; Method to determine neighbor list (simple, grid)
coulombtype	= PME	; Treatment of long range electrostatic interactions
rcoulomb	= 1.2	; Short-range electrostatic cut-off
rvdw	= 1.2	; Short-range Van der Waals cut-off
pbc	= xyz	; Periodic Boundary Conditions (yes/no)

Based on the output file, the system has a negative charge of -120 e.

System has non-zero total charge: -120.000000
 Total charge should normally be an integer.

```
gmx genion -s ions.tpr -o ions.gro -neutral -p topol.top
```

where -neutral signifies that the system will be neutralized by adding Na/Cl ions, and ions.tpr is the binary file containing the atomic-level description of our system. The types and numbers of the added ions may also be specified using the -pq -pname and -nq -nname options. When prompted, the group 13 “SOL” option was chosen to embed ions by replacing some of the solvent molecules. At this point, the system model is completely established.

3.2 Equilibration

3.2.1 Energy Minimization

The command for energy minimization in GROMACS and the content of the mdp file are given below.

```
gmx grompp -f mini.mdp -c ions.gro -p topol.top -o em.tpr
```

```
; LINES STARTING WITH ';' ARE COMMENTS
title      = Minimization      ; Title of run

; Parameters describing what to do, when to stop and what to save
integrator = steep           ; Algorithm (steep = steepest descent minimization)
emtol      = 1000.0
emstep     = 0.01             ; Energy step size
nsteps     = 50000            ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist    = 1                 ; Frequency to update the neighbor list and long range forces
cutoff-scheme = Verlet
ns_type    = grid              ; Method to determine neighbor list (simple, grid)
rlist      = 1.2               ; Cut-off for making neighbor list (short range forces)
coulombtype = PME              ; Treatment of long range electrostatic interactions
rcoulomb   = 1.2               ; long range electrostatic cut-off
vdwtype    = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw       = 1.2               ; long range Van der Waals cut-off
pbc        = xyz               ; Periodic Boundary Conditions
```

The maximum force, the criterion used to judge whether energy minimization is complete, was set to 1000 kJ/mol/nm. This means that the maximum force in the system must not exceed 1000. To minimize the energy, the following command was used.

```
gmx mdrun -v -deffnm em
```

3.2.2 Constrained Pre-equilibration

In constrained pre-equilibration, the solvent molecules and ions around the protein are equilibrated to keep the structure of the protein unchanged. The parameters of the harmonic constraint potential on heavy atoms were defined by posres.itp and generated by pdb2gmx. First, the temperature of the system was raised to the desired value (298.15 K) using the NVT ensemble and the following command.

```
gmx grompp -f nvtmdp -c em.gro -r em.gro -p topol.top -o nvt.tpr
```

The content of the mdp file is as follows.

```

define = -DPOSRES
integrator = md
dt      = 0.002 ; ps
nsteps  = 50000 ; 100ps
comm-grps = system
energygrps =
;
nstxout = 0
nstvout = 0
nstfout = 0
nstlog  = 500
nstenergy = 500
nstxout-compressed = 1000
compressed-x-grps = system

; Bond parameters
continuation      = no           ; first dynamics run
constraint_algorithm = lincs     ; holonomic constraints
constraints        = h-bonds
lincs_iter         = 1            ; accuracy of LINCS
lincs_order        = 4            ; also related to accuracy
; Neighborsearching
cutoff-scheme = Verlet
ns_type          = grid          ; search neighboring grid cells
nstlist          = 10            ; 20 fs, largely irrelevant with Verlet
rcoulomb         = 1.2           ; short-range electrostatic cutoff (in nm)
rvdw             = 1.2           ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype      = PME          ; Particle Mesh Ewald for long-range electrostatics
pme_order         = 4            ; cubic interpolation
fourierspacing   = 0.16          ; grid spacing for FFT
; Temperature coupling is on
tcoupl           = V-rescale    ; modified Berendsen thermostat
tc-grps          = Protein Non-Protein ; two coupling groups - more accurate
tau_t             = 0.2 0.2       ; time constant, in ps
ref_t             = 298.15        ; reference temperature, one for each group, in K
; Pressure coupling is off
pcoupl           = no           ; no pressure coupling in NVT
; Periodic boundary conditions
pbc              = xyz          ; 3-D PBC
; Dispersion correction
DispCorr         = EnerPres     ; account for cut-off vdW scheme
; Velocity generation
gen_vel          = yes          ; assign velocities from Maxwell distribution
gen_temp          = 298.15        ; temperature for Maxwell distribution
gen_seed          = -1           ; generate a random seed

```

The NVT simulation was run using the following command.

```
gmx mdrun -v -deffnm nvt
```

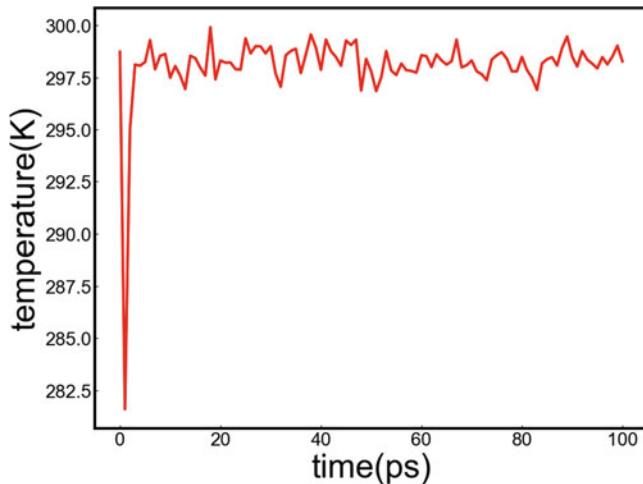


Fig. 3 The evolution of system temperature during NVT simulation

The resulting edr file contains information about the energy, temperature, pressure, box size, and density of the protein system. The value of interest was extracted using the gmx energy, and the temperature change during the simulation was obtained using the following command.

```
gmx energy -f nvt.edr -o temperature.xvg
```

“17 0” was entered at the prompt to select the temperature of the system and exit. The resulting temperature.xvg file shown in Fig. 3 indicates that the temperature of the system stabilizes at 298.15 K.

As for the pressure of the system, it was raised to the desired value (1 atm) using the NPT ensemble and the following command,

```
gmx grompp -f nptmdp -c nvt.gro -r nvt.gro -t nvt.cpt -p
topol.top -o npt.tpr
gmx mdrun -v -deffnm npt
```

where -t signifies the checkpoint file from NVT equilibration.

The content of the mdp file is as follows.

```

define = -DPOSRES
integrator = md
dt      = 0.002 ; ps
nsteps  = 50000 ; 100ps
comm-grps = system
energygrps =
refcoord-scaling = com
;
nstxout = 0
nstvout = 0
nstfout = 0
nstlog  = 500
nstenergy = 500
nstxout-compressed = 1000
compressed-x-grps = system
;
continuation = yes
pbc = xyz
constraints = h-bonds
cutoff-scheme = Verlet
vdwtype = cutoff
vdw-modifier = force-switch
rlist = 1.2
rvdw = 1.2
rvdw-switch = 1.0
coulombtype = PME
rcoulomb = 1.2
DispCorr = no
;
Tcoupl = V-rescale
tc-grps          = Protein Non-Protein
tau_t            = 0.2    0.2
ref_t            = 298.15    298.15
;
Pcoupl = Berendsen
pcoupltype = isotropic
tau_p = 0.5
ref_p = 1.0
compressibility = 4.5e-5
;
; Velocity generation
gen_vel        = no    ; velocity generation off after NVT

```

The resulting edr file was used to extract the pressure and density profiles of the system, as shown in Figs. 4 and 5. Clearly, the pressure fluctuates significantly during the simulation, but the density value remains basically unchanged and is very close to the experimental value. This indicates that the system is initially equilibrated.

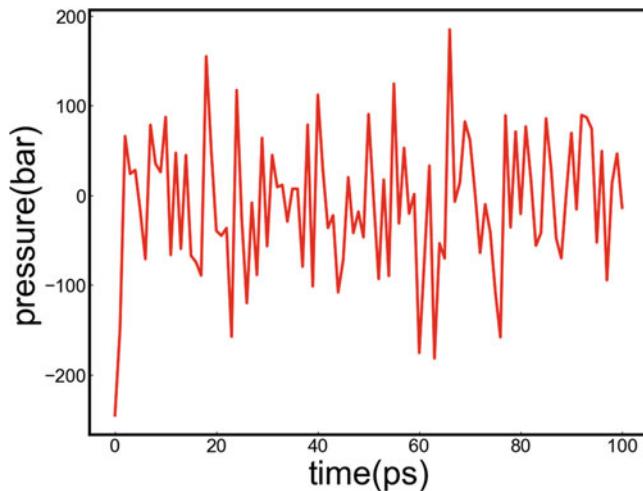


Fig. 4 The evolution of system pressure during NPT simulation

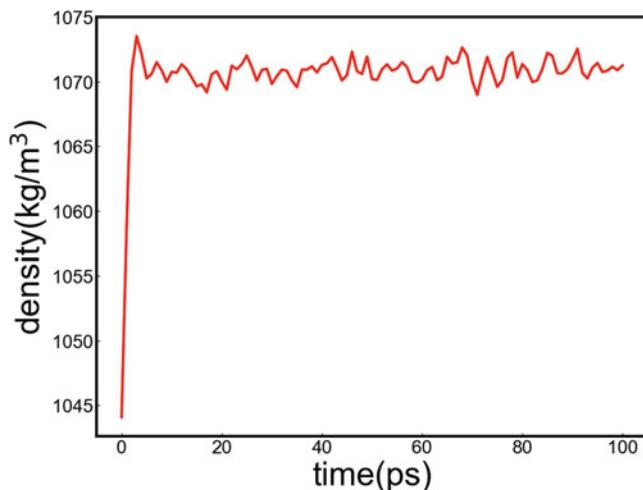


Fig. 5 The evolution of system density during NPT simulation

3.2.3 Equilibration Without Position Constraint

To achieve non-constrained equilibrium, the position constraint imposed on the protein was removed, and the simulation (5 ns) was performed using the NPT ensemble and the following command.

```
gmx grompp -f npt_nocstmdp -c npt.gro -r npt.gro -t npt.cpt -p
topol.top -o npt_nocst.tpr
gmx mdrun -v -deffnm npt_nocst
```

The content of the mdp file is as follows.

```

define =
integrator = md
dt      = 0.002 ; ps
nsteps  = 2500000 ; 100ns
;comm-grps = protein
;comm-mode = angular
energygrps =
;
nstxout = 0
nstvout = 0
nstfout = 0
nstlog  = 10000
nstenergy = 5000
nstxout-compressed = 5000
compressed-x-grps = system
;
continuation = yes
pbc = xyz
constraints = h-bonds
cutoff-scheme = Verlet
vdwtype = cutoff
vdw-modifier = force-switch
rlist = 1.2
rvdw = 1.2
rvdw-switch = 1.0
coulombtype = PME
rcoulomb = 1.2
DispCorr = no
;
Tcoupl = V-rescale
tc-grps      = Protein Non-Protein
tau_t        = 0.2    0.2
ref_t        = 298.15    298.15
;
Pcoupl = parrinello-rahman
pcoupltype = isotropic
tau_p = 2.0
ref_p = 1.0
compressibility = 4.5e-5
;
; Velocity generation
gen vel      = no

```

3.3 Production Simulation

The production simulation was conducted using the equilibrated gro and cpt files as input, and the same parameters used in equilibration simulation without pos constraint were applied, except for the simulation time (20 ns for production simulation). The trajectory obtained at the end was used for further analysis. The simulation was run using the following command.

```
gmx grompp -f mdmdp -c npt_nocst.gro -r npt_nocst.gro -t
npt_nocst.cpt -p topol.top -o md.tpr
gmx mdrun md
```

The content of the mdp file is as follows.

```
define =
integrator = md
dt      = 0.002 ; ps
nsteps   = 10000000 ; 100ns
;comm-grps = protein
;comm-mode = angular
energygrps =
;
nstxout = 0
nstvout = 0
nstfout = 0
nstlog  = 10000
nstenergy = 5000
nstxout-compressed = 5000
compressed-x-grps = system
;
continuation = yes
pbc = xyz
constraints = h-bonds
cutoff-scheme = Verlet
vdwtype = cutoff
vdw-modifier = force-switch
rlist = 1.2
rvdw = 1.2
rvdw-switch = 1.0
coulombtype = PME
rcoulomb = 1.2
DispCorr = no
;
Tcoupl = V-rescale
tc-grps      = Protein Non-Protein
tau_t        = 0.2    0.2
ref_t        = 298.15    298.15
;
Pcoupl     = parrinello-rahman
pcoupltype = isotropic
tau_p       = 2.0
ref_p       = 1.0
compressibility = 4.5e-5
;
; Velocity generation
gen_vel     = no
```

3.4 Analysis of MD Trajectories

3.4.1 RMSD

In this section, we explain how the production simulation trajectories were analyzed to obtain certain information on cage proteins.

In GROMACS, RMSD was calculated using the following command. The option 4 “Backbone” was chosen for both least-squares fitting and RMSD calculation.

```
gmx rms -s em.gro -f md.xtc -o md_rmsd.xvg -tu ns
```

It should be noted that in some cases, a step jump was observed in the RMSD profile due to the jumping of the protein out of the box during the simulation. This problem was fixed by handling the PBC using the following command. When prompted, group 1 “Protein” was chosen.

```
gmx trjconv -s md.tpr -f md.xtc -o md_nopbc.xtc -pbc mol -center
```

The fixed trajectory md_nopbc.xtc was used to recalculate the RMSD and obtain a continuous profile. Figure 6 shows the temporal evolution profile of RMSD during the simulation. The deviation of backbone atoms with respect to the reference crystal structure is about 0.17 nm and it is stable throughout the simulation, which suggests that the cage protein structure is stable.

3.4.2 RMSF

The following command was used in GROMACS to calculate the RMSF of apo-rHLFr. When prompted, the group 1 “Protein” option was chosen.

```
gmx rmsf -s em.gro -f md.xtc -o md_rmsf.xvg -res
```

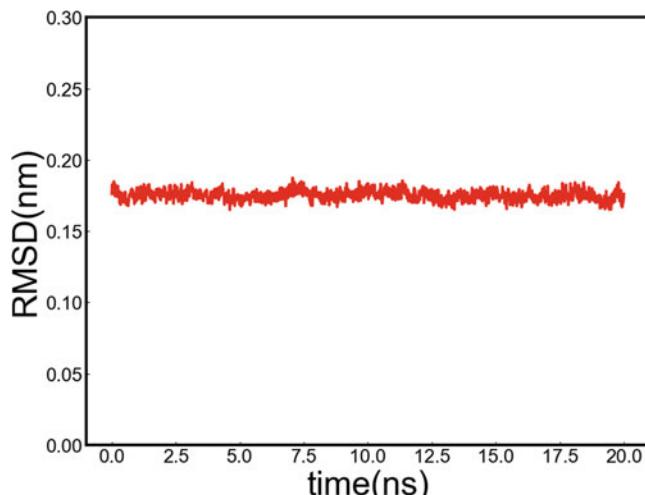


Fig. 6 Variation of apo-rHLFr RMSD as a function of simulation time

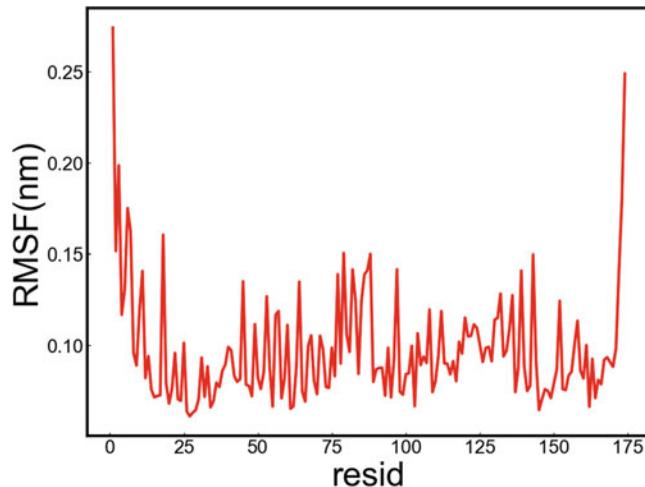


Fig. 7 The RMSF of apo-rHLFr residues during the simulation

As shown in Fig. 7, residues 6, 18, 79, and 143 exhibit the greatest fluctuation during the simulation.

3.4.3 Radius of Gyration

R_g was calculated in GROMACS using the following command and the group 1 “Protein” option.

```
gmx gyrate -f md.xtc -s md.tpr -o md_rg.xvg
```

The results shown in Fig. 8 demonstrate that the R_g of apo-rHLFr is 5.3 nm, and it is stable throughout the simulation.

3.4.4 Radial Distribution Function

The command for calculating RDF in GROMACS is as follows,

```
gmx rdf -f md.xtc -s md.tpr -o md_rdf.xvg -ref 'com of group 1'  
-sel 'group 1'
```

where -ref ‘com of group 1’ sets the reference coordinates, and -sel ‘group 1’ sets the atom group for computing RDF. The results illustrated in Fig. 9 suggest that the inner and outer radii of the apo-rHLFr cage are 3.5 and 6.5 nm, respectively, and that the thickness of the cage shell is 2.5 nm.

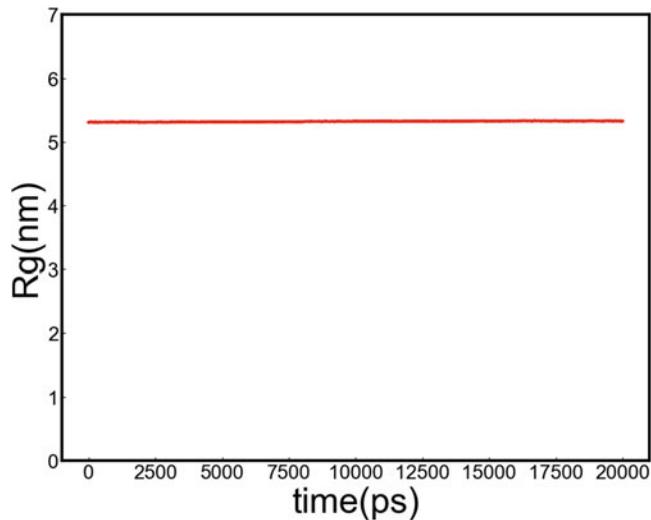


Fig. 8 Variation of apo-rHLFr Rg as a function of simulation time

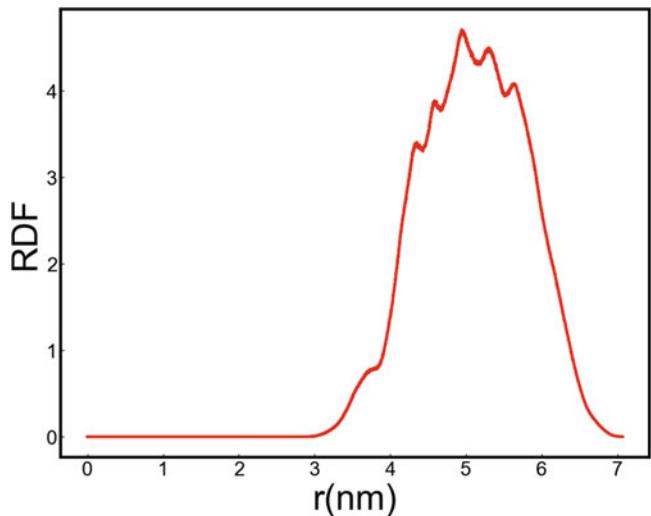


Fig. 9 Variation of apo-rHLFr RDF during simulation

3.4.5 Custom Physical Properties

As an example of how MD simulations can be used to determine custom physical properties of cage proteins, the Tcl script in VMD was applied to calculate the change in the number of water molecules during the simulation.

```

set protein [atomselect top "protein"]
set center [measure center $protein]
set center_x [lindex $center 0]
set center_y [lindex $center 1]
set center_z [lindex $center 2]
set sel [atomselect top "water and ((x-$center_x)^2+(y-$center_y)^2+(z-$center_z)^2)<50^2"]
set result [open water_num.txt w]
for {set i 1} {$i<=2002} {incr i 1} {
$protein frame $i
$protein update
$sel frame $i
$sel update
set num [expr [$sel num]/3]
puts $result "$i $num"
}
close $result

```

First, the simulation trajectory was loaded; then, the water_num.tcl file shown above was created in the same directory. Finally, the Tk console was opened and the following command was entered.

```
source water_num.tcl
```

The resulting water_num.txt file created in the working directory was used to analyze the change in the number of water molecules. As shown in Fig. 10, there are around 11,000 water molecules in apo-rHLFr, and the number remains stable throughout the simulation. Based on these results, it may be concluded that at 298.15 K, apo-rHLFr can hold up to 11,000 water molecules.

4 MD Simulation with NAMD

4.1 Preparation of the System

In this demonstration, we show how MD simulations can be conducted in NAMD, a software that makes use of the VMD graphics program to setup the simulation. The following Tcl script commands were used to generate the topology files, and the buildpsf.tcl file was opened using a text editor. As shown in the content presented below, the script calls the Charmm36 force field in the working directory to generate the topology and correct the protonation states of Glu130 and GLU137 to GLUP.

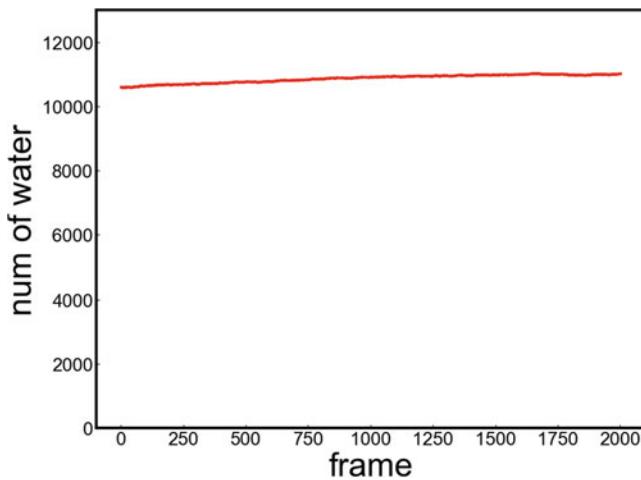


Fig. 10 Variation in the number of water molecules inside the apo-rHLFr cavity during the simulation

```

package require psfgen
topology ./toppar/top_all36_prot.rtf
pdbalias residue HIS HSE
pdbalias atom ILE CD1 CD
pdbalias atom HOH O OH2
pdbalias residue HOH TIP3

mol new 1dat_24mer.pdb waitfor all
set protein [atomselect top protein]
set chains [lsort -unique [$protein get chain]]

foreach chain $chains {
    set sel [atomselect top "chain $chain"]
    $sel writepdb 1dat_${chain}.pdb
    segment $chain {
        auto none
        pdb 1dat_${chain}.pdb
    }
    coordpdb 1dat_${chain}.pdb $chain
    patch GLUP $chain:130
    patch GLUP $chain:137
    regenerate angles dihedrals
}

guesscoord
writepdb 1dat_24mer_psfgen.pdb
writepsf 1dat_24mer_psfgen.psf

```

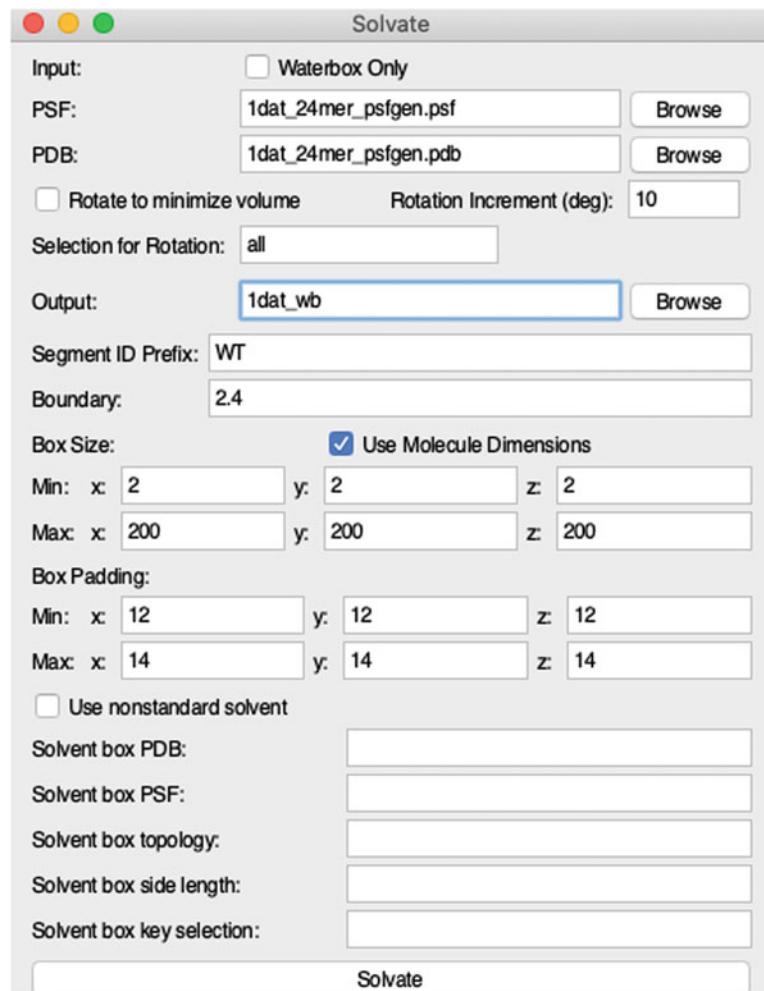


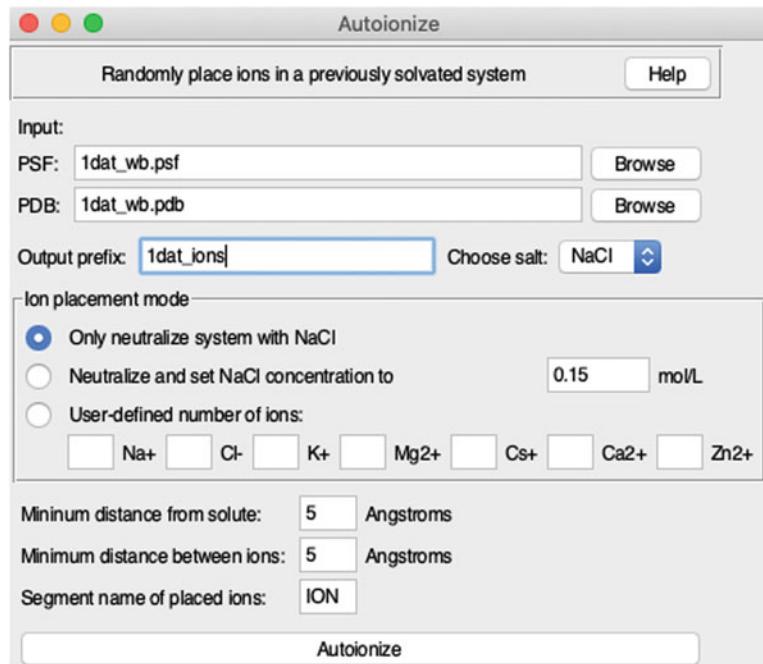
Fig. 11 Solvate Plugin

The script was executed by entering the following command in the Tk console of VMD.

```
source buildpsf.tcl
```

After running the script, the 1dat_24mer_psfgener.psf and 1dat_24mer_psfgener.pdb files were loaded into VMD, and the Solvate Plugin was used to generate a water box and solvate it based on the following settings (*see Fig. 11*).

The Autoionize Plugin was used to neutralize the system with NaCl, as per the following settings (*see Fig. 12*).

**Fig. 12** Autoionize Plugin

4.2 Equilibration

4.2.1 Energy Minimization

In NAMD, the simulation parameters are controlled through the configuration file. The following configuration file was used for energy minimization.

```

structure      ./1dat_ions.psf
coordinates    ./1dat_ions.pdb
set temperature 0
set outputname 1dat_em

firsttimestep 0
paraTypeCharmm on
parameters     ./toppar/par_all36_prot.prm
parameters     ./toppar/toppar_water_ions.str
parameters     ./toppar/par_all36_prot_mod.prm
parameters     ./toppar/par_all36_lipid.prm
parameters     ./toppar/par_all36_cgenff.prm
parameters     ./toppar/par_all36_carb.prm
temperature    $temperature
exclude        scaled1-4

```

```

1-4scaling      0.5
switching       on
cutoff          12.0
switchdist     12.0
pairlistdist   14.0
stepspercycle  20
timestep        2.0
rigidBonds     all
nonbondedFreq  1
vdwGeometricSigma yes
fullElectFrequency 2
pairlistsperCycle 2
cellBasisVector1 146.76700574159622 0 0
cellBasisVector2 0 146.7859971821308 0
cellBasisVector3 0 0 146.76499351859093
cellOrigin      73.1812744140625 73.15550231933594 73.01734161376953
PME            yes
PMEGridSpacing 1.0
wrapAll         on
outputName      $outputname
restartfreq    500
dcdfreq        200
outputEnergies 100
outputPressure 100
binaryoutput    no
binaryrestart   no
minimize       10000

```

4.2.2 Constrained Pre-equilibration

The configuration file used to perform the NVT simulation and raise the system temperature to the desired value is shown below.

```

structure      ./1dat_ions.psf
coordinates    ./1dat_em.restart.coor
extendedSystem ./1dat_em.restart.xsc
set temperature 298.15
set outputname 1dat_heat

firsttimestep 0
paraTypeCharmm on
parameters     ./toppar/par_all36_prot.prm
parameters     ./toppar/toppar_water_ions.str
parameters     ./toppar/par_all36_prot_mod.prm
parameters     ./toppar/par_all36_lipid.prm
parameters     ./toppar/par_all36_cgennff.prm
parameters     ./toppar/par_all36_carb.prm
exclude        scaled1-4
1-4scaling    0.5

```

```
cutoff      12.0
switching    on
switchdist   12.0
pairlistdist 14.0
timestep     2.0
rigidBonds   all
nonbondedFreq 1
vdwGeometricSigma yes
fullElectFrequency 2
stepspercycle 20
pairlistsperCycle 2
PME          yes
PMEGridSpacing 1.0
wrapAll      on
langevin      on
langevinDamping 1
langevinTemp   $temperature
langevinHydrogen off
fixedAtoms    on
fixedAtomsForces on
fixedAtomsFile 1dat_protein.fix
fixedAtomsCol  B
outputName    $outputname
restartfreq   500
dcdfreq       200
outputEnergies 100
outputPressure 100
binaryoutput   no
binaryrestart  no
seed          1010
numsteps      50000
temperature   0
reassignFreq   1
reassignIncr   0.01
reassignHold   $temperature
```

The NPT simulation used to increase the system pressure was conducted according to the following configuration file.

```

structure      ./1dat_ions.psf
coordinates    ./1dat_heat.restart.coor
extendedSystem ./1dat_heat.restart.xsc
velocities    ./1dat_heat.restart.vel
set temperature 298.15
set outputname 1dat_npt

firsttimestep 0
paraTypeCharmm on
parameters     ./toppar/par_all36_prot.prm
parameters     ./toppar/toppar_water_ions.str
parameters     ./toppar/par_all36_prot_mod.prm
parameters     ./toppar/par_all36_lipid.prm
parameters     ./toppar/par_all36_cgenff.prm
parameters     ./toppar/par_all36_carb.prm
exclude        scaled1-4
1-4scaling    0.5
cutoff         12.0
switching      on
switchdist     12.0
pairlistdist   14.0
timestep       2.0
rigidBonds    all
nonbondedFreq  1
vdwGeometricSigma yes
fullElectFrequency 2
stepspercycle  20
pairlistsperCycle 2
PME            yes
PMEGridSpacing 1.0
wrapAll        on
langevin        on
langevinDamping 1
langevinTemp   $temperature
langevinHydrogen off
useGroupPressure yes
useFlexibleCell no
useConstantArea no
langevinPiston  on
langevinPistonTarget 1.01325
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp $temperature
fixedAtoms      on
fixedAtomsForces on
fixedAtomsFile  1dat_protein.fix
fixedAtomsCol   B
outputName      $outputname
restartfreq    10000
dcdfreq        5000
outputEnergies  500
outputPressure  500
binaryoutput    no
binaryrestart   no
numsteps       50000

```

4.2.3 Equilibration Without Position Constraint The configuration file used for non-constrained equilibration is shown below.

```

structure      ./1dat_ions.psf
coordinates    ./1dat_npt.restart.coor
extendedSystem ./1dat_npt.restart.xsc
velocities     ./1dat_npt.restart.vel

set temperature 298.15
set outputname 1dat_npt_noest
firsttimestep 0
paraTypeCharmm on
parameters     ./toppar/par_all36_prot.prm
parameters     ./toppar/toppar_water_ions.str
parameters     ./toppar/par_all36_prot_mod.prm
parameters     ./toppar/par_all36_lipid.prm
parameters     ./toppar/par_all36_cgenff.prm
parameters     ./toppar/par_all36_carb.prm
exclude        scaled1-4
1-4scaling    0.5
cutoff         12.0
switching      on
switchdist     12.0
pairlistdist   14.0
timestep       2.0
rigidBonds     all
nonbondedFreq  1
vdwGeometricSigma yes
fullElectFrequency 2
stepspercycle  20
pairlistsperCycle 2
PME            yes
PMEGridSpacing 1.0
wrapAll        on
langevin        on
langevinDamping 1
langevinTemp   $temperature
langevinHydrogen off
useGroupPressure yes
useFlexibleCell no
useConstantArea no
langevinPiston  on
langevinPistonTarget 1.01325
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp $temperature
outputName      $outputname
restartfreq    10000
dedfreq        5000
outputEnergies  500
outputPressure  500
binaryoutput    no
binaryrestart   no
numsteps       2500000

```

4.3 Production Simulation

The production simulation was conducted using the equilibrated coor and vel files as input. The configuration file used to perform a 20-ns production simulation is shown below. The obtained trajectory was used for further analysis.

```
structure      ./1dat_ions.psf
coordinates    ./1dat_npt_nosct.restart.coor
extendedSystem ./1dat_npt_nosct.restart.xsc
velocities     ./1dat_npt_nosct.restart.vel

set temperature 298.15
set outputname 1dat_md
firsttimestep 0
paraTypeCharmm on
parameters     ./toppar/par_all36_prot.prm
parameters     ./toppar/toppar_water_ions.str
parameters     ./toppar/par_all36_prot_mod.prm
parameters     ./toppar/par_all36_lipid.prm
parameters     ./toppar/par_all36_cgenff.prm
parameters     ./toppar/par_all36_carb.prm
exclude        scaled1-4
1-4scaling    0.5
cutoff         12.0
switching      on
switchdist     12.0
pairlistdist   14.0
timestep       2.0
rigidBonds    all
nonbondedFreq  1
vdwGeometricSigma yes
fullElectFrequency 2
stepspercycle 20
pairlistsperCycle 2
PME            yes
PMEGridSpacing 1.0
wrapAll        on
langevin        on
langevinDamping 1
langevinTemp   $temperature
langevinHydrogen off
useGroupPressure yes
useFlexibleCell no
useConstantArea no
langevinPiston  on
langevinPistonTarget 1.01325
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp $temperature
outputName     $outputname
XSTfile        $outputname.xst
restartfreq    10000
dcdfreq        5000
XSTfreq        5000
outputEnergies 5000
outputPressure 5000
binaryoutput    no
binaryrestart   no
numsteps       10000000
```

References

1. Uchida M, Klem MT, Allen M et al (2007) Biological containers: protein cages as multi-functional nanoplatforms. *Adv Mater* 19: 1025–1042
2. Zhao M, Ho H, Wu Y et al (2014) Western flower thrips (*Frankliniella occidentalis*) transmits maize chlorotic mottle virus. *J Phytopathol* 162:532–536
3. Shenton W, Mann S, Cölfen H et al (2001) Synthesis of nanophase iron oxide in lumazine synthase capsids. *Angew Chem Int Ed* 40:442–445
4. Kilcoyne SH, Cywinski R (1995) Ferritin: a model superparamagnet. *J Magn Magn Mater* 140:1466–1467
5. Grant RA, Filman DJ, Finkel SE et al (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* 5:294–303
6. Ma Y, Nolte RJM, Cornelissen JJLM (2012) Virus-based nanocarriers for drug delivery. *Adv Drug Deliv Rev* 64:811–825
7. Molino NM, Wang S-W (2014) Caged protein nanoparticles for drug delivery. *Curr Opin Biotechnol* 28:75–82
8. Harrison PM, Arosio P (1996) The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim Biophys Acta (BBA)-Bioenerg* 1275:161–203
9. Chasteen ND, Harrison PM (1999) Mineralization in ferritin: an efficient means of iron storage. *J Struct Biol* 126:182–194
10. Douglas T, Dickson DPE, Betteridge S et al (1995) Synthesis and structure of an iron (III) sulfide-ferritin bioinorganic nanocomposite. *Science (80-)* 269:54–57
11. Mann S, Meldrum FC (1991) Controlled synthesis of inorganic materials using supramolecular assemblies. *Adv Mater* 3:316–318
12. Mann S, Archibald DD, Didymus JM et al (1993) Crystallization at inorganic-organic interfaces: biominerals and biomimetic synthesis. *Science (80-)* 261:1286–1292
13. Varpness Z, Peters JW, Young M, Douglas T (2005) Biomimetic synthesis of a H₂ catalyst using a protein cage architecture. *Nano Lett* 5: 2306–2309
14. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267:585–590
15. Mayor U, Guydosh NR, Johnson CM et al (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* 421:863–867
16. Dror RO, Mildorf TJ, Hilger D et al (2015) Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science (80-)* 348: 1361–1365
17. Tajkhorshid E, Nollert P, Jensen MØ et al (2002) Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science (80-)* 296:525–530
18. Plattner N, Noé F (2015) Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun* 6:1–10
19. Kubas A, Orain C, De Sancho D et al (2017) Mechanism of O₂ diffusion and reduction in FeFe hydrogenases. *Nat Chem* 9:88–95
20. Sanbonmatsu KY, Tung C-S (2007) High performance computing in biology: multimillion atom simulations of nanoscale systems. *J Struct Biol* 157:470–480
21. Lu D, Wang H, Chen M et al (2021) 86 PFLOPS Deep Potential Molecular Dynamics simulation of 100 million atoms with ab initio accuracy. *Comput Phys Commun* 259:107624
22. Shaw DE, Deneroff MM, Dror RO et al (2008) Anton, a special-purpose machine for molecular dynamics simulation. *Commun ACM* 51:91–97
23. Shaw DE, Grossman JP, Bank JA et al (2014) Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In: SC'14: proceedings of the international conference for high performance computing, networking, storage and analysis. IEEE, pp 41–53
24. Jung J, Nishima W, Daniels M et al (2019) Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *J Comput Chem* 40:1919–1930
25. Shaw DE (2013) Millisecond-long molecular dynamics simulations of proteins on a special-purpose machine. *Biophys J* 104:45a
26. Peng T, Lee H, Lim S (2012) Isolating a trimer intermediate in the self-assembly of E2 protein cage. *Biomacromolecules* 13:699–705
27. Laghaei R, Evans DG, Coalson RD (2013) Metal binding sites of human H-chain ferritin and iron transport mechanism to the ferroxidase sites: a molecular dynamics simulation study. *Proteins Struct Funct Bioinformatics* 81:1042–1050
28. Sala D, Ciambellotti S, Giachetti A et al (2017) Investigation of the iron (II) release mechanism of human H-ferritin as a function of pH. *J Chem Inf Model* 57:2112–2118

29. Ode H, Nakashima M, Kitamura S et al (2012) Molecular dynamics simulation in virus research. *Front Microbiol* 3:258
30. Peng X, Lu C, Liu Z, Lu D (2021) The synergistic mechanisms of apo-ferritin structural transitions and Au (iii) ion transportation: molecular dynamics simulations with the Markov state model. *Phys Chem Chem Phys* 23: 17158
31. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91:43–56
32. Case DA, Cheatham TE III, Darden T et al (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688
33. Phillips JC, Hardy DJ, Maia JDC et al (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* 153: 44130
34. Brooks BR, Brooks CL III, Mackerell AD Jr et al (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614
35. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys* 117:1–19
36. Jumper J, Evans R, Pritzel A et al (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:1–11
37. Baek M, DiMaio F, Anishchenko I et al (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (80-) 373:871–876
38. Krissinel E (2015) Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res* 43:W314–W319
39. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7:525–537
40. Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 40:W537–W541
41. Onufriev AV, Izadi S (2018) Water models for biomolecular simulations. *Wiley Interdiscip Rev Comput Mol Sci* 8:e1347
42. Onufriev A (2008) Implicit solvent models in molecular dynamics simulations: a brief overview. *Annu Rep Comput Chem* 4:125–137
43. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* 32:2319–2327
44. Gowers RJ, Linke M, Barnoud J et al (2019) MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. Los Alamos National Lab. (LANL), Los Alamos
45. McGibbon RT, Beauchamp KA, Harrigan MP et al (2015) MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys J* 109:1528–1532
46. Humphrey W, Dalke A, Schulter K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38



Chapter 17

Coarse-Grained Models for Vault Normal Model Analysis

D. Vijay Anand, Ronald Koh Joon Wei, and Kelin Xia

Abstract

Recent experiments have shown that the molecular complex of vault has large conformational changes at its shoulder and cap regions in solution. From the comparison of two configuration structures, it has been found that the shoulder region can twist and move outward, while the cap region will rotate and push upward correspondingly. To further understand these experimental results, in this paper, we study the vault dynamics for the first time. Since vault has an extremely large-sized structure with around 63,336 C_α atoms, traditional normal mode method with the C_α coarse-grained representation will fall short. We employ a newly invented multiscale virtual particle-based anisotropic network model (MVP-ANM). To reduce the complexity, the 39-folder vault structure is coarse-grained to about 6000 virtual particles, which significantly reduces the computational cost while still maintaining the basic structure information. Among the 14 low frequency eigenmodes from Mode 7 to Mode 20, two eigenmodes, i.e., Mode 9 and Mode 20, are found to be directly associated with the experimental observations. In Mode 9, shoulder region undergoes a significant expansion while the cap part is lifted upward. In Mode 20, a clear rotation of both shoulder and cap regions is well observed. Our results are consistent with the experimental observations. More importantly, these low frequency eigenmodes indicate that the vault waist, shoulder and lower cap regions are the most likely regions for the opening of the vault particle. And the opening mechanism is highly likely to be rotation and expansion at these regions. As far as we know, this is the first work to provide the normal mode analysis for the vault complex.

Key words Vault, Normal mode, Anisotropic network model, Eigenmodes, Conformation

1 Introduction

The elucidation of the dynamics of the giant protein cage vault is of immense interest to both scientists and engineers. The vault particles, which naturally occur in most eukaryotic organisms, might prove to be a viable platform for the transport of molecular cargo in nanomedicine [6, 13]. These nanosized particles possess many attributes such as high stability, biocompatibility, and suitable pharmaco-kinetics, that can be exploited in targeted drug delivery. The incorporation of vault associated proteins into its interior by a dynamic opening and closing mechanism has been the focus of intense research [14, 24, 25]. Uncovering the deformation

dynamics of vaults is central to the understanding of its biological functions. It has been found that recombinant vaults can be suitably engineered to allow protein packaging and cell targeting [35], indicating that vaults are not rigid impenetrable cages, but a fluctuating dynamic biomolecular complex with considerable flexibility. Further, the fact that vaults could act as a cellular transporter implies that these particles might possess a highly regulated opening mechanism in order to efficiently incorporate and deliver the specific molecular cargo with minimal toxicity [5]. Recently, two near-atomic-resolution configurations of vault in solution are obtained [11]. Through careful comparison, it has been found that the shoulder region in configuration 2 has larger conformational change, i.e., it has twisted and moved outward. More specifically, the R8 and R9 domains of the shoulder region undergoes a slightly outward bending. Simultaneously, the cap region has rotated clockwise by 2° and lifted by 10 Å [11]. These observations provide the dynamic information of vault in solution and shed light on the vault opening mechanism. Despite these experimental results, the precise opening mechanism and dynamic information of vault still remain elusive. Theoretically, statistical mechanics and molecular dynamics are the most widely used models in biomolecular dynamic and function analysis. However, with the large-sized macro-proteins and biomolecular complexes, the computational limitations of these approaches begin to reveal. To overcome the problem, normal mode analysis (NMA), especially elastic network models (ENMs) which uses a coarse-grained representation and harmonic function-based potential function, is proposed [2–4, 10, 15, 17, 19, 30]. Despite its great simplicity, the low-frequency normal modes from ENMs can efficiently capture the functionally important transition pathways of large-sized biomolecules or biomolecular complexes [10, 26]. Based on ENM, various models, including rotation-translation of blocks (RTB) [12], block normal mode (BNM) algorithm [17, 29], essential dynamics coarse graining (ED-CG) [36], NMA-based fluctuation matching method (NMA-FM) [8, 9, 32], iterative matrix projection method [22, 23]. Quantized elastic deformation model (QEDM) [7, 20, 21, 31], bend-twist-stretch (BTS) model [27], etc. [18, 20, 28, 31, 36] are proposed using different coarse-grained representation and new potential functions. These models can be used to handle even larger structures and Cryo-EM structure data. Essentially, it has been found that biomolecular conformational changes or collective global motions are largely determined by their global shapes and mass distribution [18, 20, 28, 31]. Based on the tight connection between structure global geometry and its normal modes, the multiscale virtual particle-based elastic network models (MVP-ENM) [34], including MVP-based Gaussian network model (MVP-GNM) and MVP-based anisotropic network model (MVP-ANM) are

proposed. The essential idea of MVP model is to construct a multiscale discretization model. In it, a suitable virtual particle representation that significantly reduces the computational cost and still captures the general shape of the structure, can be achieved. The MVP-ENMs have been successfully used in B-factor prediction and collective modes prediction [34]. In this paper, we apply the MVP-ANM in the normal mode analysis of the 39-folder vault. This vault structure has around 63,336 amino acids. If we coarse grain each amino acid into its C_α atom as in traditional normal mode methods, the computational cost is prohibitively high. Therefore, we consider the multiscale virtual particle representation. More specifically, to reduce the complexity, the vault structure is transformed into a density function and then coarse grained into around 6000 virtual particles (*see Note 1*). In this way, we can significantly reduce the computational cost while still preserving the basic structure information, that is required in the normal mode analysis. To analyze the long-time dynamic information, we only consider the first 14 lowest frequency eigenmodes, i.e., Mode 7 to Mode 20 (*see Note 2*). Among them, two eigenmodes, i.e., Mode 9 and Mode 20, are found to be directly associated with the conformational changes in shoulder and cap regions as observed in the recent experiments [11]. The expansion of the shoulder region together with the lift of the cap part is observed in Mode 9. The rotation of both shoulder and cap regions is found in Mode 20. Moreover, a careful study of all the 14 lowest frequency eigenmodes shows that the vault waist, shoulder, and lower cap regions are the most likely regions for the opening of the vault particle. And the opening mechanism is highly likely to be rotation and expansion at these regions (*see Note 3*).

2 Materials

Multiscale Representation of Vault The vaults are typically large barrel-shaped ribonucleoprotein complexes with a size of $41 \times 41 \times 72.5 \text{ nm}^3$. They have two caps at their ends and a hinged waist region. The barrel-shaped region between the caps is shoulder and body of the vault particle [16, 24]. Figure 1 illustrates the basic structure components, i.e., cap, shoulder, body, and waist, of a vault complex. Biologically, the vault is composed of several highly conserved proteins, including major vault proteins (PDB IDs 2QZV, 4 V60), which constitutes over 70% of the overall mass of the nanoparticle, vault poly(ADP-ribose) polymerase, telomerase associated protein 1 (TEP1) and some untranslated RNA molecules [1]. Despite the strong binding between major vault proteins, the vault has a dynamic structure that opens and closes transiently, a phenomenon referred to as breathing, to incorporate small

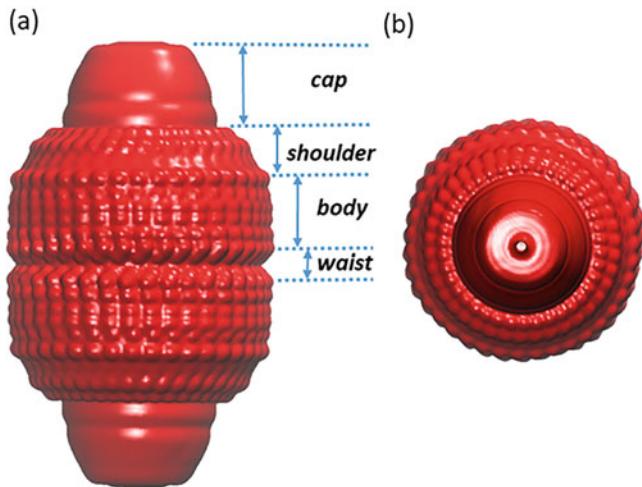


Fig. 1 The basic structure of a vault complex. Each vault is composed of several structure components, including cap, shoulder, body and waist. The vault density data illustrated are generated by using the generalized Gaussian model in Eq. 2 with a scale parameter of $\sigma = 8.0$ and $\kappa = 2$. The surface is generated by using an isovalue of 0.5. The side view and top view are shown in (a) and (b)

molecules, proteins, and other macromolecules into the inner core. In this study, the vault structure with PDB ID 4 V60 is considered. It has a dimer of half-vaults, with each half-vault comprising 39 identical major vault protein chains. Each major vault protein monomer folds into 12 domains: nine structural repeat domains, a shoulder domain, a cap-helix domain, and a cap-ring domain. The symmetry conditions are used to generate the complete structure of the vault complex, which is comprised of 63,336 amino acids. In our MVP model, a vault density data is generated by using the generalized Gaussian model in Eq. 2 with $\kappa = 2$ and a scale resolution value of $\sigma = 8 \text{ \AA}$ (*see* Methods for details). We only use the C_α atom. The grid spacing of 4 \AA is chosen to ensure the finer structural details are preserved. The vault surface in Fig. 1 is generated using isovalue 0.5. It can be noticed that the general shape of the vault structure is well captured. In our MVP model, if we define each voxel with a density value larger or equal to 0.5 as a virtual particle, the number of virtual particles will be too large. Therefore, we combine the adjacent four voxels in each direction together into a single virtual particle. This agglomeration of 64 voxels into a single virtual particle enables the vault structure to be represented by only 6194 virtual particles. The scale parameter for defining the spring constant is chosen as $\eta = 32 \text{ \AA}$. The detailed information for virtual particles and elastic network model is discussed in the following section.

3 Methods

Multiscale virtual particle model-based anisotropic network model (MVP-ANM) is employed to analyze the collective motion of the vault. The MVP-ANM is based on multiscale virtual particle model and anisotropic network model. The essence of this model is to construct a suitable coarse-grained representation of the large-sized biomolecules and a specially designed elastic network to examine the normal modes. More details on MVP-ANM and its applications can be found in our earlier works [33, 34].

3.1 Multiscale Virtual Particle-Based Anisotropic Network Model (MVP-ANM)

The key idea of the multiscale virtual particle model is to coarse-grain a large-sized biomolecular data, either in all-atom representation or in density representation, into a series of virtual particles (VPs) [34]. The size and shape of the virtual particles can vary depending on the biomolecular data, so that a balance between computational cost and model accuracy can be achieved. Elastic network models, including Gaussian network model (GNM) and anisotropic network model (ANM), can be built among these VPs, with a specially designed spring parameter, which can take into consideration of the VP properties. Generally, multiscale virtual particle model is composed of two important parts, i.e., multiscale virtual particle representation and specially designed spring parameter-based network model.

3.1.1 Multiscale Representation of Biomolecules

The biomolecular structures are usually highly complicated and exhibit multiscale characteristics. As an efficient coarse-graining technique, MVP model can be used in both PDB-based point cloud data and EMDB-based density data. The biomolecular data from PDB are of atomic resolution, thus we can define the virtual particle as various different coarse-grained elements, such as amino acid, nucleic acid, amino acid groups, subdomains, motif, etc. One of the most common approaches is to use a C_α atom as a virtual particle to represent the entire amino acid. The biomolecular density data from EMDB often do not have atomic details. Thus, virtual particles are usually generated through various different discretization approaches, such as Cartesian grid, tetrahedron mesh, Voronoi cell, etc. [34]. In essence, the high dimensional biomolecular data can be efficiently transformed into a density information, which provides a series of geometric representations at various scales of interest. In other words, a multiscale rigidity function for a data set with a total of N entries with coordinates $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, can be expressed as,

$$\mu(r, \sigma) = \sum_j^N w_j \phi(\|\mathbf{r} - \mathbf{r}_j\|; \sigma) \quad (1)$$

where w_j is the j -th weight parameter. The parameter σ is the resolution or scale parameter. The function $\phi(\|r - r_j\|; \sigma)$ is a kernel, usually chosen as the generalized exponential function,

$$\phi(\|r - r_j\|; \sigma) = e^{-\left(\frac{\|r - r_j\|}{\sigma}\right)^k} \quad (2)$$

The scale parameter σ can be used to control the “resolution” of the generated biomolecular density data. A small-scale value (high resolution) can characterize the detailed local structure of a biomolecule, while a large scale value (low resolution) will result in a coarse representation, which may capture only the general geometry of a biomolecule. It has been found that the low-frequency normal modes of a biomolecule is largely determined by its general shape or global geometry [18, 20, 28, 31] which however is preserved in the multiscale representation of the biomolecular structure [33, 34]. In this way, we can employ the transformation to generate a continuous representation which can retain the global geometry of the structure.

3.1.2 Generalized Spring Parameter

The construction of a suitable elastic connectivity between any two virtual particles, is the most critical part of the MVP model. Since the VPs might differ in their shapes and sizes, an identical spring constant employed in conventional ENMs is no longer suitable. To this end, a generalized spring parameter that accounts for density distribution and spatial influence has been proposed. More specifically, for a system of N virtual particles with coordinates r_1, r_2, \dots, r_N , and domain regions $\Omega_1, \Omega_2, \dots, \Omega_N$, the spring parameter γ_{IJ} between two virtual particles with domains I and J can be expressed as follows,

$$\gamma_{IJ} = \gamma(r_I, r_J, \Omega_I, \Omega_J, \mu(r), \eta) = \gamma_1(\Omega_I, \Omega_J, \mu(r)) \cdot \gamma_2(r_I, r_J, \eta)$$

Here r_I and r_J are coordinates for the centers of the I -th and J -th virtual particles Ω_I and Ω_J , respectively. The function $\mu(r)$ can be density function, normalized density function $\mu^s(r)$ or multiscale rigidity function as in Eq. 1. The term γ_1 accounts for the mass or density distribution and term γ_2 represents the distance influence. The γ_1 term is designed to take into consideration the shape and size of the virtual particles. It can be expressed as,

$$\gamma_1(\Omega_I, \Omega_J) = \left(1 + \alpha \int_{\Omega_I} \mu^s(r) dr\right) \left(1 + \alpha \int_{\Omega_J} \mu^s(r) dr\right)$$

Here the parameter α is a weight coefficient. $\mu^s(r)$ is the density function which is scaled with the maximum value of density. The γ_2 term represents the distance contribution part. It can be chosen as any monotonically decreasing functions. For example, we can consider the generalized Gaussian kernels,

$$\gamma_2(r_I, r_J, \eta) = e^{-\left(\frac{\|r_I - r_J\|}{\eta}\right)^\kappa}, \kappa > 0. \quad (3)$$

It is worth mentioning that the value of the spatial resolution η depends on the multiscale virtual particle model.

3.1.3 Multiscale Virtual Particle-Based Anisotropic Network Model (MVP-ANM)

The MVP-ANM is used in modeling the normal modes of large biomolecular data from PDB and EMDB. More specifically, for a system of N virtual particles with coordinates $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, and domain regions $\Omega_1, \Omega_2, \dots, \Omega_N$, we denote

$$\mathbf{r}_I = (x_I, y_I, z_I), \quad \mathbf{r}_I^d = (x_I^d, y_I^d, z_I^d) \quad \text{for } I = 1, 2, \dots, N,$$

and $\Delta R = \{\Delta x_1, \Delta y_1, \Delta z_1, \dots, \Delta x_N, \Delta y_N, \Delta z_N\}$ with $\Delta x_I = x_I^d - x_I, \Delta y_I = y_I^d - y_I$ and $\Delta z_I = z_I^d - z_I$ for $I = 1, 2, \dots, N$. A potential function can be expressed as,

$$V^{\text{MVP-ANM}} = \frac{1}{2} \Delta R^T H^{\text{MVP-ANM}} \Delta R$$

Here $H^{\text{MVP-ANM}}$ is the Hessian matrix, which composes many local 3 by 3 off diagonal matrixes,

$$H_{IJ}^{\text{MVP-ANM}} = -\frac{\gamma_{IJ}}{r_{IJ}^2} \begin{bmatrix} (x_J - x_I)(x_J - x_I) & (x_J - x_I)(y_J - y_I) & (x_J - x_I)(z_J - z_I) \\ (y_J - y_I)(x_J - x_I) & (y_J - y_I)(y_J - y_I) & (y_J - y_I)(z_J - z_I) \\ (z_J - z_I)(x_J - x_I) & (z_J - z_I)(y_J - y_I) & (z_J - z_I)(z_J - z_I) \end{bmatrix}$$

$I \neq J$. Again $\gamma_{IJ} = \gamma(\mathbf{r}_I, \mathbf{r}_J, \Omega_I, \Omega_J, \mu^s(r), \eta^{\text{MVP}})$ is the spring parameter between virtual particles at r_I and r_J . The diagonal element is the negative summation of the off-diagonal elements in the same row,

$$H_{II}^{\text{MVP-ANM}} = -\sum_{I \neq J} H_{IJ}^{\text{MVP-ANM}}$$

MVP-ANM can also be used in the B-factor prediction [34]. However, the predicted B-factors are calculated on virtual particles, which may not always be an atom, thus an interpolation is usually needed.

3.2 Vault Dynamics Analysis with MVP-ANM

The MVP-ANM model is used in the vault dynamic analysis, particularly the normal mode analysis. Using the virtual particle representation, the complexity of vault structure can be reduced significantly while the general geometry can be well preserved for normal mode analysis. In normal mode analysis, the eigenvalues are ranked from the smallest to the largest. The first six eigenvalues are always 0 and the corresponding eigenvectors are the trivial eigenmodes, i.e., translation and rotation modes. Only lowest-frequency modes from Mode 7 have dynamic information that is potentially meaningful.

In our MVP-ANM-based vault dynamic analysis, we consider 14 lowest-frequency eigenmodes, i.e., Mode 7 to Mode 20. These

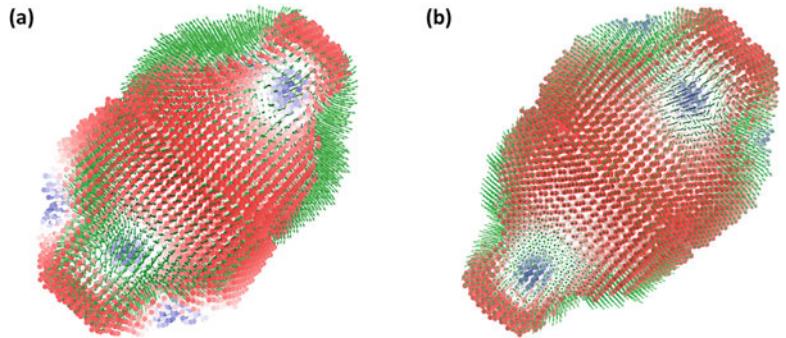


Fig. 2 The illustration of Mode 9 from the MVP-ANM modeling of vault. The eigenmode is represented by the green arrows. The virtual particle is represented by point painted with red-to-blue colormap. Blue means larger eigenmode amplitude, red indicates smaller value and white color represents the values in the middle range. Subfigures (a) and (b) represent two configuration states, i.e., contraction and expansion, from the normal mode simulation. It can be seen that the contraction and expansion dynamics are concentrated at the shoulder region

modes can be roughly grouped into two classes, i.e., the contraction-expansion modes and rotational modes. We have found that Mode 9 and Mode 20 are highly related to the recent experimental results.¹¹ More interestingly, the largest motions of the lowest-frequency eigenmodes all occur at the waist, shoulder and cap regions, indicating that these regions are potential areas that control the opening and closing of the vault complex. Two most important eigenmodes contraction-expansion type of motions widely exists in the symmetric structures with spherical geometry. Figure 2 demonstrates a contraction-expansion type of mode, observed from Mode 9. The blue and red colors indicate the amplitudes of the eigenmodes. It can be seen that the expansion and contraction happen in the shoulder regions. Subfigure (a) is the contraction of the upper shoulder and expansion of the lower shoulder. In contrast, subfigure (b) is the contraction of the lower shoulder and expansion of the upper shoulder. Moreover, the expansion of the shoulder region will induce the uplifting of the cap region. Although it is not exactly the same motion as found in the experiments, the expansion of the shoulder and lift of the cap region is consistent with the experimental results [11].

The other most important motion is observed from Mode 20 as in Fig. 3. Only one rotational configuration is demonstrated as the configuration changes are hard to show in pictures. However, four different views from two sides, top and bottom are shown in subfigures (a1) to (a4), respectively. To facilitate the visualization, we use black arrows to indicate the rotational directions. It can be

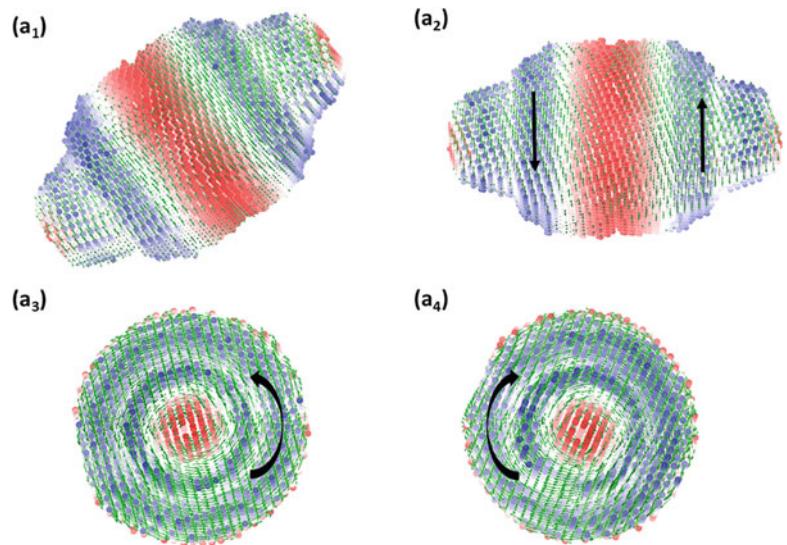


Fig. 3 The illustration of Mode 20 from the MVP-ANM modeling of vault. The eigenmode is represented by the green arrows. The virtual particle is represented by point painted with red-to-blue colormap. Blue means larger eigenmode amplitude, red indicates smaller value and white color represents the values in the middle range. Subfigures (a1) to (a4) are four different views of Mode 20. It can be seen that large rotation motion happens at the shoulder and cap regions

seen that the cap and shoulder regions have the largest motion (dark blue color) and the waist region has a relatively small rotational dynamics. Further, the upper and lower half-spheres are rotated in opposite directions, one clockwise and the other counter-clockwise. Biologically, this opposite rotational motions may induce the opening of the vault in the waist region. Moreover, the twist of the shoulder region is consistent with the experimental observations [11].

Other lowest-frequency eigenmodes: In our MVP-ANM results, the lowest-frequency eigenmodes are generally observed to be contraction-expansion and rotational type. Figure 4 illustrates the six distinct lowest-frequency modes of the vault complex. For each mode, two configurations, mostly the contraction and expansion states, from the normal mode simulations are depicted as subfigures (a) and (b). The six eigenmodes are Modes 7, 11, 13, 14, 16 and 18, denoted as subscripts 1 to 6, respectively. The green arrows represent the amplitudes and directions of the eigenmodes, the similar way as in the previous cases. Due to the symmetry of the vault structure, similar contraction-expansion motion is observed in pairs, i.e., as two modes. More specifically, the modes 7, 9, 11, 13, 17, and 19 is similar to modes 8, 10, 12, 15, 16, and 18, respectively. Modes 14 and 20 show no similar behavior as

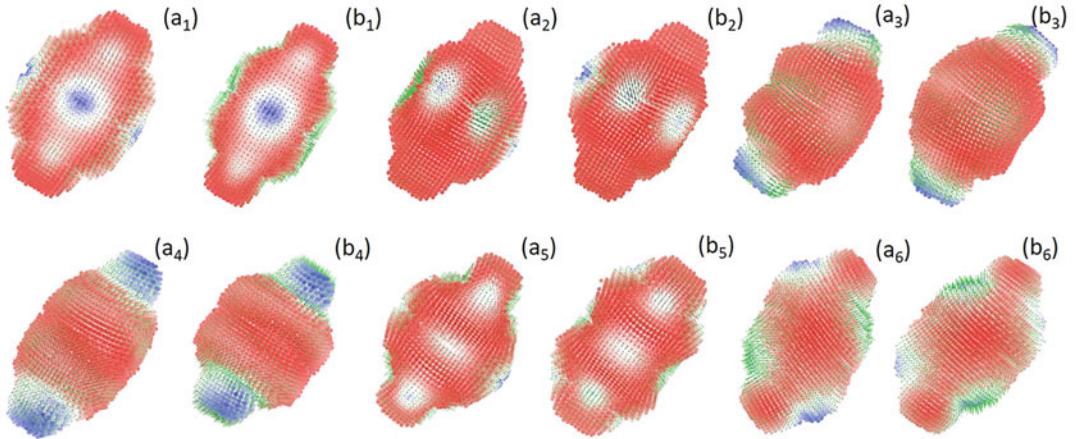


Fig. 4 The illustration of the six distinct lowest frequency eigenmodes of vault, i.e., Modes 7, 11, 13, 14, 16, and 18, denoted as subscripts 1 to 6, respectively. The green arrows represent the eigenmodes. The red and blue points represent virtual particles. The virtual particle is represented by point painted with red-to-blue colormap. Blue means larger eigenmode amplitude, red indicates smaller value and white color represents the values in the middle range. Subfigures (a) and (b) represent two configuration states, i.e., mostly contraction and expansion, from the normal mode simulation. It can be seen that the largest motions occur at only the cap, shoulder and waist regions

other eigenmodes. In this way, we consider the modes 7, 11, 13, 14, 16, and 18 for in Fig. 4.

The subfigures (a1) and (b1) for Mode 7 clearly depict the contraction-expansion type of mode. The waist region has a large deformation, i.e., bulges outwards in (a1) is retracted in (b1). A similar expansion mode but more asymmetrically along the body is noticed in Mode 11, i.e., in (a2) and (b2). The configurations observed in Mode 13 are bending which leads to compression of one side and expansion of the opposite side as seen in (a3) and (b3). The Mode 14 exhibits a stretching and compression motion in the subfigures (a4) and (b4), respectively, with a slight bending near the cap region. The bulge and retraction motion is again seen to occur in Mode 16 as shown in (a5) and (b5). However, the expansion is more spherical and the retraction motion forms a narrow waist region. The last configuration namely, Mode 18, i.e., (a6) and (b6), shows a pronounced motion of the shoulder region alone in opposite directions.

From all the above lowest-frequency eigenmodes, it can be seen that the largest deformations are concentrated in the cap, shoulder and waist regions, and are through contraction-expansion and rotation. This indicates that the most possible way of vault opening and closing is through these highly dynamic regions. Also, expansion-contraction and rotation are the most likely way of vault opening mechanism.

4 Notes

1. Due to the extremely-large size of vault, a density-based representation is considered. With the virtual particle model, the gigantic vault structure is reduced to around 6000 elements, which efficiently maintains the basic geometry of vault.
2. The 14 lowest-frequency eigenmodes from Mode 7 to Mode 20 are systematically studied. It has been found that Mode 9 and Mode 20 can be directly related to the recent experimental findings. Mode 9 shows the expansion of shoulder region together with the uplifting of cap part, while Mode 20 depicts the twist motion of shoulder and cap.
3. All lowest frequency eigenmodes have their largest variation in the waist, shoulder and cap regions, indicating that these regions are the potential part for opening of vault and the open mechanism is through expansion or rotation.

Acknowledgments

This work was supported in part by NTU-JSPS Fund M4082387, Nanyang Technological University Start-up Grant M4081842 and Singapore Ministry of Education Academic Research fund Tier 1 RG109/19 and Tier 2 MOE2018-T2-1-033.

References

1. Anderson DH, Kickhoefer VA, Sievers SA et al (2007) Draft crystal structure of the vault shell at 9 Å resolution. *PLoS Biol* 5:e318
2. Atilgan AR, Durrell SR, Jernigan RL et al (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 80:505–515
3. Bahar I, Atilgan AR, Demirel MC et al (1998) Vibrational dynamics of proteins: significance of slow and fast modes in relation to function and stability. *Phys Rev Lett* 80:2733–2736
4. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 2:173–181
5. Bhaskar S, Lim S (2017) Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Mater* 9(4):e371
6. Casanás A, Guerra P, Fita I et al (2012) Vault particles: a new generation of delivery nanodevices. *Curr Opin Biotechnol* 23:972–977
7. Chacon P, Tama F, Wriggers W (2003) Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. *J Mol Biol* 326:485–492
8. Chu JW, Voth GA (2005) Allostery of actin filaments: molecular dynamics simulations and coarsegrained analysis. *Proc Natl Acad Sci U S A* 102:13111–13116
9. Chu JW, Voth GA (2006) Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophys J* 90: 1572–1582
10. Cui Q, Bahar I (2010) Normal mode analysis: theory and applications to biological and chemical systems. Chapman and Hall/CRC
11. Ding K, Zhang X, Mrazek J et al (2018) Solution structures of engineered vault particles. *Structure* 26:619–626
12. Durand P, Trinquier G, Sanjouand YH (1994) A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers* 34:759–771
13. Goldsmith LE, Pupols M, Kickhoefer VA et al (2009) Utilization of a protein “shuttle” to

- load vault nanocapsules with gold probes and proteins. *ACS Nano* 3:3175–3183
14. Goldsmith LE, Yu M, Rome LH et al (2007) Vault nanocapsule dissociation into halves triggered at low pH. *Biochemistry* 46:2865–2875
 15. Hinsen K (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins* 33:417–429
 16. Kedersha NL, Rome LH (1986) Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J Cell Biol* 103:699–709
 17. Li GH, Cui Q (2002) A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to $\text{Ca}^{(2+)}$ -ATPase. *Biophys J* 83:2457–2474
 18. Lu MY, Ma JP (2005) The role of shape in determining molecular motions. *Biophys J* 89: 2395–2401
 19. Edward L, Jim P, Gregory AV (2008) Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophys J* 95:4183–4192
 20. Ming DM, Kong YF, Lambert MA et al (2002) How to describe protein motion without amino acid sequence and atomic coordinates. *Proc Natl Acad Sci U S A* 99:8620–8625
 21. Ming DM, Kong YF, Wakil SJ et al (2002) Domain movements in human fatty acid synthase by quantized elastic deformational model. *Proc Natl Acad Sci U S A* 99:7895–7899
 22. Na H, Jernigan RL, Song G (2015) Bridging between NMA and elastic network models: preserving all-atom accuracy in coarse-grained models. *PLoS Comput Biol* 11:e1004542
 23. Na H, Song G (2014) Bridging between normal mode analysis and elastic network models. *Proteins: Struct Funct Bioinform* 82:2157–2168
 24. Poderycki MJ, Kickhoefer KA, Kaddis CS et al (2006) The vault exterior shell is a dynamic structure that allows incorporation of vault-associated proteins into its interior. *Biochemistry* 45:12184–12193
 25. Querol-Audi J, Casanas A, Uson I (2009) The mechanism of vault opening from the high resolution structure of the N-terminal repeats of mvp. *EMBO J* 28:3450–3457
 26. Skjaerven L, Hollup SM, Reuter N (2009) Normal mode analysis for proteins. *THEOCHEM J Mol Struct* 898:42–48
 27. Stember JN, Wriggers W (2009) Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. *J Chem Phys* 131:074112
 28. Tama F, Brooks CL III (2006) Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct* 35:115–133
 29. Tama F, Gadea FX, Marques O et al (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* 41:1–7
 30. Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng* 14:1–6
 31. Tama F, Wriggers W, Brooks CL (2002) Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J Mol Biol* 321:297–305
 32. Xia F, Lu LY (2012) Multiscale coarse-graining via normal mode analysis. *J Chem Theory Comput* 8:4797–4806
 33. Anand DV, Meng Z, Xia K (2019) Complex multiscale virtual particle model based elastic network model (CMVP-ENM) for normal mode analysis of biomolecular complexes. *Phys Chem Chem Phys* 21:4359–4366
 34. Xia K, Li ZM, Mu L (2018) Multiscale persistent functions for biomolecular structure characterization. *Bull Math Biol* 80:1–31
 35. Yang J, Kickhoefer VA, Ng BC et al (2010) Vaults are dynamically unconstrained cytoplasmic nanoparticles capable of half vault exchange. *ACS Nano* 4:7229–7240
 36. Zhang Z, Voth GA (2010) Coarse-grained representations of large biomolecular complexes from low-resolution structural data. *J Chem Theory Comput* 6:2990–3002

Part IV

Translation of Protein Cages



Chapter 18

Evaluating Anti-tumor Immune Responses of Protein Nanoparticle-Based Cancer Vaccines

Enya Li, Nina Butkovich, Jo A. Tucker, Edward L. Nelson,
and Szu-Wen Wang

Abstract

Cancer vaccines displaying tumor-associated antigens (TAAs) train the immune system for enhanced tumor recognition and elimination. Nanoparticle-based cancer vaccines are ingested and processed by dendritic cells, which subsequently activate antigen-specific cytotoxic T cells, allowing them to identify and eliminate tumor cells displaying these TAAs. Here, we describe the procedures to conjugate TAA and adjuvant to a model protein nanoparticle platform (E2), followed by assessment of vaccine performance. Utilizing a syngeneic tumor model, the efficacy of *in vivo* immunization was determined by cytotoxic T lymphocyte assays and IFN- γ ELISpot *ex vivo* assays to measure tumor cell lysis and TAA-specific activation, respectively. *In vivo* tumor challenge directly allows evaluation of anti-tumor response and survival over time.

Key words Protein nanoparticle, Cancer vaccine, Tumor challenge, Cytotoxic T cells, Tumor-associated antigen, B16 melanoma model, Glycoprotein 100 (gp100)

1 Introduction

The efficacy of therapeutic cancer vaccines is dependent on the successful activation of the anti-tumor immune response, particularly of tumor antigen-specific CD8 T cells [1]. Effective delivery of antigens and activation of antigen presenting cells (APCs), such as dendritic cells and macrophages, is a critical step in this process. Caged protein nanoparticles possess several desirable qualities as delivery systems, such as: (1) stability, (2) size that is conducive to uptake by APCs, (3) geometries with multiple sites for conjugation to various immunomodulating compounds, (4) protection of sensitive compounds from degradation, and (5) high biocompatibility with low side effects [2]. By concurrently delivering a high dose of antigen and adjuvant, caged nanoparticles can effectively prime and

Enya Li and Nina Butkovich contributed equally to this chapter and are co-first authors.

activate APCs, which can carry them to the secondary lymphoid organs and subsequently elicit a strong anti-tumor immune response.

Cancer peptide vaccines aim at inducing T cell responses. Vaccines that use major histocompatibility complex (MHC) class I peptides are directed at the activation of CD8⁺ T cells [3]. CD8⁺ T cells are widely preferred for anti-tumor responses, as they mount a robust antigen-specific response, are capable of direct tumor cell killing, and form durable immunological memory [4]. These responses can be characterized in vivo using murine tumor models. Within these models, vaccine efficacy can be tested by the delay or prevention of tumor formation following immunization. Immune cell responses can be further assessed ex vivo or in vitro by collecting cells from the secondary lymphoid organs, the spleen, and draining lymph nodes. Bulk immune cells are examined for evidence of activation via immunophenotyping, cytokine release, and tumor cell lysis assays. However, tools that provide information at the single-cell level, such as enzyme-linked immunospot (ELISpot), have become the recommended methods of clinical monitoring [5].

The E2 protein nanoparticle is derived from the E2 subunit of pyruvate dehydrogenase of *Geobacillus stearothermophilus* and self-assembles into a dodecahedral cage structure composed of 60 identical subunits, allowing conjugation to the inner and outer surfaces, as well as between subunits [6]. Immunization with E2 conjugated to known immunogenic peptides and CpG, a TLR9 agonist, produced a strong cytotoxic T cell response when compared to unconjugated peptide and adjuvant [7–10]. In this chapter, we describe the assessment of E2 as a delivery system using the syngeneic B16 melanoma line as the model, with an immunogenic gp100 peptide as the antigen. The doses and immunization schedules within this chapter have been optimized for this system and may require adaptation for use with other models, peptides, or adjuvants. Although it is not within the scope of this chapter to outline combination therapy studies with the protein nanoparticle vaccines, the protocols described here establish the foundation for combination strategies and have been described elsewhere [9]. We will describe the conjugation of the E2 nanoparticle to form the vaccine, immunization schedule and methods, and assessment of the immune response, which includes cytotoxic T lymphocyte assay, interferon gamma (IFN- γ) ELISpot, and in vivo tumor challenge studies.

2 Materials

2.1 Antigen and Adjuvant Conjugation in D381C (E2)

1. D381C (E2 protein scaffold, with internal aspartate at site 381 mutated to cysteines). This protein nanoparticle and its expression, purification, and characterization are described by Dalmau et al. [6].

2. Phosphate buffer: 50 mM KH₂PO₄ and 100 mM NaCl.
3. TCEP-HCl (Tris[2-carboxyethyl]phosphine hydrochloride).
4. BMPH (N-β-maleimidopropionic acid hydrazide).
5. Zeba Spin Desalting Columns, 40K MWCO.
6. Sulfo-SMCC (sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate).
7. Adjuvant modified with aldehyde for conjugation (e.g., CpG1826 with 5' benzaldehyde).
8. Antigen peptide modified with N-terminal cysteine for conjugation (e.g., CSIINFEKL or CKVPRNQDWL).
9. Bicinchoninic acid assay (BCA) to quantify protein.

2.2 Cytotoxic T Lymphocyte (CTL) Lysis Assay

1. B16-F10 murine melanoma tumor cells (ATCC, CRL-6475).
2. Sterile media: RPMI 1640 with 10% FBS, 1 mM sodium pyruvate, 2 mM L-glutamine, 100 U/mL Pen/Strep, 50 mM 2-mercaptoethanol, and 0.1 mM nonessential amino acids.
3. Phosphate buffered solution (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄.
4. Gp100 as the relevant peptide (KVPRNQDWL) and an irrelevant peptide control (e.g., SIINFEKL).
5. Positive control, such as phytohemagglutinin, M form (PHA-M).
6. 70 µm sterile mesh cell strainers.
7. CytoTox 96 Non-Radioactive Cytotoxicity Assay (Promega). 10× lysis solution and stop solution are included in this kit.
8. 96-well clear flat bottom and round bottom culture plates.
9. Plate reader capable of measuring absorbance at 490 nm.
10. Hemocytometer and 0.4% trypan blue viability stain.

2.3 Interferon-Gamma (IFN-γ) ELISpot

1. Murine IFN-γ kit (e.g., R&D Systems, Cat. EL485), or individually purchased components including: substrate (e.g., BCIP/NBT substrate: 18.8 mg/mL nitro-blue tetrazolium chloride, 9.4 mg/mL 5-bromo-4-chloro-3-indolylphosphate toluidine salt, and 67% DMSO), avidin-horseradish peroxidase, detection antibody (biotinylated antibody for mouse IFN-γ), blocking buffer, coating buffer, capture antibody (for mouse IFN-γ), and 96-well PVDF ELISpot plates (*see Note 10*).
2. Sterile media: RPMI 1640 with 10% FBS, 1 mM sodium pyruvate, 2 mM L-glutamine, 100 U/mL Pen/Strep, 50 mM 2-Mercaptoethanol, and 0.1 mM nonessential amino acids.

3. SIINFEKL (or other irrelevant) peptide and antigen peptide of interest (e.g., gp100 KVPRNQDWL).
4. Positive control, such as phytohemagglutinin, M form (PHA-M).
5. Phosphate buffered solution (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄.
6. PBS-Tween: PBS with 0.1% Tween 20.
7. PBS-Tween-BSA: PBS with 0.1% Tween 20 and 1% BSA.
8. ELISpot reader.
9. Hemocytometer and 0.4% trypan blue viability stain.

2.4 Tumor Challenge

1. B16-F10 murine melanoma tumor cells (ATCC, CRL-6475).
2. Sterile media: DMEM with 10% FBS.
3. Calipers.
4. Insulin syringe and needle (0.3 mL syringe capacity, 31-gauge needle).
5. 25 cm² culture flask.
6. 6–8 week-old female C57BL/6 mice (The Jackson Laboratory).
7. Hemocytometer and 0.4% trypan blue viability stain.

3 Methods

3.1 Antigen and Adjuvant Conjugation to E2 Protein Nanoparticle

E2 protein from the pyruvate dehydrogenase complex of *Geobacillus stearothermophilus* was mutated with cysteines in the internal cavity (D381C) to enable chemical functionalization. Aldehyde-modified adjuvants can be covalently loaded into the interior of D381C, and cysteine-modified antigens can be conjugated to the exterior lysines [6].

1. Incubate D381C in phosphate buffer with ten-fold excess of TCEP for 30 min at room temperature (RT) to reduce the cysteines in the internal cavity.
2. Incubate the reduced D381C in phosphate buffer with ten-fold excess of N-(b-maleimidopropionic acid) hydrazide (BMPH) linker for 2 h at RT (*see Note 1*).
3. Remove the unreacted linker with (Zeba) spin desalting columns.
4. Add two-fold excess of aldehyde-modified adjuvant (e.g., CpG 1826 with 5' benzaldehyde) and let incubate overnight at RT (*see Note 1*).
5. Remove the unreacted adjuvant with desalting columns.

6. To conjugate peptide onto the exterior of D381C, add 20-fold excess of sulfo-SMCC linker that is freshly dissolved in water, and incubate for 30 min at RT.
7. Remove the unreacted sulfo-SMCC with desalting columns.
8. Reduce the cysteine-modified peptide (e.g., gp100 with N-terminal cysteine) with ten-fold excess of TCEP for at least 30 min at RT.
9. Add ten-fold excess of the reduced peptide to the sulfo-SMCC-functionalized D381C and incubate at RT for 2 h (*see Note 2*). Keep the sample at 4 °C for overnight incubation.
10. Remove the unreacted peptide with desalting columns.
11. Store the protein nanoparticle vaccines at 4 °C for short-term (up to a week) or –80 °C for longer term.
12. Confirm and quantify conjugation with dynamic light scattering (particle size), SDS-PAGE, BCA, and/or mass spectrometry.

3.2 CTL Lysis Assay

CTL lysis assay is performed to determine the antigen-specific lytic activity of splenocytes from immunized mice. It is recommended to test different target-to-effector cell ratios to understand the dose-dependent effect.

1. Immunize C57BL/6 mice subcutaneously with the vaccine construct from **step 3.1** (for example, CpG-gp100-E2, which is the E2 nanoparticle with encapsulated CpG and a gp100 peptide epitope on the surface). A typical dose is 5 µg/dose of peptide antigen (approximately 50 µg of the vaccine).
2. After 7 days (defined as day 0 of splenocytes culture), harvest the spleen and process it by mechanical digestion, mincing, and crushing the spleen in a 10 cm petri dish to separate the splenocytes from the pulp. As the tissue breaks apart, a red solution of cells will form, and white to cream colored connective tissue will be evident and removable. Pipette vigorously and rinse the dish with PBS and pass the solution through a 70 µm sterile mesh cell strainer to obtain a single cell suspension.
3. Centrifuge cells at $300 \times g$ for 5 min to pellet the splenocytes. Remove supernatant and lyse the red blood cells by resuspending in 10 mL of ACK lysing buffer and incubating for 3–5 min at RT. Centrifuge the cells again at $300 \times g$ for 5 min, remove supernatant, and rinse with 10 mL of PBS to remove lysis buffer. Spin again to pellet cells.
4. Resuspend the cells in 5–10 mL media. To determine number of viable cells, mix 20 µL of cells with an equal volume of 0.4% trypan blue and count viable cells with a hemocytometer. Dilute to set the single cell suspension concentration at 5×10^5 splenocytes/mL in complete RPMI.

5. Culture 8×10^6 cells/well in 24-well plates in complete RPMI, with a total volume of 1 mL per well. Pulse the splenocytes with 10 µg/mL of the relevant peptide (gp100) (or irrelevant peptide control) and incubate at 37 °C in 5% CO₂ for 24 h.
6. On day 1, centrifuge the 24-well plates at $300 \times g$ for 5 min to pellet the splenocytes. Wash the splenocytes with PBS twice, centrifuging as described in **step 3** to remove the unbound peptide. Add complete RPMI back to the culture and incubate for 48 h.
7. On day 3, add 0.4 ng/mL of purified mouse IL-2 to the culture.
8. On day 5, plate the target cells (B16-F10) at a density of 5×10^3 cells/well in a round bottom 96-well tissue culture-treated plate. Collect and count the stimulated splenocytes and add into wells of the plate at different effector-to-target cell ratios (100:1, 50:1, 25:1). Target cells stimulated by PHA-M and media-only wells should be included as positive and negative controls, respectively. A target cell line with the same MHC but antigen-negative should also be added as a control (*see Notes 3–5* for the relevant controls).
9. Incubate the plate containing the cell co-culture at 37 °C for 4 h or longer if needed.
10. To each well of the plate, add 10 µL of 10× lysis solution that is included in the CytoTox 96 Non-Radioactive Cytotoxicity Assay kit for every 100 µL of the cells.
11. Incubate the plate with lysis solution for 45 min in the 37 °C incubator.
12. After the incubation step, remove 50 µL of the supernatant, and transfer the supernatant to a new 96-well clear flat bottom plate.
13. Add 50 µL of the CytoTox 96 reagent and incubate the plate for 30 min at RT. Protect the plate from light.
14. Add 50 µL of the stop solution to each well.
15. Record the absorbance of the wells at 490 nm within 1 h (*see Note 6*).
16. Take the average values for each condition and subtract the average absorbance value for the background (defined as Experimental LDH release).
17. Determine the percentage of cell lysis cytotoxicity by using the following formula:

$$\text{Cytotoxicity (\%)} = \frac{\text{experimental LDH release}}{\text{maximum LDH release}} \times 100$$

3.3 IFN- γ ELISpot

IFN- γ ELISpot is a highly sensitive ex vivo assay used to measure antigen-specific IFN- γ secretion at the single-cell level. IFN- γ is a proinflammatory cytokine released by T cells that is associated with an effective anti-tumor response.

1. Obtain splenocytes of mice by harvesting spleens from immunized mice as described above (*see Subheading 3.2, steps 1 and 2*).
2. Prepare single-cell suspensions of splenocytes as described above (*see Subheading 3.2, steps 2 through 4, Notes 7 and 8*).
3. The day before harvesting splenocytes, incubate the wells with 50 μ L capture antibody (2 μ g/mL) diluted in 4 °C PBS. Seal the plate to prevent evaporation and store at 4 °C overnight. In the next day, wash wells three times with 100 μ L sterile PBS-Tween followed by three times with 100 μ L sterile PBS; each wash should be 5 min on a platform shaker with strong agitation. Next, block the wells using media for 1–2 h in a 37 °C incubator while collecting splenocytes. Alternatively, if using an IFN- γ ELISpot kit that has been pre-coated with the capture antibody (e.g., R&D Systems, Cat. EL485), wet all wells of the microplate with 200 μ L of media for 20 min at RT while collecting splenocytes (*see Notes 9 and 10*).
4. Quickly invert the plate in a flicking motion over a waste container to remove media, and add 100 μ L of fresh media. The use of a multichannel or repeating pipette is suggested to prevent the plate membrane from drying during this and subsequent washing steps. Add 1 \times 10⁶, 0.4 \times 10⁶, or 0.2 \times 10⁶ splenocytes/well in 100 μ L to plates pre-coated with anti-mouse IFN- γ antibody (*see Notes 11 and 12*). Include technical replicates, typically three or more, per condition. Conditions may include cells incubated with specific peptide, cells incubated with irrelevant peptide, cells alone, media alone, or cells with positive control such as 1.5% PHA-M. The final volume per well should be ~200 μ L. Add 10 μ g/mL of relevant antigen peptide, irrelevant peptide, media alone (negative control), or 1.5–2% PHA-M (positive control).
5. Carefully wrap the plate in foil for the duration of the incubation (24 h at 37 °C) to avoid evaporation. Do not disturb the plate, as any movement may affect spot definition (*see Note 13*).
6. Flick the plate to quickly remove supernatant and wash away cells and unbound cytokine by incubating with 200 μ L PBS-Tween per well, incubating for 10 min at RT. Repeat two additional times without the 10 min incubation. For washing, pipette solution into each well, and remove solution by inverting and disposing the wash into a waste solution with 10% bleach or other biohazard container.

7. Add 100 μ L of biotinylated monoclonal detection antibody specific for mouse IFN- γ (0.25–2 μ g/mL) diluted in PBS-Tween-BSA to wells and incubate for 2 h at RT or overnight at 4 °C.
8. Wash the plate (as described in **step 6**) three times with 200 μ L PBS-Tween.
9. Incubate with 100 μ L of avidin-horseradish peroxidase (HRP) conjugate in PBS-BSA for 1–2 h at RT.
10. Wash the plate three times as in **step 6** with 200 μ L PBS-Tween, then three times with 200 μ L PBS.
11. Add 200 μ L of the HRP substrate solution (e.g., BCIP/NBT substrate solution) and monitor for spot formation.
12. Carefully remove bottom flexible plastic underdrain and stop reaction by vigorously rinsing with deionized water for at least 5 min.
13. Completely dry the plate overnight at RT or at least 2 h at 37 °C (*see Note 14*).
14. Count spots using an automated imaging unit, such as the Cellular Technology ImmunoSpot ELISpot reader and analysis software. *See Fig. 1* for an example of typical data.

3.4 Tumor Challenge

Tumor challenge is a direct method of determining vaccine efficacy and is performed *in vivo*, to verify the *in vitro* and *ex vivo* results. Tumor challenge using a syngeneic model is an important analysis method for cancer vaccine work, as it best models a natural tumor response. An example of such data is shown in *Fig. 2*, for which the protocol is described below.

1. B16-F10 tumor cells are cultured in DMEM with 10% FBS media and passaged at approximately 70% confluence in a T75 culture flask (*see Notes 15 and 16*).
2. A series of steps are used to collect and count B16-F10 tumor cells (*see Note 17*).
 - (a) Remove media supernatant. Rinse the flask with 10 mL of PBS to remove any remaining media. Dispose of the removed PBS and media into a waste solution with 10% bleach or use another approved biohazard disposal method.
 - (b) Release adherent cells by adding 3 mL of warm (37 °C) 0.25% Trypsin-EDTA to the flask and tilt to cover all the cells.
 - (c) Place the flask in 37 °C incubator for 3 min or until cells start to lift from surface when tilted. Add 10 mL of media to stop the trypsin reaction.

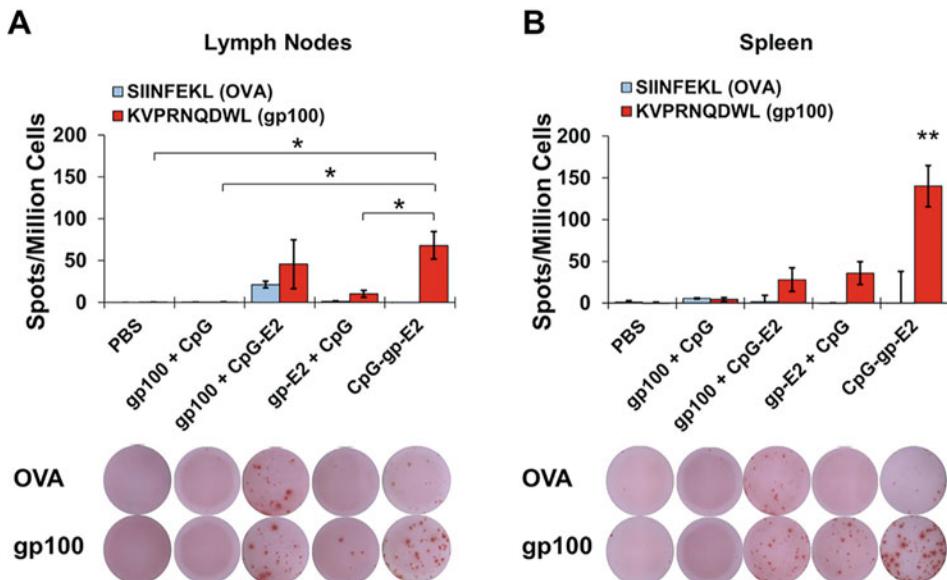


Fig. 1 Gp100 (TAA of mouse melanoma)-specific cytotoxic T cell response is enhanced following immunization with CpG-gp-E2. Lymph nodes and spleens were collected from mice treated with PBS (negative control), gp100 + CpG, gp100 + CpG-E2, gp-E2 + CpG, or whole vaccine CpG-gp-E2. Lymphocytes (a) and splenocytes (b) were harvested and analyzed with IFN- γ ELISpot. Graphs (top) show combined data from ELISpot plates (examples at bottom) with wells stimulated with irrelevant OVA peptide or gp100. Whole vaccine is shown to enhance TAA-specific IFN- γ cell secretion from draining lymph nodes and spleen. Plotted data are average \pm S.E.M., with $N \geq 3$ biological replicates. Statistics: ANOVA and Dunnett's test. * $p < 0.05$, ** $p < 0.01$. CpG: CpG 1826 adjuvant. gp100: gp100_{25–33}, abbreviated gp. (Reproduced with permission from Elsevier from Molino et al., *Biomaterials* 86:83–91 (2016) [7])

- (d) Remove and retain this solution in a conical tube. Rinse the flask with media to collect all cells and add to the same conical tube. Centrifuge the tube at $300 \times g$ for 5 min.
- (e) Remove the supernatant and resuspend the cell pellet in 10 mL of PBS. Count cells and assess viability; mix 90 μ L of cells with 10 μ L of 0.4% trypan blue, then count live and dead cells with a hemocytometer. Dilute to yield a cell concentration of 1×10^6 live cells/mL.
- 3. For inoculation, 6–8 week-old C57BL/6J mice are injected with 7.5×10^3 viable B16-F10 melanoma cells subcutaneously into the right flank in 100 μ L volume using an insulin syringe. We consider this day 0 of our experiment (Fig. 2). The number of animals depends on the number of groups (typically, at least five animals per group).
- 4. Mice are separated into different treatment groups at 5 mice per group. For the example shown in Fig. 2, this is PBS (vehicle control), 1 vaccine immunization (day -28), and 2 vaccine immunizations (day -28 and day -14). On day -28, the groups

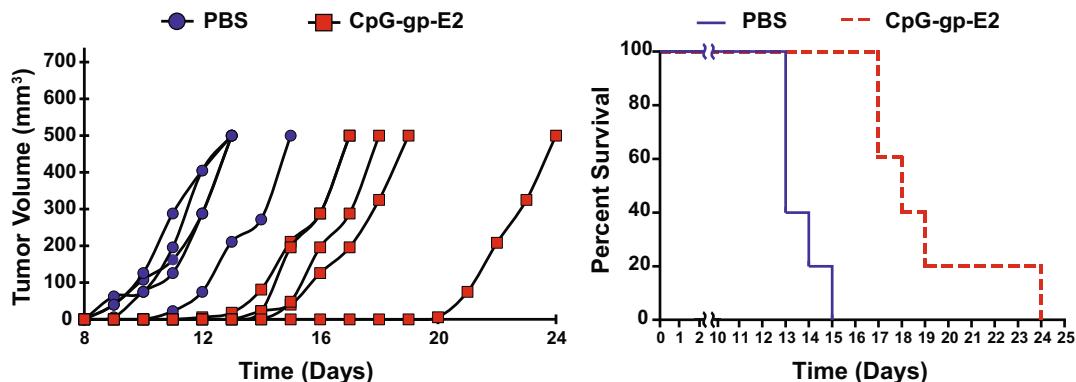


Fig. 2 Immunization with E2 vaccine displaying gp100 epitope enhances anti-tumor immunity. Mice were immunized with PBS negative control or E2 encompassing CpG 1826 adjuvant and displaying gp100 peptide (CpG-gp-E2) on Days -28 and -14, followed by B16-F10 tumor challenge on Day 0. (a) The tumor volumes of individual mice were monitored over time, showing a delay in tumor growth for vaccinated mice. Tumor volumes were approximated as $(0.5 \times \text{shortest diameter}^2 \times \text{longest diameter})$. (b) Vaccinated mice survived for longer times after tumor challenge, further verifying the efficacy of the vaccine formulation. This vaccine formulation has also been used for a number of vaccination schedules in combination with immune checkpoint therapy, further improving performance [9]. $N = 5$ mice per group, with 5 μg peptide per immunization. Statistics: log-rank test, $p < 0.002$. (Reproduced with permission from Elsevier from Molino et al., *Biomaterials* 86:83–91 (2016) [7])

are injected with 50 μg of vaccine (CpG-gp-E2) in 120 μL of PBS or the same volume of PBS subcutaneously at the base of the tail. On day -14, the two immunization groups and the PBS group receive the same dose again (see Note 18).

5. Mice are monitored at least every other day for tumor growth or distress. Once tumors appear, tumor volume are measured with calipers daily. Upon completion of experimental schedule or the growth of tumor reaching 1 cm in diameter, both primary and secondary termination methods are performed according to approved IACUC protocols (see Notes 19–21). Results of the in vivo tumor challenge may be displayed as a growth curve or a Kaplan-Meier survival curve to elucidate vaccination benefits.
6. Tumors and spleens are collected from euthanized mice for immunophenotyping using flow cytometry. Populations of interest may include natural killer cells, dendritic cells, B cells, macrophages, T cells, CD8+ T cells, CD4+ T cells, and CD4+ FoxP3+ T cells.

4 Notes

4.1 Antigen and Adjuvant Conjugation

1. Stock solutions of BMPH and sulfo-SMCC should be made fresh every time prior to use. Dissolve BMPH linker in DMSO, and make the BMPH linker concentration at 5 mg/mL. Bring sulfo-SMCC to room temperature before opening the bottle, and make fresh sulfo-SMCC stock solution at 5 mg/mL in water.
2. If the peptide for the conjugation is insoluble in aqueous solution, try to dissolve the peptide first in organic solvents (e.g., DMSO). Then add the peptide into the E2 solution very slowly, and gently swirl the pipette tip while adding the peptide. The maximum amount of solvent is 10% of solvent in aqueous solution.

4.2 CTL Assay

3. Cellular controls should include: (1) a cell line that is of the same MHC but antigen-negative to show antigen specificity, (2) a cell line that is antigen-positive but a different MHC (a negative control to show proper peptide presentation may also be included, but it is not shown in this example), (3) unstimulated effector cell control to show that the cells are being activated by the addition of peptide, and (4) target cells stimulated by a positive control such as PHA-M.
4. For background control wells, use media-only.
5. Multiple target-to-effector cell ratios should be used to detect a dose-dependent effect.
6. Remove any bubbles that are in the well with a syringe needle prior to reading the absorbance, if needed.

4.3 IFN- γ ELISpot

7. Subheading 3.3, steps 1 through 4 must be performed in a sterile environment. Contamination may cause false positive spots.
8. Collected splenocytes can also be analyzed using flow cytometry to quantify cell populations using markers such as CD11c+ (dendritic cells), CD3+ (T cells), and CD3+CD4+ or CD3+CD8+ (CD4+ or CD8+ T cells). However, ensure enough splenocytes are collected for any additional analysis.
9. Once ELISpot wells are wet, do not let them dry out until the final step. Drying may cause high background.
10. ELISpot conditions can vary significantly between reagent/kit manufacturers. Refer to product-specific protocols for optimal outcome.
11. Do not touch the bottom of the wells with any pipette tip. Doing so may lead to a streak of false positive color.

12. Multiple cell dilutions are recommended to ensure that data signal is in the linear range and not oversaturated.
 13. During this incubation, the plate is highly sensitive to any shaking. Be slow and cautious while taking the plate out of the incubator, as well as any opening and closing of the incubator door.
 14. Do not dry the plate at temperatures higher than 37 °C, as doing so may cause membrane cracking.
- 4.4 Tumor Challenge**
15. Carry out all animal studies in accordance with institution-approved animal protocols and guidelines. It is recommended to check the status of mice at least every other day to ensure animal well-being and monitor tumor growth. Tumor growth rate will depend on model and number of cells used for inoculation.
 16. After each cell passage, the genetic makeup of the B16-F10 cells could change. Similarly, there may be differences based on the commercial source of the cells. Therefore, confirming the presence of the chosen tumor-associated antigen and/or expression level when acquiring new cells is recommended.
 17. Tumor cells can be frozen in 10–20% DMSO in FBS (to help prevent ice crystal formation) in a cryogenic vial and stored with liquid nitrogen. To thaw cryopreserved B16-F10 cells, hold the ~1 mL vial in a 37 °C water bath until half of ice pellet remains. Then, quickly transfer contents to ~9 mL media. Spin cells at $300 \times g$ for 5 min, discard supernatant, and re-suspend in media. Cells may be cultured or inoculated. Ensure cells are maintained in a sterile environment at each step.
 18. Depending on the tumor model and TAA, the use of humanized model mice may be necessary.
 19. In accordance to protocols approved by the Institute for Animal Care and Use Committee (IACUC) at the University of California, Irvine, mice must be euthanized by the time tumor diameters reach 1.5 cm or tumor volumes reach 1500 mm³ [7, 9].
 20. Tumor volume can be approximated as $(0.5 \times \text{shortest diameter}^2 \times \text{longest diameter})$.
 21. To more easily monitor tumor growth, shave hair at the site of injection.

Acknowledgements

This work was supported by the National Institutes of Health (1R01EB027797 and P30CA062203) and the Department of Education (GAANN Fellowship for N.B.). We gratefully acknowledge Dr. Medea Neek for helpful discussions.

References

1. Schlor J, Hodge JW, Palena C, Tsang KY, Jochems C, Greiner JW, Farsaci B, Madan RA, Heery CR, Gulley JL (2014) Therapeutic cancer vaccines. *Adv Cancer Res* 121:67–124
2. Molino NM, Wang S-W (2014) Caged protein nanoparticles for drug delivery. *Curr Opin Biotechnol* 28:75–82
3. Hirayama M, Nishimura Y (2016) The present status and future prospects of peptide-based cancer vaccines. *Int Immunopharmacol* 28(7):319–328
4. Slingluff CL Jr (2011) The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? *Cancer J* 17(5):343–350
5. Nagorsen D, Scheibenbogen C, Thiel E, Keilholz U (2004) Immunologic monitoring of cancer vaccine therapy. *Expert Opin Biol Ther* 4(10):1677–1684
6. Dalmau M, Lim S, Chen HC, Ruiz C, Wang SW (2008) Thermostability and molecular encapsulation within an engineered caged protein scaffold. *Biotechnol Bioeng* 101(4):654–664
7. Molino NM, Neek M, Tucker JA, Nelson EL, Wang SW (2016) Viral-mimicking protein nanoparticle vaccine for eliciting anti-tumor responses. *Biomaterials* 86:83–91
8. Neek M, Tucker JA, Kim TI, Molino NM, Nelson EL, Wang SW (2018) Co-delivery of human cancer-testis antigens with adjuvant in protein nanoparticles induces higher cell-mediated immune responses. *Biomaterials* 156:194–203
9. Neek M, Tucker JA, Butkovich N, Nelson EL, Wang S-W (2020) An antigen-delivery protein nanoparticle combined with anti-PD-1 checkpoint inhibitor has curative efficacy in an aggressive melanoma model. *Adv Ther* 3(12):2000122
10. Molino NM, Anderson AKL, Nelson EL, Wang S-W (2013) Biomimetic protein nanoparticles facilitate enhanced dendritic cell activation and cross-presentation. *ACS Nano* 7(11):9743–9752



Chapter 19

Construction Protocol of Drug-Protein Cage Complexes for Drug Delivery System

Yuichi Nakahara, Yuta Endo, and Ippei Inoue

Abstract

Ferritin is one of the most promising drug delivery system (DDS) carriers because of its uniform nanosize, biodistribution, efficient cellular uptake, and biocompatibility. Conventionally, a disassembly/reassembly method that requires pH change has been used for the encapsulation of molecules in ferritin protein nanocages. Recently, a one-step method in which a complex of ferritin and a targeted drug was obtained by incubating the mixture at an appropriate pH, was established. Here, we describe two types of protocols, the conventional disassembly/reassembly method, and the novel one-step method for the construction of a ferritin-encapsulated drug using doxorubicin as an example molecule.

Key words Ferritin, Protein cage, Nanomedicine, Drug delivery system, Cancer chemotherapy, Doxorubicin, Tumor, Protein purification

1 Introduction

Human ferritin, a major iron-storage protein, is composed of 24 identical subunits that assemble into a robust cage-shaped structure with an external size of 12 nm [1, 2]. The cage-shaped structure capable of holding iron ions within its inner cavity is approximately 8 nm in diameter. Owing to their attractive characteristics such as uniform nanosize, solution dispersity, thermal stability (melting temperature up to 80 °C), pH stability (pH 3–12), easy functionalization, and low-cost for large-scale production [2–4]; ferritin has been applied for the construction of nanoparticles in various fields from nanoelectronic devices [5–7] to medical applications [8–12].

One of the medical applications is to use ferritin as a drug carrier/DDS to encapsulate various molecules such as doxorubicin (DOX) [13], cisplatin [14], paclitaxel [15], and oligonucleotides [16] by using a disassembly/reassembly method. The disassembly/reassembly method utilizes pH-dependent conformational changes

in ferritin. The 24-mer ferritin cage shape becomes disassembled into its subunits within acidic pH environments and is reassembled following return to neutral pH environments [17]. To encapsulate a targeted drug, the ferritin disassembled under acidic pH was mixed with the drug. After the disassembled ferritin is reassembled by pH neutralization, the drug is stochastically loaded into the ferritin nanocage during the reassembly process.

Another medical application was bioimaging using metal-complex-loaded ferritin [18–21]. The iron storage mechanism of ferritin was utilized for the construction of medical-metal-complex-loaded ferritin. The ferritin nanocage contains a threefold channel as a pathway for the transfer of iron ions driven by the potential difference between the ferritin inner surface and the protein shell. The transported iron ions bind a negatively charged metal-binding site in the inner cavity of ferritin and grow into iron compound nanoparticles [22]. To deposit a target metal compound, ferritin and the target metal ions are mixed in the optimal buffer. After incubation, ferritin takes the target ion and forms a metal complex nanoparticle in its inner cavity. This method is simple, but it requires optimization of the pH, ion concentration, reaction temperature, and additives.

Recently, a one-step method for obtaining ferritin-encapsulated small-molecule drugs has been developed by utilizing a method for obtaining a metal compound-ferritin complex [12]. In the one-step method, the encapsulation reaction proceeds efficiently at a pH near the pK_a of the target small molecules. A reaction mechanism model has been proposed in which the transport of the organic molecules into the inside of ferritin proceeds by passing through a fourfold channel larger than the threefold channel, depending on the chemical potential difference between the inside and outside of the ferritin nanocage (*see Fig. 1*).

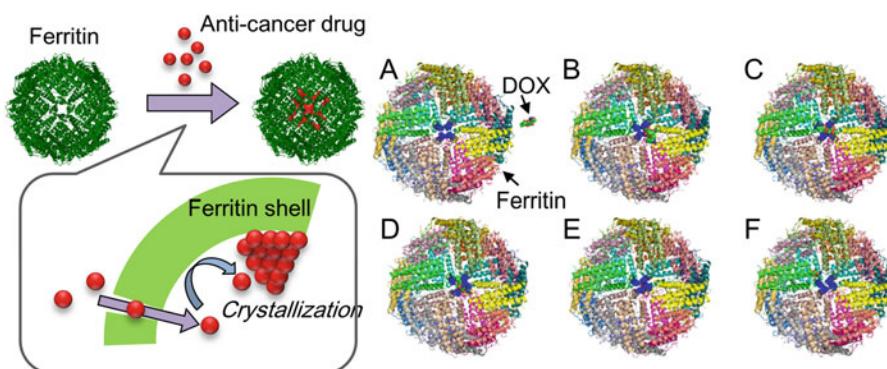


Fig. 1 Mechanism of the one-step method. An electrostatic potential within the ferritin cage facilitated the passage through a fourfold channel and deposition of the small molecule in the inner cavity. (Reproduced from Ref. [12] with permission from the Royal Society of Chemistry)

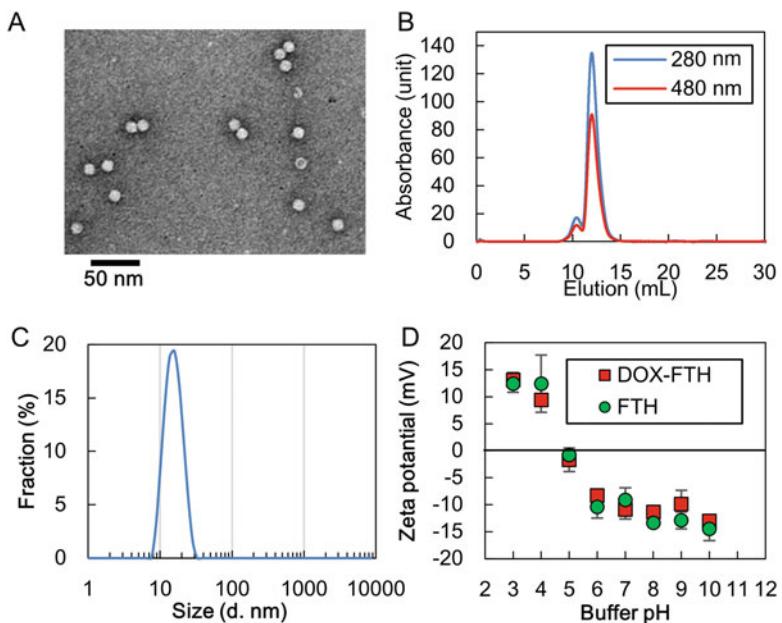


Fig. 2 The character of DOX-FTH by the one-step method. **(a)** TEM image of DOX-FTH. **(b)** SEC analyses of DOX-FTH. **(c)** DLS analysis of DOX-FTH size distribution. The diameter of DOX-FTH was 14.61 nm using the intensity mode; the Pdi value was 0.065. **(d)** Zeta potential of DOX-FTH; red squares and green circles represent zeta potential of the DOX-FTH and apo-ferritin, respectively (means \pm standard deviation, $n = 8$). (Reproduced from Ref. [12] with permission from the Royal Society of Chemistry)

Here, we introduce the disassembly/reassembly method and the one-step method using DOX as a model molecule. The one-step method was superior to the disassembly/reassembly method in terms of loading capacity, encapsulation efficiency, and human ferritin heavy chain (FTH) recovery rate [12]. Figure 2 shows the characteristics of ferritin-encapsulated DOX (DOX-FTH) using the one-step method. Transmission electron microscopy (TEM) analysis indicated that DOX-FTH formed a cage-shaped structure with a diameter of 12 nm, similar to apo-ferritin (*see* Fig. 2a). The mono-dispersibility of DOX-FTH in the solution was confirmed by size exclusion chromatography (SEC) and dynamic light scattering (DLS) analysis (*see* Fig. 2b, c). The pI of DOX-FTH was approximately pH 5, and the zeta potential of DOX-FTH at pH 7 was -10.9 ± 1.8 mV (*see* Fig. 2d). DOX-FTH showed high stability at 4 °C for 6 months (*see* Fig. 3). DOX-FTH was also stable in human plasma solution at 37 °C for 24 h. In contrast, DOX was rapidly released from the

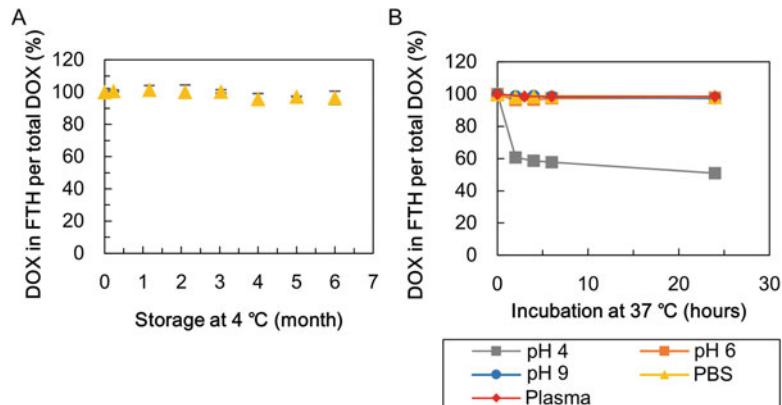


Fig. 3 Stability of DOX-FTH. **(a)** The stability of DOX-FTH was evaluated in PBS at 4 °C in the dark (mean \pm standard deviation, $n = 3$). **(b)** DOX-FTH stability at 37 °C. DOX-FTH was stored in pH 4.0 (grey squares), pH 6.0 (orange squares), pH 7.4 (PBS, blue circles), pH 9.0 (yellow triangles), and human plasma (red diamonds) for 24 h in the dark (mean \pm standard deviation, $n = 3$). (Reproduced from Ref. [12] with permission from the Royal Society of Chemistry)

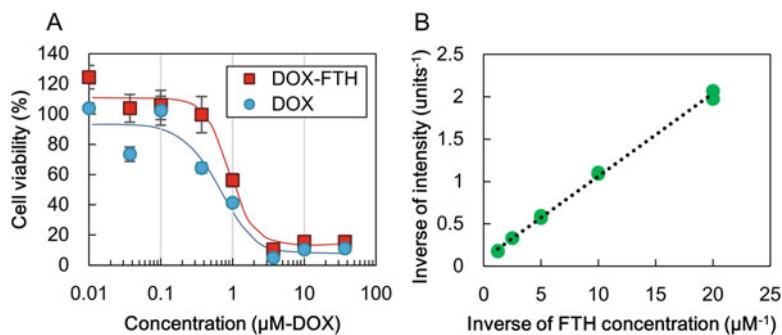


Fig. 4 **(a)** Dose-dependent cytotoxicity of DOX-FTH. Cell viability of SKBR-3 cells treated with drugs was measured by the WST-8 ($n = 3$). **(b)** A double-reciprocal plot of FTH concentration and fluorescence intensities of SKBR-3 cells incubated with uranine dye-loaded FTH for 24 h at 37 °C. The fluorescence intensity of the cells was measured by flow cytometry ($R^2 = 1.00$). (Reproduced from Ref. [12] with permission from the Royal Society of Chemistry)

FTH nanocage in a low-pH environment. Therefore, DOX-FTH is stable during the delivery of its encapsulated drug through the systemic circulation. After delivery into the target cells, DOX-FTH releases DOX into endosomes. Since TfR1-expressing cells were efficiently able to take up DOX-FTH by endocytosis, ferritin is a promising drug delivery system (DDS) carrier (see Figs. 4 and 5). This one-step method is a powerful production process for the construction of drug-loaded ferritin.

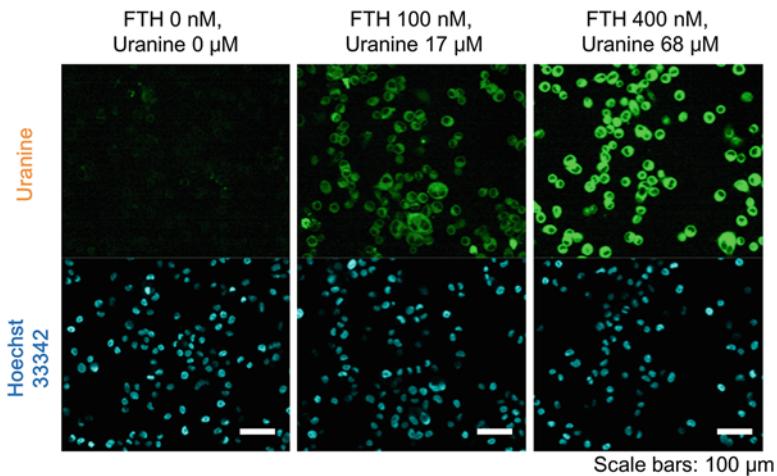


Fig. 5 Cellular uptake of the drug-FTH complex. Confocal microscopy images of the SKBR-3 cells incubated with uranine dye (green)-loaded FTH. Nuclei were labeled with Hoechst33342 (cyan). (Reproduced from Ref. [12] with permission from the Royal Society of Chemistry)

2 Materials

2.1 Ferritin Production

1. Human ferritin heavy chain gene (FTH, NCBI Gene ID: 2495).
2. pET-20b(+) plasmid (Merck Millipore).
3. FTH in F primer; 5'- aaggagatat acatatgacg accgcgtcca cctcg- 3'.
4. FTH in R primer; 5'- ctcgaattcg gatccttagc tttcatttac actgtc- 3'.
5. pET20b-in F primer; 5'- aggatccgaa ttcgagctcc gtcg- 3'.
6. pET20b-in R primer; 5'- catatgtata tctcattttt aaagttaaac- 3'.
7. PrimeSTAR HS DNA Polymerase (Takara Bio, Inc.).
8. *Escherichia coli* JM109 Competent Cells (Takara Bio, Inc.).
9. ECOS™ competent *Escherichia coli* BL21(DE3) (Nippon Gene Co., Ltd.).
10. 50 mM Tris-HCl buffer, pH 8.0.
11. 50 mM Tris-HCl/1 M NaCl buffer, pH 8.0.
12. 10 mM Tris-HCl buffer, pH 8.0.
13. LB medium (Luria-Bertani, Becton) with 100 µg/mL ampicillin.
14. LB-agar plates with 100 µg/mL ampicillin.
15. in-fusion system (Takara Bio, Inc.).

16. Wizard® SV Gel and PCR Clean-Up System (Promega).
17. Wizard® Plus SV Minipreps DNA Purification Systems (Promega).
18. HiTrap Q HP 5 mL (Cytiva).
19. HiPrep 26/60 Sephadryl S-300 HR (Cytiva).
20. Steriflip-GV (0.22 µm pore, Merck).
21. Millex-GV filter (0.22 µm pore, Merck).
22. Vivaspin 20-100K (MWCO 100 kDa, Cytiva).
23. 0.5 L flask.
24. 50 mL centrifuge tube.

2.2 Protein Analysis

1. Buffer A: Water containing 0.1% TFA (trifluoroacetic acid).
2. Buffer B: 80 vol% acetonitrile containing 0.1% TFA.
3. PBS, pH 7.4.
4. PLRP-S 300A (3 µm, 150 mm × 4.6 mm; Agilent Technologies).
5. Superdex 200 Increase 10/300GL column (Cytiva).
6. PBS, pH 7.4 (FUJIFILM Wako Pure Chemical Co.).

2.3 Drug-Ferritin Complex Construction

1. Doxorubicin hydrochloride (DOX, CAS RN. 25316-40-9).
2. 1 M Glycine buffer (pH adjusted to 2.3 with HCl).
3. 1 M Tris-HCl, pH 9.0 (Nacalai Tesque, Inc.).
4. PD-10 (Disposable PD-10 Desalting Column, with Sephadex G-25. resin) (Cytiva).
5. Superdex 200 Increase 10/300GL column (Cytiva).

2.4 Equipment

1. Thermal cycler.
2. Centrifuge.
3. Sonicator.
4. Water bath.
5. AKTA pure 25 system (Cytiva).
6. HPLC system.
7. Transmission Electron Microscope (TEM).
8. Dynamic light scattering (DLS) system.
9. UV/visible spectrophotometer.

3 Methods

3.1 Plasmid Construction

1. Amplify human ferritin heavy chain gene (FTH) using PCR with PrimeSTAR HS DNA Polymerase, primers (FTH in F and FTH in R) and the synthesized FTH gene as a template.
2. Amplify the pET-20b(+) plasmid using PCR with PrimeSTAR HS DNA Polymerase, primers (pET20b-in F and pET20b-in R) and pET-20b(+) plasmid as template.
3. Purify the PCR products of **steps 1** and **2** using Wizard® SV Gel and PCR Clean-Up System.
4. Introduce the FTH gene into a pET-20b(+) plasmid using an in-fusion system with purified PCR products.
5. Transform *E. coli* JM109 with the **step 4** mixture and cultivate the transformant (JM109/pET-FTH) in LB-agar plates with 100 µg/mL ampicillin at 37 °C for 14 h (*see Note 1*).
6. Inoculate and incubate a grown colony of JM109/pET-FTH in 1 mL of LB medium with 100 µg/mL ampicillin at 37 °C for 14 h.
7. Purify the pET-FTH plasmid from **step 6** broth using Wizard® Plus SV Minipreps DNA Purification Systems.

3.2 Protein Production

1. Transform ECOS™ competent *Escherichia coli* BL21(DE3) with the purified pET-FTH and incubate in LB-agar plates with 100 µg/mL ampicillin at 37 °C for 12 h.
2. Incubate a grown colony of BL21(DE3)/pET-FTH in 100 mL of LB medium with 100 µg/mL ampicillin while shaking in a 0.5 L flask at 37 °C for 24 h (*see Notes 2 and 3*).
3. Centrifuge the **step 2** broth using to 50 mL centrifuge tubes at 5000 × *g* for 5 min and discard the supernatant (*see Note 4*).
4. Suspend the **step 3** pellet in 20 mL of 50 mM Tris–HCl buffer, pH 8.0.
5. Combine the suspended pellet in one tube.
6. Centrifuge at 5000 × *g* for 5 min and discard the supernatant.
7. Suspend the **step 6** pellet in 10 mL of 50 mM Tris–HCl buffer, pH 8.0.
8. Disrupt the suspended cells using an ultrasonic wave (140 W) at 4 °C for 5 min.
9. Collect the disrupted cells in a new 50 mL centrifuge tube and centrifuge at 8000 × *g* for 5 min.
10. Transfer the supernatant to a new 50 mL tube and boil at 60 °C for 20 min.
11. Cool the boiled **step 10** solution on ice for over 30 min.

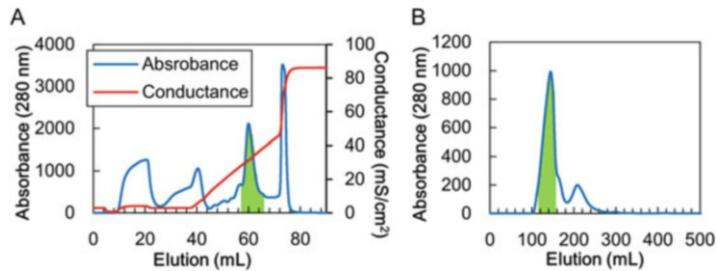


Fig. 6 Purification of FTH using (a) anion exchange chromatography and (b) size filtration chromatography. FTH eluted in the green area in each chromatogram

12. Centrifuge the cold solution at $8000 \times g$ for 15 min.
13. Filter the supernatant of **step 12** with Steriflip-GV (*see Notes 5 and 6*).
14. Load the filtered solution into an anion-exchanger column (HiTrap Q HP 5 mL). FTH is eluted with an NaCl gradient (0–0.7 M) in 50 mM Tris–HCl buffer, pH 8.0, and 50 mM Tris–HCl/1 M NaCl, pH 8.0, at a flow rate of 3 mL/min. FTH is detected at 280 nm (*see Fig. 6a*). Elution is controlled using the AKTA pure 25 system (*see Note 7*).
15. Pool and concentrate the eluted FTH to 10 mL using membrane ultrafiltration (Vivaspin 20-100K, MWCO 100 kDa).
16. Load the concentrated FTH into a gel filtration column (HiPrep 26/60 Sephadryl S-300 HR). FTH is eluted with 10 mM Tris–HCl buffer, pH 8.0, at a flow rate of 1.3 mL/min. FTH is detected at 280 nm absorption (*see Fig. 6b* and **Note 8**).
17. Pool and filter sterilize FTH using Millex-GV filter.
18. Measure the concentration and purity of FTH by RPC with buffers A and B (*see Fig. 7a* and **Notes 9–11**).
19. Characterize FTH using TEM, SEC, and DLS (*see Fig. 7* and **Notes 12–15**).

3.3 Disassembly/Reassembly Method

1. Mix FTH and DOX in 1 mL of 50 mM Glycine-HCl buffer, pH 2.3, to a final concentration of FTH (1 mg/mL, 2 μ M) and DOX (0.5 mg/mL, 0.9 mM) (*see Notes 16 and 17*).
2. Incubate the **step 1** mixture at 25 °C for 15 min.
3. Add 100 μ L of 1 M Tris–HCl buffer, pH 9.0, after **step 2** incubation (*see Notes 18 and 19*).
4. After 60 min of incubation, add 1.5 mL of 50 mM Tris–HCl, pH 8.0, to the **step 3** mixture.

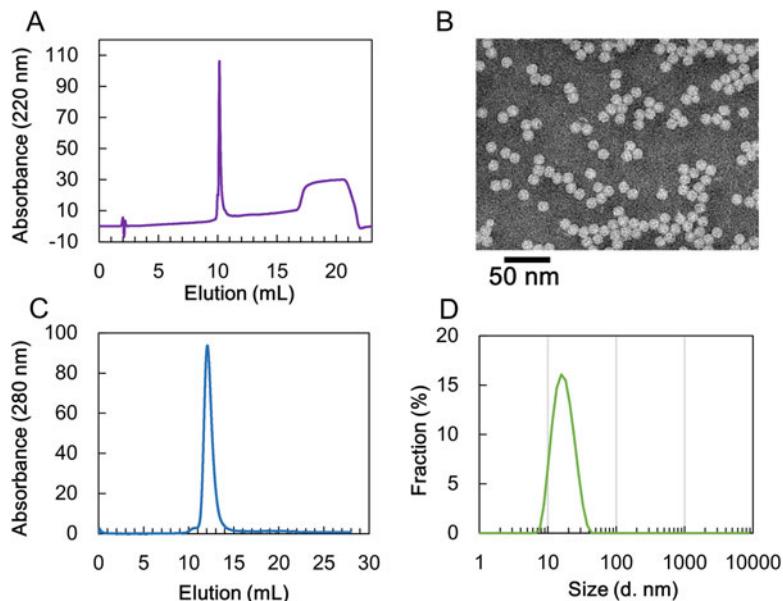


Fig. 7 The character of the purified FTH. FTH was analyzed by (a) RPC, (b) TEM, (c) SEC, and (d) DLS

5. Load 2.5 mL of **step 4** mixture to the PD-10 column equilibrated with 50 mM Tris–HCl, pH 8.0. Discard the flow-through (*see Note 20*).
6. Elute FTH encapsulated DOX with 3.5 mL of 50 mM Tris–HCl, pH 8.0, and the eluate.

3.4 One-Step Method

1. Mix 1 M Tris–HCl buffer, pH 9.0, FTH and DOX in 1 mL of reactant at a final concentration of 50 mM Tris–HCl buffer, FTH (1 mg/mL, 2 μ M) and DOX (0.3 mg/mL, 0.5 mM).
2. Incubate the **step 1** mixture at 60 °C for 60 min at water bath (*see Note 21*).
3. Add 1.5 mL of 50 mM Tris–HCl, pH 8.0.
4. Load 2.5 mL of the **step 3** mixture to the PD-10 column equilibrated with 50 mM Tris–HCl, pH 8.0. Discard the flow-through (*see Note 22*).
5. Elute FTH encapsulated DOX with 3.5 mL of 50 mM Tris–HCl, pH 8.0, and pool the eluate.

3.5 Evaluation of Drug-Protein Complex

1. Measure the nanostructure, solution dispersibility, and surface potential of DOX-FTH by TEM, SEC, and DLS (*see Fig. 2* and **Notes 12–17**).

Table 1
An example of reaction results (see Notes 23–27)

	Disassembly/reassembly (%)	One-step (%)
Loading capacity	4	16
Encapsulation efficiency	5	49
FTH recovery rate	61	78

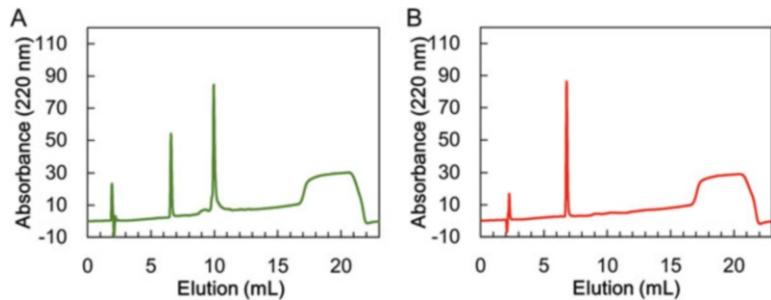


Fig. 8 RPC analysis of DOX-FTH. (a) RPC chromatogram of DOX-FTH solution. RPC was able to separate DOX (7 min) and FTH (10 min) of DOX-FTH. (b) RPC chromatogram of 0.1 mg/mL DOX solution

- Measure concentration, drug loading capacity, and encapsulation efficiency using a UV/visible spectrophotometer or RPC (see Table 1 and Fig. 8).

4 Notes

- JM109 should be used for amplification of pET-FTH. The plasmid amplified in BL21(DE3) decreased the FTH production activity.
- The shaking speed depends on the culture system. For example, the speed is set to 120 rpm.
- Increasing the number of flasks for the preparation of a large amount of FTH.
- When cultivation using multiple flasks, repeat **step 3** for combining into two tubes.
- When the supernatant contained a large amount of debris, using a 0.45 µm pore filter before 0.2 µm pore filtration.
- Before column purification, confirm the expression and recovery of FTH by SDS-PAGE analysis.

7. Depending on the condition of the sample and column, elute the FTH at a solution conductance of approximately 30 mS/cm².
8. FTH eluted at around 0.35 CV.
9. The yield of FTH was 50–100 mg/L. The purity of the FTH is over 95%.
10. Apply Samples to an RPC column at a flow rate of 1.0 mL/min with a linear elution gradient from 24% to 56% acetonitrile containing 0.1% trifluoroacetic acid (TFA) at 70 °C. The eluate is monitored at the absorbance of 220 nm.
11. FTH can be stored stably at 4 °C or –80 °C.
12. JEM-2200FS (JEOL) is used for TEM analysis. To prepare for TEM, FTH samples (0.2 mg/mL, 2 µL) are dropped on a carbon-coated copper grid. After 1 min, the excess solution is wicked with filter paper and washed with 2 µL of water. The samples on the grid are stained with 1% phosphotungstic acid (PTA).
13. SEC analyses are conducted using AKTA pure (Cytiva) with a Superdex 200 Increase 10/300GL column. PBS, pH 7.4, is used as the elution buffer at a flow rate of 0.8 mL/min. Protein and DOX are monitored at 280 nm and 480 nm, respectively.
14. DLS analyses are conducted using a Zetasizer Nano (Malvern Instruments). Samples containing 0.5 mg/mL of FTH are measured in 50 µL PBS, pH 7.4, at 25 °C.
15. Zeta potential analyses are also conducted using a Zetasizer Nano. Samples containing 0.1 mg/mL FTH are measured in 750 µL of buffer at 25 °C. The sample pH is adjusted by adding 50 mM phosphate buffer, pH 3, 6, and 7, 50 mM acetate buffer, pH 4 and 5, 50 mM Tris–HCl buffer, pH 8 and 9, and 50 mM carbonate-bicarbonate buffer, pH 10.
16. Ferritin supramolecule is disassembled into identical subunits in acidic conditions.
17. HCl is also used for adjusting pH.
18. Ferritin subunits are reassembled into 24-mer ferritin supramolecule at around natural pH.
19. NaOH is also used for adjusting pH.
20. Optionally, debris in the **step 4** mixture is removed using a centrifuge before PD-10 purification.
21. A stirrer is used for a large volume reaction.
22. Optionally, debris in the **step 3** solution is removed using a centrifuge before PD-10 purification.
23. A loading capacity is calculated as follows:

(weight of encapsulated drug) per (weight of ferritin-encapsulated drug) × 100.

24. Encapsulation efficiency is calculated as follows:

(weight of encapsulated drug) per (weight of drug used for reaction) × 100.

25. FTH recovery rate is calculated as follows:

(weight of FTH after purification) per (weight of FTH for reaction) × 100.

26. To determine the DOX concentration using a spectrophotometer, DOX-FTH is incubated at pH 2.3, by adding 50 mM Glycine-HCl buffer for 15 min. After releasing DOX from the FTH nanocage, the concentration of encapsulated DOX is determined by measuring the absorbance at 480 nm (DU800 UV/Visible Spectrophotometer, Beckman Coulter).
27. To determine the DOX concentration using RPC, samples are applied to an RPC column at a flow rate of 1.0 mL/min with a linear elution gradient from 24% to 56% acetonitrile containing 0.1% trifluoroacetic acid (TFA) at 70 °C. The eluate was monitored at the absorbance of 220 nm.

References

1. Truffi M, Fiandra L, Sorrentino L, Monieri M, Corsi F, Mazzucchelli S (2016) Ferritin nanocages: a biological platform for drug delivery, imaging and theranostics in cancer. *Pharmacol Res* 107:57–65
2. Theil EC (1987) Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annu Rev Biochem* 56: 289–315
3. Martsev SP, Vlasov AP, Arosio P (1998) Distinct stability of recombinant L and H subunits of human ferritin: calorimetric and ANS binding studies. *Protein Eng* 11:377–381
4. Crichton RR, Declercq JP (2010) X-ray structures of ferritins and related proteins. *Biochim Biophys Acta* 1800:706–718
5. Yamashita I, Iwahori K, Kumagai S (2010) Ferritin in the field of nanodevices. *Biochim Biophys Acta* 1800:846–857
6. He D, Marles-Wright J (2015) Ferritin family proteins and their use in bionanotechnology. *Nat Biotechnol* 32:651–657
7. Inoue I, Yamauchi H, Okamoto N, Toyoda K, Horita M, Ishikawa Y, Yasueda H, Uraoka Y, Yamashita I (2015) Thermo-stable carbon nanotube-TiO₂ nanocomposite as electron highways in dye-sensitized solar cell produced by bio-nano-process. *Nanotechnology* 26: 285601
8. Zhang C, Zhang X, Zhao G (2020) Ferritin nanocage: a versatile nanocarrier utilized in the field of food, nutrition, and medicine. *Nanomaterials (Basel)* 10:1894
9. Fan K, Cao C, Pan Y, Lu D, Yang D, Feng J, Song L, Liang M, Yan X (2012) Magnetoferritin nanoparticles for targeting and visualizing tumour tissues. *Nat Nanotechnol* 7:459–464
10. Uchida M, Flenniken ML, Allen M, Willits DA, Crowley BE, Brumfield S, Willis AF, Jackiw L, Jutila M, Young MJ, Douglas T (2006) Targeting of cancer cells with ferrimagnetic ferritin cage nanoparticles. *J Am Chem Soc* 128: 16626–16633
11. Monti DM, Ferraro G, Merlini A (2019) Ferritin-based anticancer metallodrug delivery: crystallographic, analytical and cytotoxicity studies. *Nanomedicine* 20:101997
12. Inoue I, Chiba M, Ito K, Okamatsu Y, Suga Y, Kitahara Y, Nakahara Y, Endo Y, Takahashi K, Tagami U, Okamoto N (2021) One-step construction of ferritin encapsulation drugs for cancer chemotherapy. *Nanoscale* 13:1875–1883
13. Huang X, Chisholm J, Zhuang J, Xiao Y, Duncan G, Chen X, Suk JS, Hanes J (2017) Protein nanocages that penetrate airway mucus and tumor tissue. *Proc Natl Acad Sci U S A* 114:E6595–E6602

14. Falvo E, Tremante E, Fraioli R, Leonetti C, Zamparelli C, Boffi A, Morea V, Ceci P, Giacomini P (2013) Antibody–drug conjugates: targeting melanoma with cisplatin encapsulated in protein-cage nanoparticles based on human ferritin. *Nanoscale* 5:12278–12285
15. Li R, Ma Y, Dong Y, Zhao Z, You C, Huang S, Li X, Wang F, Zhang Y (2019) Novel paclitaxel-loaded nanoparticles based on human H chain ferritin for tumor-targeted delivery. *ACS Biomater Sci Eng* 5:6645–6654
16. Li L, Muñoz-Culla M, Carmona U, Lopez MP, Yang F, Trigueros C, Otaegui D, Zhang L, Knez M (2016) Ferritin-mediated siRNA delivery and gene silencing in human tumor and primary cells. *Biomaterials* 98:143–151
17. Kim M, Rho Y, Jin KS, Ahn B, Jung S, Kim H, Ree M (2011) PH-dependent structures of ferritin and apoferritin in solution: disassembly and reassembly. *Biomacromolecules* 12:1629–1640
18. Uchida M, Terashima M, Cunningham CH, Suzuki Y, Willits DA, Willis AF, Yang PC, Tsao PS, McConnell MV, Young MJ, Douglas T (2008) A human ferritin iron oxide nanocomposite magnetic resonance contrast agent. *Magn Reson Med* 60:1073–1081
19. Ruggiero MR, Alberti D, Bitonto V, Geninatti Crich S (2019) Ferritin: a platform for MRI contrast agents delivery. *Inorganics* 7:33
20. Neburkova J, Rulseh AM, Chang SLY, Raabova H, Vejpravova J, Dracinsky M, Tarabek J, Kotek J, Pingle M, Majer P, Vymazal J, Cigler P (2020) Formation of gadolinium-ferritin from clinical magnetic resonance contrast agents. *Nanoscale Adv* 2: 5567–5571
21. Kitagawa T, Kosuge H, Uchida M, Iida Y, Dalton RL, Douglas T, McConnell MV (2017) RGD targeting of human ferritin iron oxide nanoparticles enhances in vivo MRI of vascular inflammation and angiogenesis in experimental carotid disease and abdominal aortic aneurysm. *J Magn Reson Imaging* 45:1144–1153
22. Chasteen ND, Harrison PM (1999) Mineralization in ferritin: an efficient means of iron storage. *J Struct Biol* 126:182–194



Chapter 20

Protein Cage Relaxivity Measurement for Magnetic Resonance Imaging Contrast Agents

Ambrish Kumar, Bhargy Sharma, and Sierin Lim

Abstract

Contrast agents are employed to enhance the differentiation of diseased cells or lesions from normal tissues in magnetic resonance imaging (MRI). Protein cages have been explored as templates to synthesize superparamagnetic MRI contrast agents for decades. The biological origin imparts natural precision in forming confined nano-sized reaction vessels. With natural capacity to bind divalent metal ions, ferritin protein cages have been used for the synthesis of nanoparticles containing MRI contrast agents inside their core. Furthermore, ferritin is known to bind transferrin receptor 1 (TfR1) which is overexpressed on specific cancer cell types and could be used for targeted cellular imaging. In addition to iron, other metal ions such as manganese and gadolinium have been encapsulated within the core of ferritin cages. To compare the magnetic properties of ferritin loaded with contrast agents, a protocol for calculating the contrast enhancement power of protein nanocage is required. The contrast enhancement power is demonstrated as relaxivity and can be measured using MRI and solution nuclear magnetic resonance (NMR) methods. In this chapter, we present methods for measuring and calculating the relaxivity of ferritin nanocages loaded with paramagnetic ions in solution (in tube) with NMR and MRI.

Key words Contrast agent, NMR, MRI, Relaxivity, Ferritin

1 Introduction

Magnetic resonance imaging (MRI) is a medical imaging modality that uses magnetic fields and radio waves to produce anatomical images. These anatomical images provide information on the health state of a patient's tissues and cells. The commonly used magnetic fields in a clinical MRI scanner are 1.5 and 3 T. Higher magnetic fields (e.g., 7, 9.4, 10.5, 11.7, 14, 21.1 T) are being explored for enhancing resolutions and are in research stage. In MRI, the high magnetic fields align the hydrogen protons of water molecule (most abundant) in a patient's body. Subsequent irradiation with a pulse of radio waves disrupts the alignment. Following the radio wave pulse, the protons return to their relaxed state and realigned to the magnetic fields. The time that the protons take to return to relaxed

state is called relaxation time that consists of two components that are longitudinal (T_1) and transverse (T_2) relaxation time. The MR images are produced as a result of the different relaxation time of the hydrogen proton of water molecules in the tissues. To enhance the image contrast during a MRI procedure, contrast agents are employed. The enhancement enables better distinction between normal and diseased tissues.

MRI contrast agents are mainly paramagnetic ions whose magnetic moments interact with proton nuclei and eventually shorten the T_1 and T_2 relaxation time [1]. The effect of paramagnetic ions on proton nuclei relaxation in water molecule depends on their electronic configuration which determines the magnetic moments. In MRI contrast agents, these paramagnetic ions are either chelated or coated with a polymer scaffold which further modulates the water proton relaxation process. Scaffolds (silica, micelles and polymers) have been developed to encapsulate the paramagnetic ions and the resulting nanoparticles are used as MRI contrast agents [2, 3]. Ferritin cages pose less toxicity due to their biological origin and are widely explored for the synthesis or encapsulation of MRI contrast agents within the core [4–6]. Twenty four subunits of ferritin form a 12 nm nanocage with an internal diameter of 8 nm [7]. It is a ubiquitous protein whose main function is the maintenance of homeostatic iron (Fe) levels which is achieved by storing the excess iron in the core [8]. Interestingly, ferritin cages also bind non-native metal ions such as gadolinium (Gd) and manganese (Mn) which are of interest as MRI contrast agents [5]. The use of ferritin cages loaded with metal ions for cellular imaging, especially for immune cells such as macrophages, have been demonstrated previously [4, 9, 10]. Furthermore, cancerous cells overexpressing the transferrin receptor 1 (TfR1) show higher uptake of ferritin compared to normal cells, which makes it advantageous for evaluating cancer progression and stages [11].

The contrast enhancement property of a contrast agent is mainly represented as relaxivity (r) with a unit of $\text{mM}^{-1} \text{ s}^{-1}$. r_1 and r_2 correspond to longitudinal and transverse relaxivity, respectively. A higher relaxivity value shows higher contrast property of a contrast agent under the same magnetic field strength and conditions. Therefore, relaxivity is an important parameter to compare the contrast enhancement property of different contrast agents. The relaxivity of a contrast agent depends on the nature of its magnetic domain as well as its scaffold, which affects the interactions of water protons with the magnetic domain. Methods to measure and calculate r_1 and r_2 of ferritin embedded in an agarose matrix have been presented elsewhere [12]. This chapter presents a detailed method for calculating the relaxivity of protein cages loaded with metal ions in a buffer solution using solution NMR and MRI modalities. The acquisition method and raw data for calculating the T_1 and T_2 values from NMR and MRI require

different approaches. In solution NMR, the relaxivity is derived from the peak intensities for water protons, whereas in MRI method, it is derived from the enhancement of image contrast. Subsequently, both peak intensities and MRI images are used to calculate the relaxation time, T_1 and T_2 . T_1 and T_2 are calculated at different concentrations of the MRI contrast agents. The inverse of relaxation time, ($1/T$), known as relaxation rate, are plotted against concentrations of the contrast agent followed by a simple linear regression analysis. The slope of the linear fit gives the respective relaxivity (r_1 and r_2) of the contrast agent [13].

2 Materials

Prepare buffer solutions in deionized water with analytical-grade chemicals. Filter (0.22 µm) and degas the buffer. Store all the buffers at room temperature unless otherwise mentioned.

2.1 Consumables

1. 5 mm NMR tubes.
2. MRI glass tubes.
3. Disposable glass capillaries.
4. Kimwipes.
5. Agarose.
6. Parafilm®.
7. 200 µL pipette and pipette tips.
8. 1.5 mL microcentrifuge tubes.
9. 15 mL tubes.
10. 100 kDa MWCO centrifugal filter device.

2.2 Protein Cage Loaded with Metal Ions

Ferritin cage from *Arachaeoglobus fulgidus* modified with alanine amino acid at positions 150 and 151 (AfFtnAA) is used as an example in this protocol for the calculation of relaxivity. The protein production, purification and iron loading protocol are described in published reports by Lim et al. [6, 9].

2.3 Chemicals

1. HEPES salt.
2. Sodium chloride.
3. Copper sulphate.
4. 99.9% D₂O.
5. Concentrated HNO₃.
6. Deionised water.
7. 1% agarose gel solution kept at 60 °C until use.
8. Fomblin® PFPE lubricants.

2.4 Buffer

Prepare the buffer for relaxivity calculations with composition of 25 mM HEPES, 50 mM NaCl, pH 7.4. Dissolve 5.96 g of HEPES and 2.9 g of NaCl salts in 1 L of deionised water. Adjust the pH of the solution with 4 M NaOH to 7.4. Filter the buffer with a 0.22 µm membrane using a vacuum filtration device.

2.5 Equipment

1. Bruker Avance NEO 400 MHz NMR equipped with Broad-band Observe (BBO) probe with Z-gradient.
2. Bruker Avance III HD 600 MHz (14.1 T) vertical wide-bore system with water-cooled triple-axis Micro 2.5 wide-bore (WB) gradient system with a maximum gradient strength per direction as 1.5 T/m at 60 A. Any other vertical bore system or horizontal bore system equipped with small animal volume coils can be used to generate comparable data. MicWB40 micro-imaging probe with two independent rf channels and a 30 mm quadrature coil can be used for MR imaging [14].
3. Inductively coupled plasma mass spectrometer (ICP-MS) (PerkinElmer).
4. Centrifuge machine (Eppendorf 5804R).

3 Methods**3.1 Iron Loading**

The iron salt was loaded into ferritin in aerobic and anaerobic (deoxygenated and 65 °C) conditions as reported previously [6, 9]. Iron (II) sulphate salt was added to a 0.1% HCl solution to prepare a 100 mM stock solution. For aerobic iron loading, 4800 moles per cage (maximum capacity) equivalent of iron (II) sulphate solution was added to a 0.1 mg/mL apo-ferritin sample in a 100 mM HEPES, 50 mM NaCl, pH 7.5 buffer solution. The iron salt solution was added in drop-wise manner to prevent protein aggregation and iron precipitation due to sudden increase in local iron concentration. For anaerobic iron loading, the HEPES buffer with apo-ferritin was first degassed by passing nitrogen (N_2) gas through the solution at 65 °C in a sealed flask for 30 min. Next, the 4800 moles per cage equivalent iron salt solution was added along with H_2O_2 in a 3:1 molar ratio using a syringe pump. Both the aerobic and anaerobically loaded samples were stirred gently with a magnetic bead while loading the iron salt solution to ensure homogenous iron concentrations throughout the solution. After loading iron, both samples were kept at room temperature for 1 h followed by overnight incubation at 4 °C. The following day, both samples were passed through to a desalting column (High Prep 26/10; Cytiva) to remove unbound iron. The iron-loaded ferritin samples were concentrated with 100 kDa MWCO centrifugal filter device (see Note 1).

3.2 Concentration Estimation of Ferritin Cage Bound Iron Using ICP-MS

1. Take 100 μL of ferritin protein sample loaded with metal ions in a 15 mL tube and add 300 μL of concentrated HNO_3 followed by incubation at room temperature for 4 h (*see Note 2*).
2. Add deionized water to 5 mL and filter with 0.22 μm syringe filter.
3. Measure the iron concentration with ICP-MS (*see Note 3*).

3.3 Sample Preparation for Relaxivity Measurement

Calculate the concentrations of ferritin cage-bound iron from ICP-MS results. Prepare 5–6 dilutions of ferritin cage with respect to bound iron concentration in the range of 50 μM –2.0 mM in HEPES buffer in 1.5 mL microcentrifuge tubes (*see Note 4*). Make these dilutions in a volume of 500 μL and add 10% D_2O for NMR measurements. For MRI, the samples are prepared similarly without D_2O .

3.4 Water Proton Relaxation Time Measurements using NMR

3.4.1 T_1 Relaxation

Put each sample dilution in a 5 mm NMR tube using a Pasteur pipette. Insert the sample in the NMR tube holder and place the tube in 400 MHz NMR equipped with BBO probe with Z-gradient (*see Note 5*).

1. Bruker NMR has in-built Topspin software package which controls the acquisition and processing of the experiments. Use the modified inversion recovery pulse for T_1 calculation as reported [15]. This pulse sequence overcomes the radiation damping that arises due to high water signal.
2. Set the variable delay (vd)-list with 12–15 delay points ranging from 0.1 to 15 s (*see Note 6*).
3. Set the number of delay points in the vd-list equal to number of points (TD) in F1 dimension.
4. Set the recycle delay (d1) to 15 s (*see Note 7*).
5. Repeat the **steps 1–4** for each ferritin cage dilution sample and run the experiments.

3.4.2 T_2 Relaxation

1. Use standard Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence for T_2 calculation.
2. Set the echo time (d20) to 1 millisecond (ms) and define the vc-list with 12–15 delay points ranging from 1 to 1500 ms (*see Note 6*).
3. Set the number of delay points in vc-list equal to the TD in F1 dimension and d1 = 15 s.
4. Repeat **steps 1–3** for each ferritin cage dilution sample and run the experiments.

3.4.3 Calculation of T_1 and T_2 Values

1. Use Topspin's in-built relaxation module for extracting the water proton peak intensity and fitting the curve to calculate the relaxation time, T_1 and T_2 . Use the longest delay point free induction decay (FID) in T_1 and the shortest delay point FID in T_2 experiments for phase correction. Extract the water peak (around 4.7 ppm) and save the peak intensity (area) for fitting the curve.
2. Use Topspin inversion recovery (invrec) and exponential decay (uxnmrt2) in-built functions for fitting the intensity (area) to calculate the T_1 and T_2 relaxation time, respectively (see Note 7).
3. Otherwise, one can use Origin/Matlab software to calculate the T_1 and T_2 values manually from the peak intensities of water protons at different delay points.

Plot the water proton intensities against delay points in vd-list and vc-list. To calculate the T_1 relaxation time, fit the curve to a non-linear inversion equation,

$$I(t) = a + b e^{(-t/T_1)}$$

where, $I(t)$ is the intensity of water proton peak at different time points (t) in vd-list and a and b are constants (Fig. 1a).

4. For T_2 , fit the water proton intensity to mono-exponential decay equation to calculate the relaxation time. The equation is

$$I(t) = a e^{(-t/T_2)} + c$$

where $I(t)$ is the intensity of water proton peak at different time points (t) in vc-list, a and c are constants (Fig. 1b).

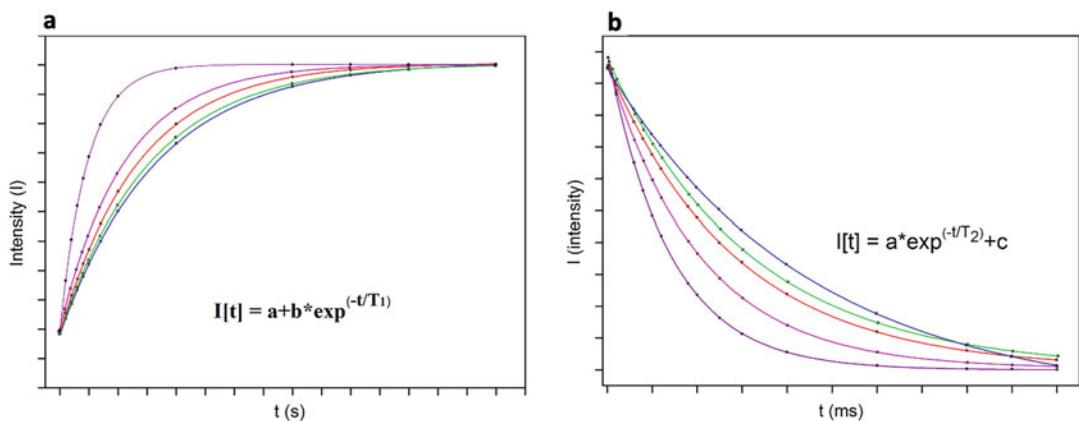


Fig. 1 Relaxation time calculation: typical curve fitting of intensity against time plot for (a) T_1 and (b) T_2 calculations

3.5 Water Proton Relaxation Time Measurement Using MRI

3.5.1 T_1 and T_2 Pulse Sequences Used for the Generation of MRI Data

3.5.2 Protocol for T_1 and T_2 Relaxation Measurements

Spin-echo sequences are used to acquire quantitative data for T_1 and T_2 time calculations. During spin-echo pulse sequences, a 90° radiofrequency (RF) pulse is followed by a delay of half an echo before a 180° re-focusing pulse is applied. In multi-slice multi-echo (MSME) and rapid acquisition with relaxation enhancement (RARE) sequence, a slice-selective RF pulse generates multiple individually phase-encoded echoes, creating a T_2 -weighted image. The echo time (TE) is the time from the beginning of the RF pulse to the maximum induced signal. The repetition time (TR) is the duration of one entire pulse. MSME and RARE experiments, both based on the CPMG pulse sequence, are used to acquire quantitative T_2 relaxation data for *in vitro* studies. With variable repetition times, RARE can also generate quantitative information for T_1 relaxation. While T_2 and T_2^* imaging are considered conventional for iron-loaded ferritin seen as hypointense or dark spots in MRI, incorporation of metals such as Gd and Mn allows for T_1 imaging, leading to hyperintense or bright regions [16] (*see Note 8*).

1. *Probe set-up for MRI:* A suitable RF coil is mounted on the microimaging probe, aligned with the trimmer rods on the probe and fixed using non-magnetic screws. The trimmer rods should be unobstructed for proper tuning and matching of the probe.
2. *Phantom experiments:* A 10 mm NMR tube filled with 20% H_2O , 80% D_2O and 1 g/L $CuSO_4$ is used to acquire basic localizer scans to identify the optimal sample location at the centre of the magnet and for the overall set-up of basic parameters.
3. *Preparation of samples:* The glass capillaries are attached in front of a 200 μL pipette tip using narrow tubing. Each dilution sample of iron-loaded ferritin protein cages is suctioned into each capillary and the capillary tube is sealed using either parafilm or heat. Between 5 and 10 of such capillaries can be imaged simultaneously depending on their diameter. Once the samples are bundled, the capillaries are tied together using parafilm to ensure no movements during MR image acquisition. The samples are then transferred to the MR tube which is placed into the RF coil.
4. *MRI tube setup:* A MRI compatible glass tube with outer diameter 30 mm is filled with 1% agarose gel to a height of 2.5 cm from the bottom to increase the base height for aligning the samples with the centre of the magnet. The 30 mm tube fits inside the Bruker micro-imaging RF-inserts. Samples prepared in **step 3** are fixed on the top of the agarose gel scaffold. The location of each tube with respect to the control (buffer) is noted and the capillary tubes are fixed in their location using

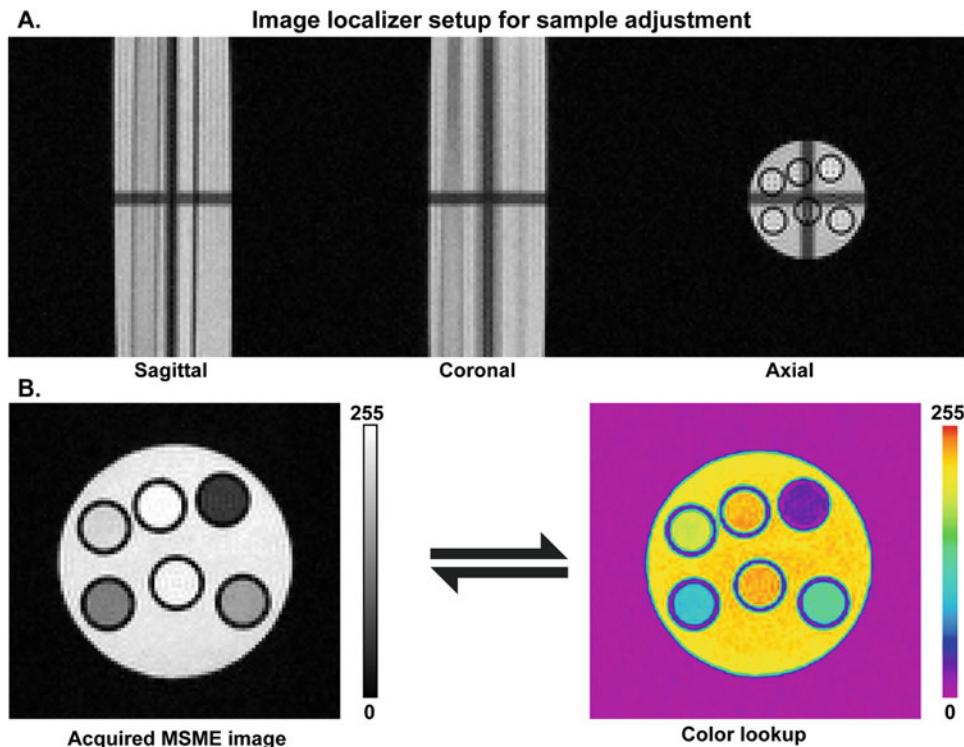


Fig. 2 Resulting images from MRI data acquisition for different concentrations of ferritin samples. (a) Localizer setup in all three planes to identify slice location to be imaged. (b) T_2 Image acquired from MSME scan shows hypointense regions in samples with higher iron content. Contrast differences may be easily detected in colored lookup display

Parafilm® and tissue (see Note 9). The glass tube is covered at the top using a parafilm membrane and mounted on the MRI probe.

5. *Connecting the MRI probe:* The MRI probe is mounted within the gradient coil by inserting from the bottom of the vertical bore machine such that the centre of the sample tubes aligns with the centre of the magnet. Once the probe is fixed at the location, MRI data acquisition can start.
6. *Starting a study:* Paravision v6.0.1 software is used for MRI data acquisition. A new study is created with the description of the material with the date and location. An examination window is opened within this study to run all the scan adjustments and start the acquisition.
7. *Scan local adjustments:* A simple fast low angle shot (FLASH) scan is acquired as the localizer experiment in X, Y, and Z directions (axial, sagittal or coronal; equivalent to planes XY, XZ and YZ) to ensure that the sample tube is well aligned to the centre of the magnet (Fig. 2a). Any required arrangements of

the slice selection and orientation are done based on this experiment. If the sample region in the tube is too far from the centre, it is re-arranged to align with the magnet and gradient centre to avoid MRI artefacts. Adjustment settings including the wobble adjustment and shim values are optimised. The RF pulse power is calibrated as the reference. The global shim values for the study and local shim values for the field of view (FOV) in each scan are optimised using the B0 map.

8. *MR image acquisition for T_1 and T_2 relaxation:* For T_1 and T_2 data collection in vitro, MSME or RARE scans are acquired (Fig. 2b). The parameters which need to be adjusted include TE, TR, the total number of echoes and repetitions, FOV, number of slices, slice thickness and orientation and matrix resolution. In the case of RARE pulse sequences, the RARE factor can be adjusted (as multiples of 2, within the limits of 2–8 to avoid motion artefacts) so that the overall acquisition time may be decreased, if needed. For T_1 calculation, each slice is acquired with a single echo with variable repetition times, whereas T_2 can be calculated when multiple echoes are acquired.

3.5.3 Data Processing

Generating quantitative data from MRI images: Once the images are successfully acquired, the data are exported to the “Image Display and Processing” tool within Paravision, by selecting “View in Image Display.” Each sample condition within the image is selected as a region of interest (ROI) with an identical area and avoiding any selection near the edges or outside each capillary tube (Fig. 3). Image sequence analysis on the selected ROIs allows us to identify the data points to generate T_1 and T_2 relaxation curves. These values can be exported for further analysis in Microsoft Excel, MATLAB, Origin, or other data processing software as mentioned in Subheading 3.4.3 (Fig. 1).

3.6 Relaxivity r_1 and r_2 Calculation from T_1 and T_2 Values

Plot the inverse of relaxation time (s^{-1}), $1/T_1$ and $1/T_2$ values on the Y-axis against ferritin cage iron concentrations and fit to a straight line. The slope of the line is the relaxivity (r_1 or r_2) value with unit $\text{mM}^{-1} \text{ s}^{-1}$ for the ferritin cage contrast agent (Fig. 4).

4 Notes

1. Ferritin cages can be loaded with other non-native paramagnetic ions such as Gd and Mn [5]. We can use the same protocol for measuring the relaxivity of other protein cages loaded with non-native metals.
2. The concentration of ferritin cage loaded with metal ions should be adjusted to the detection limit of ICP-MS.

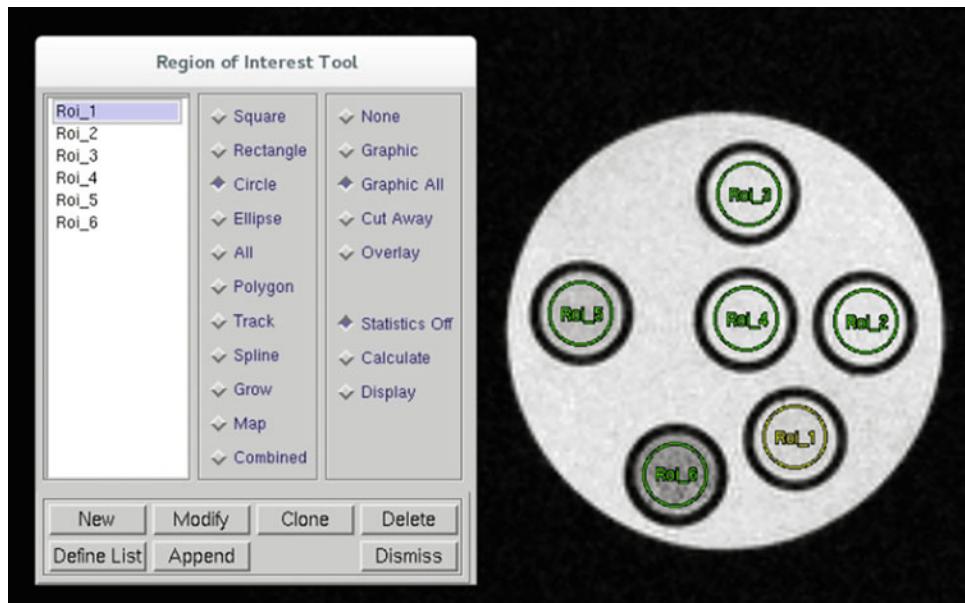


Fig. 3 Selection of regions of interest (ROI) for image sequence analysis. Each sample concentration is selected as separate ROI to obtain relaxation plot values, which are fitted into relaxation equation curves

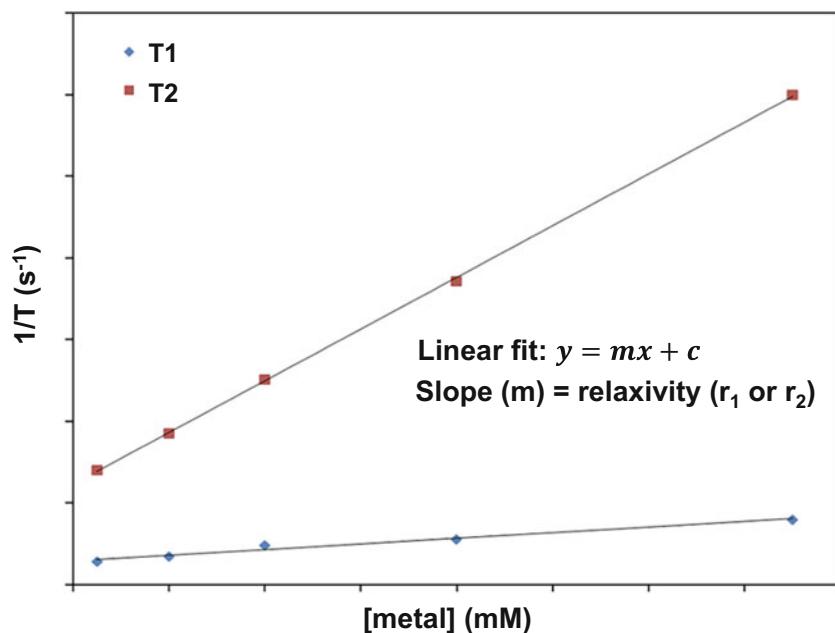


Fig. 4 Relaxivity measurement: relaxation rates ($1/T$) are plotted against the contrast agent concentrations for T_1 (blue diamond) and T_2 (red square)

3. ICP-MS gives the metal ions concentration in ppm or ppb. Divide the metal concentration by iron metal molar mass (metal conc. (mg/mL)/55.85 (mg/mol) to calculate the molar concentration of iron in the sample.
4. The dilutions concentration range of a contrast agent should be adjusted according to its contrast enhancement property and the magnetic field strength of the equipment. Lower concentration is used for a contrast agent with higher relaxivity. Moreover, higher magnetic field NMR/MRI machine requires lower concentration of contrast agents.
5. The relaxivity of a contrast agent also depends on the magnetic field strength of the MRI/NMR machine. Therefore, we cannot compare the relaxivity of contrast agents obtained from two different magnetic field strengths.
6. Provide enough number of delay points in the vd- and vc-lists which would help in the fitting of water peaks intensity for the calculation of T_1 and T_2 relaxation time.
7. Ideally, the relaxation delay ($d1$) should be five times that of T_1 relaxation time of water protons in NMR experiments. Topspin in-built functions, invrec and uxnmrt2, use the following equations to fit the water proton peak intensities to curve.

$$\text{invrec} = \text{Inversion recovery } (T_1) : I(t) = I(0) \left(1 - 2Ae^{-\frac{t}{T_1}} \right)$$

$$\text{uxnmrt2} = \text{CPMG}(T_2) : I(t) = Pe^{-\frac{t}{T_2}}$$

where, $I(t)$ = intensity of water proton at time point (t), A and P are constants.

8. Gradient Echo pulse sequences are used to calculate iron content in ferritin proteins inside living tissues. In a gradient echo pulse sequence, each RF pulse is applied during a slice-select gradient, using a variable flip angle with frequency-encoded gradient and iterating phase-encoding gradients which generate gradient echoes. For a FLASH sequence, the TE and TR are both short time intervals and the contrast of the MR image is largely independent of T_1 and T_2 time of the sample. A multiple gradient echo (MGE) sequence consists of a train of identically phase-encoded gradient echoes after each slice-selective excitation. MGE can allow for the measurement of iron concentration and apparent transverse relaxation time (T_2^*) in different tissues, which are relevant to the study of ferritin in biological samples [17].
9. Note: Kimwipes immersed in Fomblin® can also be used to fix the capillaries if they are not too delicate. The void in the glass tube outside the capillaries can be filled with Fomblin® to a volume slightly above the samples to improve the contrast-to-noise ratio.

Acknowledgement

We acknowledge the funding support from NTU-Northwestern Institute for Nanomedicine located at the International Institute for Nanotechnology, Northwestern University, USA and the Nanyang Technological University, Singapore (Grant No. M4081504.F40.706022). MRI experiments were conducted at NTU Centre for High Field NMR Spectroscopy and Imaging.

References

1. Wood ML, Hardy PA (1993) Proton relaxation enhancement. *JMRI: J Magn Reson Imaging* 3:149–156
2. Wahsner J, Gale EM, Rodriguez-Rodriguez A, Caravan P (2019) Chemistry of MRI contrast agents: current challenges and new Frontiers. *Chem Rev* 119:957–1057
3. Pellico J, Ellis CM, Davis JJ (2019) Nanoparticle-based paramagnetic contrast agents for magnetic resonance imaging. *Contrast Media Mol Imaging* 2019:1845637
4. Uchida M, Terashima M, Cunningham CH, Suzuki Y, Willits DA, Willis AF, Yang PC, Tsao PS, McConnell MV, Young MJ, Douglas T (2008) A human ferritin iron oxide nanocomposite magnetic resonance contrast agent. *Magn Reson Med* 60:1073–1081
5. Ruggiero MR, Alberti D, Bitonto V, Crich SG (2019) Ferritin: a platform for MRI contrast agents delivery. *Inorganics* 7:33
6. Nandwana V, Ryoo SR, Kanthala S, Kumar A, Sharma A, Castro FC, Li Y, Hoffman B, Lim S, Dravid VP (2017) Engineered ferritin nanocages as natural contrast agents in magnetic resonance imaging. *RSC Adv* 7:34892–34900
7. Klem MT, Young M, Douglas T (2005) Biomimetic magnetic nanoparticles. *Mater Today* 8:28–37
8. Lawson DM, Treffry A, Artymiuk PJ, Harrison PM, Yewdall SJ, Luzzago A, Cesareni G, Levi S, Arosio P (1989) Identification of the ferroxidase center in ferritin. *FEBS Lett* 254: 207–210
9. Kumar A, Nandwana V, Ryoo S-R, Ravishankar S, Sharma B, Pervushin K, Dravid VP, Lim S (2021) Magnetoferitin enhances T2 contrast in magnetic resonance imaging of macrophages. *Mater Sci Eng C* 128:112282
10. Kim HS, Woo J, Lee JH, Joo HJ, Choi Y, Kim H, Moon WK, Kim SJ (2015) In vivo tracking of dendritic cell using MRI reporter gene, ferritin. *PLoS One* 10:e0125291
11. Li L, Fang CJ, Ryan JC, Niemi EC, Lebron JA, Bjorkman PJ, Arase H, Torti FM, Torti SV, Nakamura MC, Seaman WE (2010) Binding and uptake of H-ferritin are mediated by human transferrin receptor-1. *Proc Natl Acad Sci U S A* 107:3505–3510
12. Sana B, Lim S (2015) Determining the relaxivity values of protein cage-templated nanoparticles using magnetic resonance imaging. In: Orner BP (ed) Protein cages: methods and protocols. Springer New York, New York, pp 39–50
13. Laufer RB (1987) Paramagnetic metal-complexes as water proton relaxation agents for NMR imaging—theory and design. *Chem Rev* 87:901–927
14. Oerther T (2012) Micro imaging for AVANCE III systems user guide. Bruker BioSpin GmbH, Rheinstetten
15. Eykyn TR, Payne GS, Leach MO (2005) Inversion recovery measurements in the presence of radiation damping and implications for evaluating contrast agents in magnetic resonance. *Phys Med Biol* 50:N371–N376
16. Szulc DA, Lee XA, Cheng HYM, Margaret HL (2020) Bright ferritin-a reporter gene platform for on-demand, longitudinal cell tracking on MRI. *iScience* 23:101350
17. Gossuin Y, Muller RN, Gillis P (2004) Relaxation induced by ferritin: a better understanding for an improved MRI iron quantification. *NMR Biomed* 17:427–432



Chapter 21

Crystalline Biohybrid Materials Based on Protein Cages

Hendrik Böhler, Michael Rütten, Laurin Lang, and Tobias Beck

Abstract

Highly ordered superstructures of nanomaterials can be synthesized using protein cages as templates for the assembly of inorganic nanoparticles. Here, we describe in detail the creation of these biohybrid materials. The approach involves computational redesign of ferritin cages, followed by recombinant protein production and purification of the new variants. Metal oxide nanoparticles are synthesized inside the surface-charged variants. The composites are assembled using protein crystallization to yield highly ordered superlattices, which are characterized, for example, with small angle X-ray scattering. This protocol provides a detailed and comprehensive account on our newly established strategy for the synthesis of crystalline biohybrid materials.

Key words Ferritin protein cage, Protein crystallization, Highly ordered nanoparticle superlattices, Metal oxide nanoparticles, Protein design

1 Introduction

The synthesis of nanomaterials with defined composition and structure is a challenging task, but paramount for the development of multifunctional materials with applications in sensing, information transport, catalysis, energy storage and light harvesting [1–5]. Recently, nanoparticle-based materials have received great interest, as the combination of different nanoparticles in the solid state can create materials with novel properties that emerge from the interactions of the building blocks [6].

Biomolecules such as protein cages offer a unique approach for the assembly of nanomaterials, because they have a defined size and shape. We have recently shown that redesigned protein cages can serve as an atomically precise ligand shell for the assembly of inorganic nanoparticles into highly ordered binary superlattices [7]. To this end, we utilized two variants of the protein cage ferritin: One of them carries a positive surface charge, the other one a negative surface charge. The protein cage as a template overrides any imperfection of the nanoparticles. By self-assembly of

protein cages loaded with metal oxide nanoparticles, we create nanoparticles superlattices with high order. Moreover, these superlattices show catalytic activity in the solid state [8].

Here, we describe the protocol for the assembly of inorganic nanoparticles into binary superlattices using ferritin cages. We start with the protein cage redesign using the Rosetta software to create two protein variants, namely $\text{Ftn}^{(\text{pos})}$ and $\text{Ftn}^{(\text{neg})}$, with positive or negative surface charge, respectively. The production and purification are described in detail in the following section. In the next step, the purified variants are utilized as a size-constrained vessel for the synthesis of metal oxide nanoparticles. Subsequently, protein crystallization techniques are applied to assemble the composite building blocks into highly ordered binary superlattices. These superlattices are characterized, for example with small angle X-ray scattering, as described in the final section.

The discussed approach represents a generalizable strategy for the assembly of cargo molecules into highly ordered materials. As protein cages can be packed with a diverse set of cargos – such as nanoparticles [9] as well as molecules such as fluorophores or catalytic complexes – a range of multifunctional materials based on the presented strategy can be envisioned in the future [10–13].

2 Materials

2.1 Materials for Computational Design of Supercharged Ferritin Variants

The latest version of the Rosetta software can be downloaded at www.rosettacommons.org. The license is free of charge for academic users and can be acquired on the same website (*see Notes 1–3*). Additionally, the crystal structure of the desired protein is required in PDB format. In this example, the structure file for human heavy chain ferritin (PDB accession code: 2CEI) is used.

2.2 Materials for Protein Production and Purification of Supercharged Ferritin Variants

Prepare all solutions using ultrapure water (e.g., prepared with a Purelab Flex 2 system, resistivity 18.2 MΩ cm) and analytical grade reagents where possible unless indicated otherwise. Prepare and store all reagents at room temperature (unless indicated otherwise). All buffer solutions as well as raw protein stocks (before applying FPLC methods) are filtered by either syringe or bottle top filtration (syringe filter units: PVDF, sterile, pore size 0.22 µm; bottle top filter membranes: mixed cellulose esters, pore size 0.2 µm).

2.2.1 Cell Culture

All cell culture work is conducted while observing appropriate aseptic techniques.

1. Ampicillin stock solution (100 mg/mL): 4 g of ampicillin sodium salt (cell culture grade) are dissolved in 40 mL ultrapure water. The resulting solution is sterilized using sterile

syringe filter units with 0.22 µm pore size. Sterilized solution is distributed into 20 × 2 mL sterile plastic tubes and stored at –20 °C.

2. 0.2 M IPTG stock solution: 1.9 g isopropyl-β-D-thiogalactopyranoside are dissolved in 10 mL of ultrapure water. The resulting solution is sterilized using sterile syringe filter units with 0.22 µm pore size. Sterilized solution is distributed into 5 × 2 mL sterile plastic tubes and stored at –20 °C.
3. Culture media: LB as well as TB media are prepared from premixed powder according to supplier instructions using deionized water in 1–2 L glass bottles utilizing ~70% of the bottle volume to minimize spill over. Stirring is applied until the powder is completely dissolved. Bottles are labelled with autoclave tape and loosely covered with their lids to ensure successful sterilization in the autoclave. A suitable autoclave and program setting are then used to sterilize the media. Sterile media is stored at room temperature.
4. Vector: pET-22b(+) vector containing the Ftn^{(pos)/(neg)} gene.
5. Cells: *E. coli* BL21-Gold (DE3).

2.2.2 Protein Purification

1. RNase A: Store at –20 °C.
2. Ammonium sulfate solution, saturated: Ammonium sulfate is added to 600 mL of ultrapure water in a 1 L glass bottle while strong stirring is applied, until a substantial amount remains undissolved to ensure a saturated solution has been prepared. This process can take up to 20 min as the solubility of ammonium sulfate in water of ~750 g/L is relatively high.
3. Ftn^(neg) ion exchange buffer (low salt), 50 mM Tris–HCl, pH 7.5, 0.15 M NaCl: 6.06 g Tris-(hydroxymethyl)-aminomethane (Tris) and 8.8 g sodium chloride (NaCl) are weighed into a 1 L bottle. 900 mL ultrapure water are added and the pH is adjusted to pH 7.5 using 5 M HCl. The final volume is adjusted to 1000 mL using ultrapure water.
4. Ftn^(neg) ion exchange buffer (high salt), 50 mM Tris–HCl, pH 7.5, 1.0 M NaCl buffer: 6.06 g Tris and 58.44 g NaCl are weighed into a 1 L bottle. 900 mL ultrapure water is added and pH is adjusted to pH 7.5 using 5 M HCl. The final volume is adjusted to 1000 mL using ultrapure water.
5. Ftn^(pos) ion exchange buffer (low salt), 50 mM MES, pH 6, 0.5 M NaCl: 9.6 g 2-morpholin-4-ylethanesulfonic acid monohydrate (MES) and 26.3 g NaCl are weighed into a 1 L bottle. 900 mL ultrapure water are added and the pH is adjusted to pH 6.0 using 5 M sodium hydroxide solution (NaOH). The final volume is adjusted to 1000 mL using ultrapure water.

6. $\text{Ftn}^{(\text{pos})}$ ion exchange buffer (high salt), 50 mM MES, pH 6, 1.5 M NaCl: 9.6 g MES and 78.9 g NaCl are weighed into a 1 L bottle. 900 mL ultrapure water are added and the pH is adjusted to pH 6.0 using 5 M NaOH solution. The final volume is adjusted to 1000 mL using ultrapure water.
7. $\text{Ftn}^{(\text{pos})}$ size exclusion buffer: same as item 4.
8. $\text{Ftn}^{(\text{neg})}$ size exclusion buffer, 50 mM Tris–HCl, pH 7.5, 0.3 M NaCl: 6.06 g Tris and 17.53 g sodium chloride (NaCl) are weighed into a 1 L bottle. 900 mL ultrapure water are added and the pH is adjusted to pH 7.5 using 5 M HCl. The final volume is adjusted to 1000 mL using ultrapure water.
9. Ion exchange columns: $\text{Ftn}^{(\text{pos})}$ – 5 mL Sepharose HiTrap SP HP cation exchange column (GE Healthcare)/ $\text{Ftn}^{(\text{neg})}$ – 5 mL Sepharose HiTrap SP Q anion exchange column (GE Healthcare).
10. Size exclusion column: HiLoad 16/600 Superdex 200 PG gel filtration column (GE Healthcare).
11. Centrifugal filter units: Amicon Ultra-15 centrifugal filter units, 30 kDa Molecular weight cut-off (Merck Millipore).

2.3 Materials for Synthesis of Metal Oxide Nanoparticles Inside the Ferritin Cavity

2.3.1 Iron Oxide Nanoparticles

Prepare all solutions using ultrapure water (e.g., prepared with a Purelab Flex 2 system, resistivity 18.2 MΩ cm) and analytical grade reagents. Prepare and store all reagents at room temperature (unless indicated otherwise). All solutions are filtered and depleted of O₂ by bubbling N₂ through the solutions for at least 15 min before usage. Diligently follow all waste disposal regulations when disposing waste materials. Amicon Ultra-15 centrifugal filter units (30 kDa MWCO) by Merck Millipore are used for rebuffering.

1. $\text{Ftn}^{(\text{neg})}$ buffer, 50 mM Tris–HCl, pH 7.5, 0.3 M NaCl: 6.06 g Tris-(hydroxymethyl)-amino methane (Tris) and 17.53 g sodium chloride (NaCl) are weighed into a 1 L bottle. 900 mL ultrapure water are added and the pH is adjusted to pH 7.5 using 5 M HCl. The final volume is adjusted to 1000 mL using ultrapure water.
2. $\text{Ftn}^{(\text{pos})}$ buffer, 50 mM Tris–HCl, pH 7.5, 1.0 M NaCl: 6.06 g Tris and 58.44 g NaCl are weighed into a 1 L bottle. 900 mL ultrapure water are added, then the pH is adjusted to pH 7.5 using 5 M HCl. The final volume is adjusted to 1000 mL using ultrapure water.
3. H₂O₂: H₂O₂ is added to ultrapure water until a concentration of 6 mM is reached.
4. (NH₄)₂Fe(SO₄)₂: 353 mg of ammonium iron(II)sulfate hexahydrate are dissolved in 50 mL ultrapure water to achieve a concentration of 18 mM.

5. Sodium citrate: Prepare a 300 mM sodium citrate stock solution by dissolving 4.41 g sodium citrate dihydrate in 50 mL ultrapure water.

2.3.2 Cobalt Oxide Nanoparticles

1. $\text{Ftn}^{(\text{neg})}$ and $\text{Ftn}^{(\text{pos})}$ buffer, 100 mM HEPES, pH 8.3, 0.2 M NaCl: 2.38 g *N*-2-hydroxyethylpiperazin-*N'*-2-ethane sulfonic acid (HEPES) and 1.17 g NaCl are weighed into a 250 mL bottle. 200 mL ultrapure water are added and the pH adjusted to pH 8.3 using 5 M NaOH. The final volume is adjusted to 250 mL using ultrapure water.
2. H_2O_2 : H_2O_2 is added to ultrapure water until a concentration of 10 mM is reached.
3. $(\text{NH}_4)_2\text{Co}(\text{SO}_4)_2$: 593 mg of ammonium cobalt(II) sulfate hexahydrate are dissolved in ultrapure water to achieve a concentration of 30 mM.
4. Sodium citrate: Prepare a 300 mM sodium citrate stock solution by dissolving 4.41 g sodium citrate dihydrate in 50 mL ultrapure water.

2.3.3 Cerium Oxide Nanoparticles

1. $\text{Ftn}^{(\text{neg})}$ buffer with 50 mM Tris-HCl pH 7.5, 0.3 M NaCl: 6.06 g Tris-(hydroxymethyl)-amino methane (Tris) and 17.53 g sodium chloride (NaCl) are weighed into a 1 L bottle. 900 mL ultrapure water are added and the pH is adjusted to pH 7.5 using 5 M HCl. The final volume is adjusted to 1000 mL using ultrapure water.
2. $\text{Ftn}^{(\text{pos})}$ buffer with 50 mM Tris-HCl pH 7.5, 1.0 M NaCl buffer: 6.06 g Tris and 58.44 g NaCl are weighed into a 1 L bottle. 900 mL ultrapure water are added, then the pH is adjusted to pH 7.5 using 5 M HCl. The final volume is adjusted to 1000 mL using ultrapure water.
3. H_2O_2 : H_2O_2 is added to ultrapure water until a concentration of 15 mM is reached.
4. CeCl_3 : 559 mg of Cer(III)-chloride heptahydrate are dissolved in 50 mL ultrapure water to achieve a concentration of 30 mM.
5. EDTA: Prepare a 500 mM ethylenediaminetetraacetic acid (EDTA) stock solution by dissolving 7.31 g EDTA in 50 mL ultrapure water.

2.4 Materials for Protein Crystallization

1. Reservoir solution PDB ID 5JKL, 0.19 M Magnesium formate.
2. Reservoir solution PDB ID 5JKM, 0.2 M Magbesium format, 0.1 M Tris-HCl, pH 8.5.

3 Methods

3.1 Methods for Computational Design of Supercharged Ferritin Variants

In the first part, the computational design of a supercharged protein variant of the protein nanocage ferritin will be described using the Rosetta molecular modelling software suite (*see* Fig. 1). Supercharging with the help of the Rosetta software was first described by Miklos et al. for supercharging of globular proteins (antibodies) [14]. Here, as an example, the protocol will be described for a positively supercharged variant of human heavy chain ferritin [15]. Figure 1 shows the schematic workflow for introducing mutations using Rosetta software.

3.1.1 Preparing the Input

To prepare the structure file, every entry in the PDB-file except the lines starting with ATOM needs to be deleted. Also, heteroatoms (lines starting with HETATM) should be removed. For symmetric molecules, such as the protein cage ferritin, it is often advisable to generate a symmetry-definition file. In this case, the lines starting from CRYST1 to the first ATOM line remain in the PDB file.

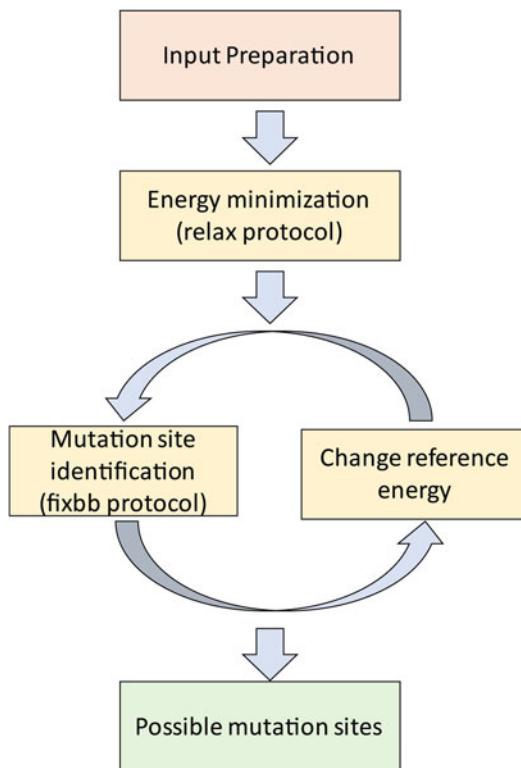


Fig. 1 Flow diagram for computational protein supercharging

3.1.2 Generate a Symmetry Definition File

In the case of a symmetric molecule, a symmetry definition file can be created with the *make_symmdef_file.pl* script, which is part of the Rosetta software suit.

1. Copy the input PDB file in a new directory
2. Run the following command to generate the symmetry definition file:

```
$your_Path_to_Rosetta/main/source/src/apps/public/symmetry/make_symmdef_file.pl -m CRYST -r 12.0 -s P 4 3 2 -p 2CEI.pdb > symmetry_definition-file.symm (see Note 4)
```

This command generates a symmetry definition file for the ferritin cage. This file can be added to every future Rosetta protocol to generate six additional surrounding subunits. Therefore, the original subunit in the middle has the same environment (neighboring protein chains) as there are in a full cage. This approach makes sure that the simulations are quite close to the true structure but saves resources because not the full cage needs be calculated.

Setting up a symmetry definition file is not trivial and very different from protein to protein. Especially if the input PDB contains different chains, a lot of trial and error is needed to generate a proper symmetry definition file.

3.1.3 Energy Minimization of the Input Structure

To get reliable results, it is advisable to carry out an energy minimization step of the input structure before further use within Rosetta. To do this, the Rosetta molecular modeling package offers a pre-defined protocol called *relax*.

1. Add the prepared input file and if desired the symmetry definition file in a directory where you wish to perform the energy minimization.
2. Prepare an options file for the planned Rosetta run by creating a text file containing the following options (*see Note 5*).

```
# Path to the Rosetta databases
-database $your_Path_to_Rosetta/main/database

# define the input structure for the relax run
-s Cleaned-PDB.pdb

# Rotamer packing options (see Note 6)
-ex1
-ex2
-use_input_sc
```

```

# Add coordinate constraints to the backbone based
on the # input structure
-constrain_relax_to_start_coords

# sets the number of output structures
-nstruct 50

# Define symmetry definition file
- symmetry::symmetry_definition
symmetry_definition-file.symm

```

In the following, this text file will be named *options_relax*, but in principle, it can have any name.

3. Start the program with the following command

```
$your_Path_to_Rosetta/main/source/bin/relax.linuxgccrelease @options_relax (see Note 7)
```

4. When the calculation finished successfully, new PDB files named like the input PDB with a number at the end ranging from 1 to the number given with the nstruct option, and a file called score.sc are created in the directory. The score.sc file contains for each PDB output the total Rosetta score and all energy terms. This file can be ranked from the lowest to the highest total score with the following command.

```
grep -H 'SCORE' score.sc |sort -k2 -n|head -n20|nl -w 3 -s '' >Output-file
```

The total score of your structures should be negative assuring a stable protein. As an example, the total Rosetta score of a single subunit of human heavy chain ferritin has a total Rosetta score of around -700. If a symmetry definition file is used, the total score is around -1250 for the ferritin cage. Since the Rosetta run is based on random changes of the backbone, multiple structures (called trajectories) need to be generated. For the *relax* run, at least 50 structures should be generated.

If the total scores of the trajectories with the lowest energy in the ordered score file have roughly the same values, the number of trajectories is sufficient. The protein structure with the lowest total score is used as an input for the next step. Important: If a symmetry definition file was used, all additional copies of the original subunit need to be deleted from that PDB.

3.1.4 Identification of Possible Mutation Sites

To determine possible mutations sites for the introduction of charged amino acids, the Rosetta fixed backbone (*fixbb*) protocol is used. Here, the protein backbone is fixed and only the sidechains

and their conformation are altered by the program. In contrast to the *relax* protocol, Rosetta is also allowed to change the side chains of amino acids at specified positions.

1. Add the protein structure with the minimized energy and if desired the symmetry definition file in a directory where you wish to perform the calculation. In contrast to the *relax* protocol, you need two additional files, namely a weights file and a residue file.
2. The weights file contains the weight for the different energy terms of the Rosetta scoring function. For the first run, it is enough to just copy the most recent weights file from rosettacommons.org/demos/latest/tutorials/scoring/scoring. On this site, more in-depth information about the different terms of the weights file can be found. In the following, the weights file will be named *weights.wts*
3. The residue file or *resfile* is used to specify the amino acid positions within the protein chain that are allowed for mutations. An example for the first lines of the *resfile* for the supercharging of ferritin is shown in the following:

```
# default command for every position not included
in the # resfile

NATAA # allow only the native amino acid (NATive
Amino Acid)

# list of positions in the protein which derivates
from the # default behavior

start
19 A PIKAA AKR
...
```

After the line with *start*, all positions in the input structure where a mutation is allowed are listed. Besides the position number also the chain identifier—in this case A—is needed. There are different ways to specify the possible mutations (*see Note 8*). In this case, the *PIKAA* (pick amino acid) command is chosen, after which a list of amino acids (in one letter code) is given. In the case of the positive supercharging of ferritin, it is arginine (R) and lysine (K). To give Rosetta the choice to keep the native amino acid, also the native amine acid, here alanine (A), is added to the list. Only these amino acids can be introduced at this position. For all other positions, the default behavior specified beforehand with *NATAA* is applied, which allows no further mutation and therefore keeps the native amino acid.

4. Before the calculations can be started, a new option file with the following content must be prepared.

```
# Path to the Rosetta databases
-database $your_Path_to_Rosetta/main/database

# define the input structure for the fixbb run
-s energy-minimized.pdb

# define the residue file
-resfile name_of_resfile

# define the weights file
-score:weights weights.wts

# Rotamer packing options (see Note 7)
-ex1
-ex2
-use_input_sc

# sets the number of output structures
-nstruct 500

# Define symmetry definition file
-symmetry::symmetry_definition symmetry_definition-file.symm
```

The file is saved as *fixbb_options* and stored in the same directory as the other files.

5. Now the working directory should contain the energy-minimized PDB file, the *resfile*, the *weights.wts* file, the *fixbb_options* file and if needed also the symmetry definition file. The fixbb protocol can be started with the following command

```
$your_path_to_Rosetta/main/source/bin/fixbb.linuxgccrelease @options_fixbb
```

6. When the protocol finishes successfully, new protein structure files and a score.sc file should appear in the directory. Similar to the relax run, the score.sc file should be sorted according to the total score. The amino acid sequence for the best structures can be extracted from the PDB files with the following command

```
perl $your_Path_to_Rosetta/main/source/ scripts/python/public/pdb2fasta.py List_of_PDBs > AA_sequence.fasta
```

The sequences are compared to the wild type and any introduced mutations can be detected.

7. In the case of the positively supercharged ferritin, only three mutation sites (C90K, C102K and H104K) are found in the initial run. To increase the number of possible mutation sites, the reference energies in the *weights.wts* file can be manipulated to favor the desired amino acids arginine and lysine. In the weights file, there is a line starting with “METHOD_WEIGHTS ref” followed by 20 numbers. Each number represents a reference energy for an amino acid (ordered alphabetically using the one letter code). By reducing the reference energy value (i.e., making it more negative) for a desired amino acid, it is more likely introduced in the corresponding Rosetta run. It is advisable to alter the energy in small increments to get an understanding which mutation are more favorable (*see Note 9*) [15].
8. The steps 1–6 can be repeated with different sets of METHOD_WEIGHTS reference energies. The amino acid sequences of structures with the lowest total Rosetta score can be listed after one another and based on this, the desired mutations can be selected. With this method, additional mutations (A18K, N25R, C90K, N98R, C102K, H105K, N109K, D123K, E162R) were selected to generate a positive supercharged variant of the human heavy chain variant. This variant will be called Ftn^(pos) in the following.

To engineer the negatively supercharged variant, the same procedure is repeated, but instead mutations for glutamic and aspartic acid were allowed. Mutations were carried out at the position A18E, C90E, C102E, H105E (*see Note 10*). This variant will be called Ftn^(neg) in the following.

After the mutation sites are identified, the mutations can be introduced in the ferritin gene with common PCR techniques, for example, QuikChange [16]. Alternatively, by using commercial gene synthesis or overlap extension PCR, the gene for the newly designed variants can be synthesized and cloned into the expression vector of interest [17].

3.2 Methods for Protein Production and Purification of Supercharged Ferritin Variants

In the following part, the production and purification of the newly designed supercharged ferritin variants are described.

Carry out all procedures at room temperature unless otherwise specified. All procedures on cell culture, protein production as well as purification are carried out in biosafety level S1 laboratories. Diligently follow all waste disposal regulations when disposing waste materials, especially concerning biosafety regulations for bacterial culture/gene modification work.

3.2.1 Protein Production of *Ftn*^(pos) While working with cells, observe proper aseptic techniques.

1. Chemically competent production strain *E. coli* cells are transformed with a pET-22b(+) vector containing the *Ftn*^(pos) gene. Transformed cells are plated on ampicillin containing LB agar plates to select for successful transformation. Plates are incubated at 37 °C overnight and then checked for colonies. Plates containing colonies are stored at 4 °C for 2–4 weeks.
2. Precultures are prepared by touching a single colony on one of the prepared LB-Agar plates with a sterile disposable pipette tip and subsequent ejection of that tip into a 15 mL plastic centrifuge tube containing 10 mL LB medium supplemented with 15 µL of 100 mg/mL ampicillin stock solution.
3. Precultures are incubated overnight (after being prepared in the afternoon ~16 h) at 37 °C while shaking at 230 rpm.
4. Before starting the production, culture incubators are pre-heated to 37 °C.
5. 400 mL of LB medium are filled into sterile 2 L conical cell culture flasks. The media is supplemented with 600 µL of 100 mg/mL ampicillin stock solution. The production culture is started by adding 4 mL (1% v/v) of previously prepared preculture to the culture flask (*see Note 11*).
6. Prepared culture flasks are incubated at 37 °C while shaking at 230 rpm. Optical density at 600 nm (OD600) should reach 0.15–0.2 within 2–3 h. OD600 should be measured every 30–45 min at the start of the culture and at shorter intervals when near OD600 of 0.15–2.
7. When OD600 of 0.15–0.2 is reached, 500 µL of IPTG stock solution are added to the culture flask to induce the protein production. Following induction, incubation at 37 °C and shaking are continued for another 5 h.
8. Cells are harvested by centrifugation for 20 min at 4 °C and 5000 g in 0.5 L centrifuge flasks. Cells are then washed and resuspended with 10 mL *Ftn*^(pos) size exclusion buffer (*see Note 12*).
9. Resuspended cells are then transferred into 50 mL plastic centrifugation tubes and centrifuged for another 20 min at 4 °C and 14,000 g.
10. Supernatant is discarded and cell pellets can either be processed directly or stored at –20 °C for prolonged periods of time.

3.2.2 Protein Production of *Ftn*^(neg)

1. LB agar plates with production bacteria transformed with the correct vector and gene are prepared identically to the first step of Subheading 3.2.1.

2. Precultures are prepared by touching a single colony of previously transformed production strain bacteria with a sterile disposable pipette tip and subsequent ejection of that tip into a 15 mL plastic centrifuge tube containing 10 mL LB medium supplemented with 15 μ L of 100 mg/mL ampicillin stock solution.
3. Precultures are incubated overnight (after being prepared in the afternoon ~16 h) at 37 °C while shaking at 230 rpm.
4. Before starting the production, culture incubators are pre-heated to 37 °C.
5. 200 mL of TB medium are filled into sterile 1 L conical cell culture flasks. The media is supplemented with 300 μ L of 100 mg/mL ampicillin stock solution. The production culture is started by adding 2 mL (1% v/v) of previously prepared preculture to the culture flask (*see Note 11*).
6. Prepared culture flasks are incubated at 37 °C while shaking at 230 rpm. Optical density at 600 nm (OD600) should reach 0.6–0.7 within 2–3 h. OD600 should be measured every 30–45 min at the start of the culture and at shorter intervals when near OD600 of 0.6–0.7.
7. When OD600 of 0.6–0.7 is reached 250 μ L of IPTG stock solution are added to the culture flask. Following induction, the temperature is reduced to 18 °C and incubation is continued for 48 h.
8. Cells are harvested by centrifugation for 20 min at 4 °C and 5000 g in 0.5 L centrifuge flasks. Cells are then washed and resuspended with 10 mL Ftn^(neg) size exclusion buffer (*see Note 12*).
9. Resuspended cells are then transferred into 50 mL plastic centrifugation tubes and centrifuged for another 20 min at 4 °C and 14,000 g.
10. Supernatant is discarded and cell pellets can either be processed directly or stored at –20 °C for prolonged periods of time.

3.2.3 Protein Purification

Fast Protein Liquid Chromatography (FPLC) is performed with an ÄKTA Go chromatography system (GE Healthcare). All buffer solutions are degassed under vacuum and strong stirring for at least 20 min before use during FPLC. UV absorption (280 nm) as well as conductivity sensors are required to monitor the purification. A sample loading pump is helpful, but not strictly necessary. Flowrate is set to 5 mL/min during ion exchange chromatography and 1 mL/min during size exclusion chromatography unless otherwise specified. Inlets are washed with their respective buffers before starting purification.

Purification of Ftn^(pos)

1. Cell pellets are resuspended in 10 mL Ftn^(pos) size exclusion buffer (*see Note 13*).
2. Lysis is performed by sonication in an ice bath to prevent heat-induced denaturation with a Vibra-Cell VCX-130 ultrasonic processor (Sonics). Amplitude is set to 60% and it is cycled four times: 1 min on -> 1 min off. On/off cycling is done to allow sufficient cooling of the solution to avoid denaturation (*see Note 14*).
3. The resulting suspension is centrifuged for 20 min at 4 °C and 14,000 g. Pelleted cell debris is discarded, while the supernatant is decanted into a fresh centrifuge tube.
4. RNase A is added to the supernatant and incubated for 3 h at 37 °C to allow for digestion of RNA (*see Note 15*).
5. The solution is now immersed in a 65 °C water bath for 10 min to denature remaining *E. coli* proteins and subsequently centrifuged to remove these for 20 min at 14,000 g and 4 °C. Pelleted denatured proteins are discarded and the supernatant is decanted into a fresh centrifuge tube (*see Note 16*).
6. Saturated ammonium sulfate solution is added to the supernatant to a final concentration of 70%.
7. The solution should turn turbid when its properly mixed (gently invert the centrifuge tube to mix). Precipitated protein is pelleted by centrifugation for 20 min at 4 °C and 14,000 g. A white pellet should be discernible. The supernatant can be discarded.
8. The pellet is resuspended with 5 mL Ftn^(pos) size exclusion buffer. Pipetting the solution up and down on top of the pellet should be sufficient to dissolve the pellet completely. A second precipitation is performed by adding 11.7 mL of saturated ammonium sulfate solution to the supernatant (final concentration of 70%), followed by centrifugation for 20 min at 4 °C and 14,000 g. A white pellet should again be discernible. The supernatant can be discarded.
9. Resuspension is done with 50 mL Ftn^(pos) ion exchange buffer (low salt)
10. Sterile filtration of the protein solution is done with 0.22 µm pore size, PVDF membrane, sterile syringe filter units. It can be necessary to use multiple filter units as the raw solution quickly blocks the membrane.
11. The cation exchange column is mounted to the system and prepared by washing with water until conductivity as well as absorption at 280 nm reach a stable baseline. ~5 column volumes (25 mL) should suffice. This step is also performed when a used column was stored under 20% ethanol.

12. Washing steps are also performed with the $\text{Ftn}^{(\text{pos})}$ ion exchange buffer (low salt) followed by $\text{Ftn}^{(\text{pos})}$ ion exchange buffer (high salt), followed by equilibration with $\text{Ftn}^{(\text{pos})}$ ion exchange buffer (low salt). Each step should require ~25 mL.
13. $\text{Ftn}^{(\text{pos})}$ solution, sterile filtered (*see step 9*), is then either injected onto the column by syringe and injection valve or sample loading pump. Conductivity and absorption at 280 nm should be monitored during sample application as excess ammonium sulfate remaining from the previous step can inhibit $\text{Ftn}^{(\text{pos})}$ binding to the column.
14. Unbound impurities are washed out with 20 mL of 50 mM MES pH 6, 0.5 M NaCl buffer.
15. Gradient elution is then performed from low to high salt buffer (0.5 M NaCl to 1.5 M 50 mM MES pH 6) over 12 min. All peaks with significant UV absorption should be collected, although $\text{Ftn}^{(\text{pos})}$ should elute at a conductivity of ~65–75 mS/cm.
16. Washing with 1.5 M 50 mM MES pH 6 is continued for 15 mL after gradient completion. The column is then re-equilibrated with 0.5 M 50 mM MES pH 6 for the next run or washed with water followed by 20% ethanol solution for storage. As an example, the complete chromatogram for an IEC purification of $\text{Ftn}^{(\text{pos})}$ is shown in Fig. 2.

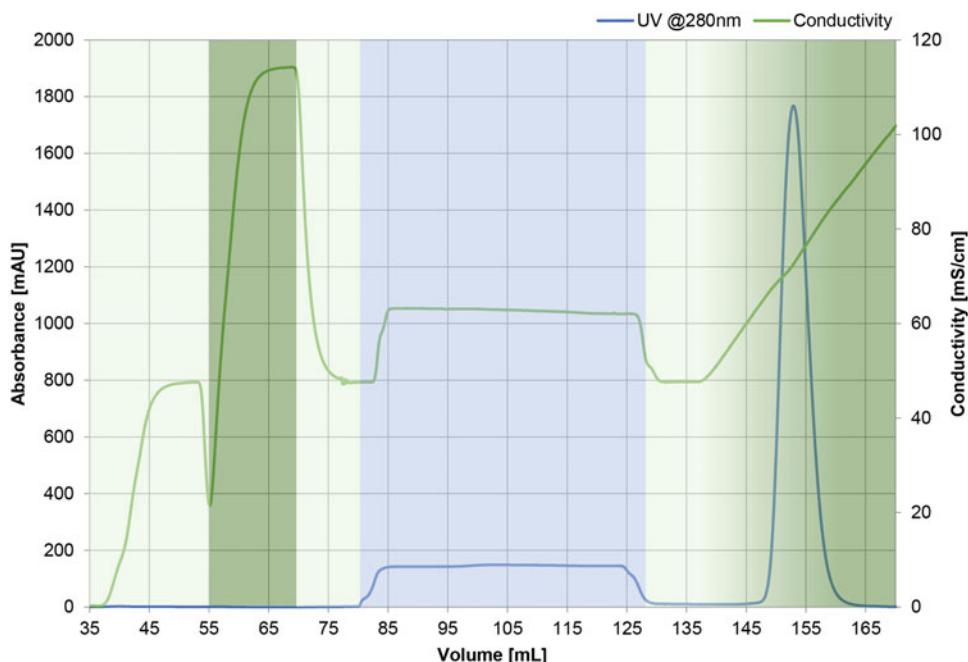


Fig. 2 Purification of $\text{Ftn}^{(\text{pos})}$ via ion exchange chromatography. Washing steps (at ~35–80 mL elution volume) and sample application (at ~80–130 mL) are shown before gradient elution (at ~135–170 mL)

Purification of Ftn^(neg)

1. Cell pellets are resuspended in 10 mL Ftn^(neg) size exclusion buffer.
2. Lysis is performed by sonication in an ice bath to prevent heat induced denaturation with a Vibra-Cell VCX-130 ultrasonic processor (Sonics). Amplitude is set to 60% and it is cycled eight times: 1 min on ->1 min off. (The cycle amount is increased since the amount of bacteria is larger when compared to the Ftn^(pos) protocol.)
3. **Step 3** is identical to Ftn^(pos), while **step 4** can be omitted completely. All subsequent steps, starting at Ftn^(pos) **step 5**, are performed equivalently with the respective buffers for Ftn^(neg).

Size Exclusion Chromatography

1. Fractions from IEC are concentrated to 2 mL using AmiconUltra-15 centrifugal filter units.
2. Mount the size exclusion column to the system and wash with 1.5 column volumes of water.
3. The system flow path before the column is washed with respective size exclusion buffer (SEC buffer) before starting purification.
4. 5 mL sample loops are washed first with 20 mL of water followed by 20 mL of respective SEC buffer.
5. Sample is loaded into the sample loop, followed by injection of 1 mL of respective SEC buffer to push all sample out of connecting tubes between injection valve and loop.
6. Equilibration is performed with 1.2 column volumes of respective SEC buffer.
7. Sample is then injected onto the column and run for 1.2 column volumes. The first SEC of each sample routinely shows two distinct peaks at 50 and 55 mL corresponding to 24-meric correctly assembled protein cages (later elution) and aggregated subunits (earlier elution) (*see Fig. 3a*). Second size exclusion chromatograms (only injecting fractions from the second peak of the first SEC) usually display a substantially reduced aggregate peak (*see Fig. 3b*).
8. Fractions containing the second peak are collected and combined and stored at 4 °C.

3.3 Methods for Synthesis of Metal Oxide Nanoparticles Inside the Ferritin Cavity

The ferritin nanocage has a cavity that can be loaded with different types of cargo, for example small molecules such as drugs, catalysts or nanoparticles. As an example, in the next part of this protocol, the synthesis of metal oxide nanoparticles inside the protein cavity is shown.

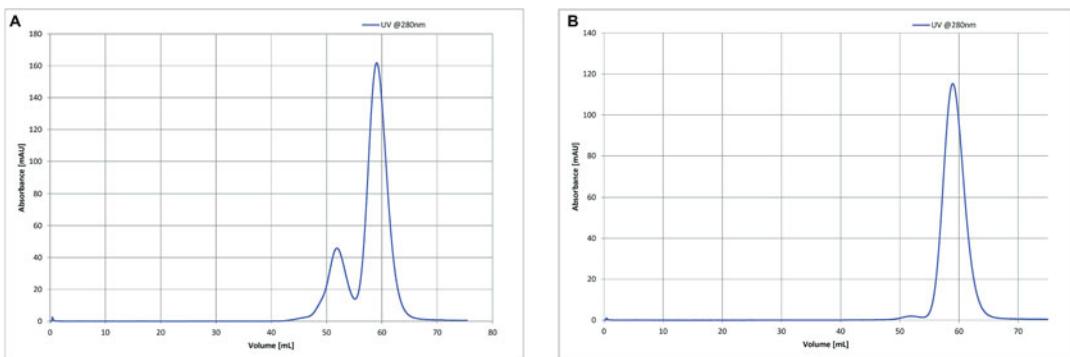


Fig. 3 Purification of $\text{Ftn}^{(\text{pos})}$ via size exclusion chromatography. **(a)** First SEC run showing aggregates at ~50 mL and singular cages at ~60 mL. **(b)** Second SEC run using only fractions containing the peak around ~60 mL of the first run

3.3.1 Iron Oxide Nanoparticle Synthesis

1. Preheat protein buffer in a 65 °C oil bath for 20 min in oxygen-free atmosphere in a two-neck round-bottom flask (25 mL) with constant stirring.
2. 4 mg of protein in buffer is added to a total volume of 10 mL.
3. After 10 min, H_2O_2 (6 mM) and $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ solution (18 mM) are added in equal volumes using KDS 100 Legacy syringe pumps (KD Scientific) (*see Note 17*).
4. Up to 5000 Fe(II) ions are injected per ferritin cage at a flow rate of 100 ions per minute per ferritin cage (*see Note 18*).
5. After complete addition, the solution is kept stirring for another 15 min.
6. 400 μL sodium citrate (300 mM stock solution) are added to chelate residual iron ions.
7. After additional 15 min at 65 °C, the solution was centrifuged for 10 min at 14,000 g and 4 °C.
8. The supernatant was rebuffered with protein buffer five times with a centrifugal filter unit to remove excess reagents.

3.3.2 Cobalt Oxide Nanoparticle Synthesis

Synthesis of cobalt oxide nanoparticles is similar to iron oxide nanoparticle synthesis with some modifications. Instead of Tris buffer, a HEPES buffer is used.

1. Here, H_2O_2 is added to reach a concentration of 10 mM and $(\text{NH}_4)_2\text{Co}(\text{SO}_4)_2$ solution is added until a concentration of 30 mM is reached.
2. Up to 2250 Co(II) ions are injected per ferritin cage at a flow rate of 45 ions per minute per ferritin cage (*see Notes 17 and 18*).

3. 1.5 mL sodium citrate (300 mM) are added at the end of reaction.
4. Afterwards, centrifuge the sample several times for 10 min at 20,000 g and 4 °C to remove small aggregates.

3.3.3 Cerium Oxide Nanoparticle Synthesis

Synthesis of cerium oxide nanoparticles is done in the same buffer as for the synthesis of iron oxide nanoparticles mentioned above. This synthesis is carried out with some modifications noted here.

1. Here, H₂O₂ is added to reach a concentration of 15 mM and CeCl₃ solution is added until a concentration of 30 mM is reached.
2. Up to 2250 Co(II) ions are injected per ferritin cage at a flow rate of 37 ions per minute per ferritin cage (*see Notes 17 and 18*).
3. Here, 1.2 mL EDTA (500 mM stock solution) are added at the end of the reaction.
4. Centrifuge the sample several times for 10 min at 20,000 g and 4 °C to remove small aggregates.

3.3.4 Gel Filtration for Nanoparticle Loaded Protein Cages

1. Nanoparticle-loaded protein cages are purified via gel filtration. Prior to that, the sample is centrifuged to remove aggregates and other larger contaminations.
2. The loading efficiency can be monitored by comparing the characteristic nanoparticle absorption with the protein absorption at 280 nm in the monomeric protein cage fractions. The size exclusion chromatogram of Ftn^(pos) loaded with cerium oxide nanoparticles is shown in Fig. 4 as an example.
 - (a) Iron loading efficiency is monitored by 350 nm absorption.
 - (b) Cobalt loading efficiency is monitored by 400 nm absorption.
 - (c) Cerium loading efficiency is monitored by 322 nm absorption.

3.3.5 Sucrose Gradient Centrifugation of Nanoparticle-Loaded Samples

1. To separate empty or sparsely loaded cages from nanoparticle-loaded cages, sucrose gradient centrifugation is performed.
2. 2 mL of loaded samples are layered on top of a stepwise sucrose gradient (20%, 40%, 70%) in 20 mL ultracentrifugation tubes.
3. Nanoparticle-loaded samples are centrifuged for 3 h at 193,000 g and 4 °C using an Optima XPN-80 ultracentrifuge (Beckman Coulter).
4. 1 mL fractions are extracted using a pipette, starting from the top. This procedure ensures to recover the nanoparticle-loaded cages (*see Note 19*).

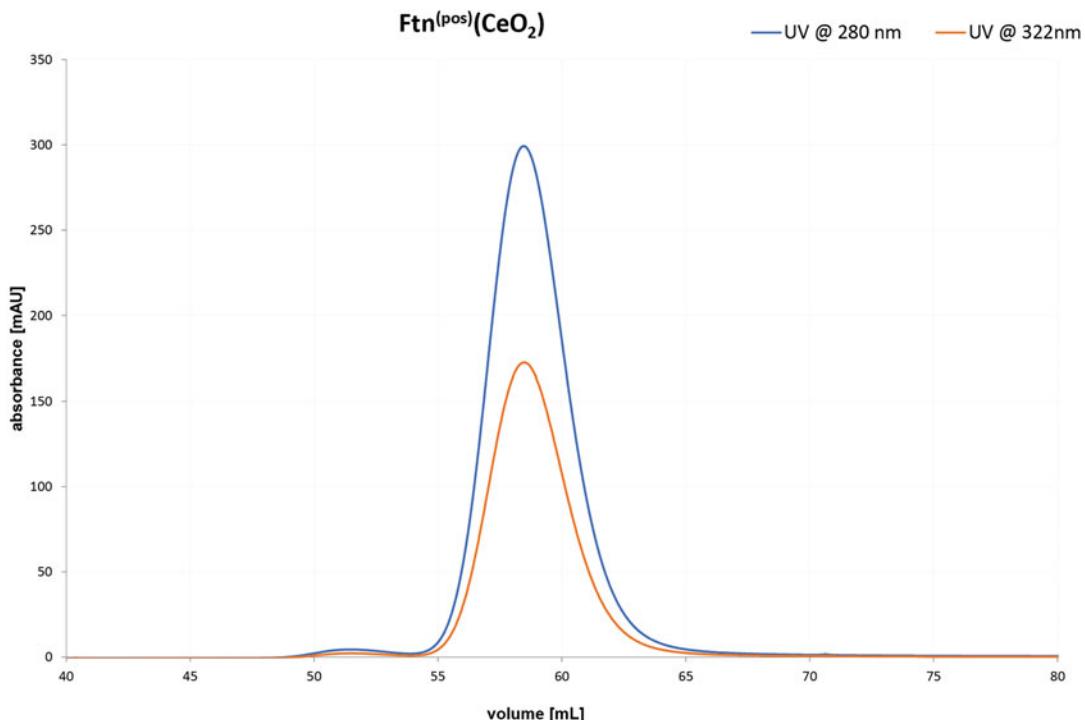


Fig. 4 Cerium oxide loaded Ftn^(pos) sample purified via size-exclusion chromatography. Absorption is shown at 280 nm (protein) and 322 nm (cerium oxide nanoparticles)

5. For each fraction, the UV/Vis absorption spectrum is measured with a NanoDrop 2000 spectrophotometer (ThermoScientific) (*see Note 20*).
6. Nanoparticle loading is monitored by comparing, e.g., the 350 nm absorption for iron oxide with the 280 nm absorption (protein).
7. Nanoparticle-containing samples are washed with protein buffer using centrifugal filter units to remove sucrose.
8. Protein concentration is determined by Bradford method [18].

3.4 Methods for Protein Crystallization

To assemble the protein cages in a highly ordered superlattice, protein crystallization techniques are used. These protocols and conditions can be used to assemble both empty or loaded ferritin cages. This fact indicates that the protein template governs the structure independent from the type of cargo, and the protein cages can be used to assemble highly ordered nanoparticle superlattices. Moreover, the crystallization setup used in protein crystallography elegantly enables control over interaction strength between building block because, for example, salt concentration and pH value can be rationally controlled.

1. Ftn^(neg) and Ftn^(pos) samples, which have been purified by two rounds of SEC, are concentrated to 4, 6 and 8 mg/mL (*see Notes 21 and 22*).
2. Protein stocks are centrifuged for 10 min at 20,000 g to remove any particulates that might remain in the samples.
3. Reservoir solutions for crystallization are prepared according to one of the following protocols, each corresponding to a different crystal structure. Each reservoir solution can be either prepared by freshly mixing sterile-filtrated stock solutions or as a larger batch of stock solutions beforehand:

The following steps are identical for each crystallization condition.

4. Each well on a pregreased Crystalgen SuperClear crystallization plate (Jena BioScience) is filled with 500 µL of reservoir solution.
5. 2 µL of reservoir solution are pipetted onto a 22 mm diameter siliconized cover slip to form a droplet. 1 µL of Ftn^(pos) stock solution is added to the droplet, followed by 1 µL of Ftn^(neg) stock solution. Both protein stock solution should have the same concentration (*see Note 23*).
6. The cover slip is picked up with tweezers and placed upside down onto a well containing the respective reservoir solution. Prepared plates are stored in a 20 °C incubator.
7. Crystals with several 10–100 µm in size are formed after 1–3 days (*see Fig. 5*).

3.5 Methods for Small Angle X-Ray Scattering (SAXS) Measurement

3.5.1 Protein Crystal Fixation

1. Protein crystals are fixated in the hanging drop of the crystallization setup (*see Notes 24 and 25*).
2. The cover slide with the crystals is briefly removed to add 10 µL of 25% glutaraldehyde to the reservoir solution [19] and followed by mixing the reservoir solution with a pipette tip (final glutaraldehyde concentration in reservoir solution: 0.5%)
3. Afterwards, the cover slide with the crystallization drop is put back on top of the crystallization well.
4. Incubation up to 48 h (glutaraldehyde diffuses slowly into the crystallization drop).

To determine the superstructure of the crystalline samples, small angle X-ray measurements are carried out. The sample preparation, data collection and analysis are described in the following section.

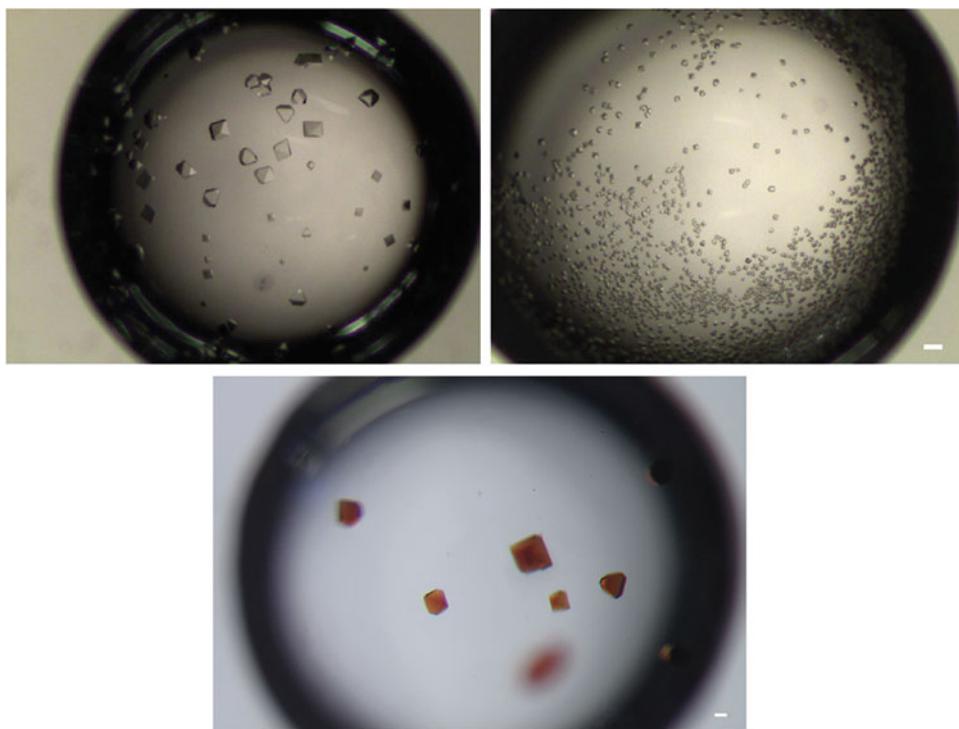


Fig. 5 Binary protein crystals from $\text{Ftn}^{(\text{pos})}$ and $\text{Ftn}^{(\text{neg})}$. Upper left: empty cages; Upper right: Cerium oxide loaded $\text{Ftn}^{(\text{pos})}$ with empty $\text{Ftn}^{(\text{neg})}$. Bottom: $\text{Ftn}^{(\text{pos})}$ empty with iron oxide filled $\text{Ftn}^{(\text{neg})}$ (Scalebar:100 μm)

3.5.2 SAXS Measurement

Data Collection

1. Fixated crystals are transferred from mother liquor to a borosilicate glass capillary (*see Notes 26 and 27*).
2. During data collection, the capillary is rotated around its vertical axis by an electrical motor (Speed: 12.6 rotations per minute).
3. SAXS data are recorded with X-ray microfocus source (Cu K_α , $\lambda = 1.54 \text{ \AA}$) and a suitable detector in a q range from 0.004 to 0.22 \AA^{-1} at room temperature (*see Notes 28 and 29*).
4. Here, the sample-to-detector distance is set to 2.59 m.
5. Standard sample: silver behenate is used for calibration of the length of the scattering vector q .

Data Analysis

1. SAXSGUI (Rigaku Innovative Technologies, Inc. JJ X-Ray Systems ApS) is used to analyze collected data.
2. The 2D scattering data are azimuthally averaged to generate one-dimensional SAXS data (*see Fig. 6*).
3. The magnitude of the scattering vector q is given by $q = 4\pi \sin \theta / \pi$, where 2θ is the scattering angle.

The unit cell parameters are determined using the scattering vector q for reflections (001) and (100).

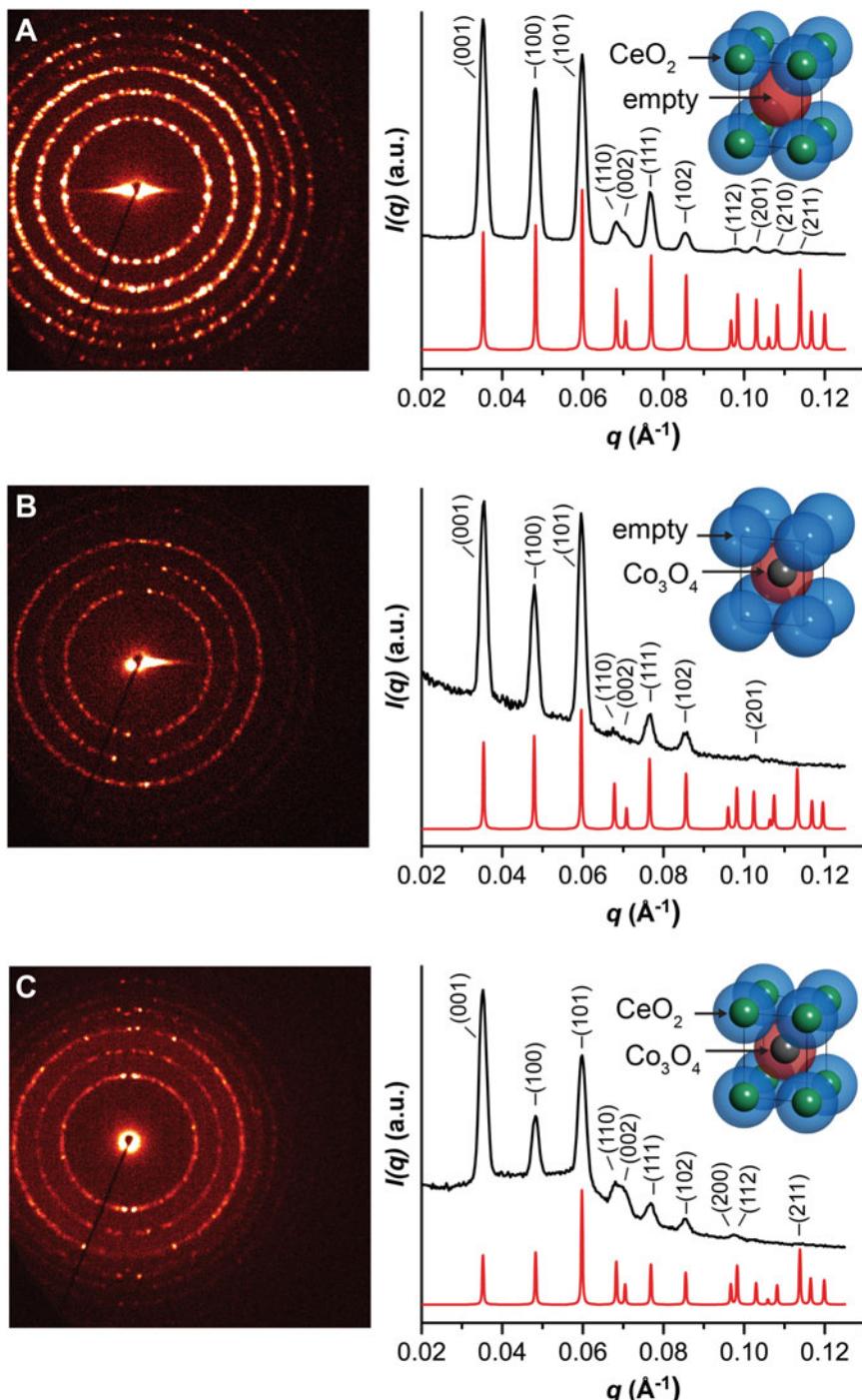


Fig. 6 2D and radially averaged 1D SAXS data for binary nanoparticle crystals. Binary crystals composed of (a) Ftn^(pos) CeO₂ and Ftn^(neg) empty. (b) Ftn^(pos) empty and Ftn^(neg) Co₃O₄. (c) Ftn^(pos) CeO₂ and Ftn^(neg) Co₃O₄. Experimental data are shown in black, and simulated diffraction patterns in red. A unit cell of the corresponding crystal composition is shown. (Reprinted with permission from M. Künzle, T. Eckert, T. Beck, *J. Am. Chem. Soc.* **2016**, *138*, 12731–12734. Copyright 2016, American Chemical Society)

- Combination of scattering vector and quadratic Bragg equation for tetragonal crystal system result in:

$$q_{hkl} = \frac{2\pi}{a} \sqrt{h^2 + k^2 + \left(\frac{a}{c}\right)^2 l^2} \quad (1)$$

- Cell parameters a and c can be determined using:

$$a = \frac{2\pi}{q_{100}} \quad (2)$$

$$c = \frac{2\pi}{q_{001}} \quad (3)$$

- The experimental q value can directly be compared for reflection (101) with the calculated value using the derived cell parameters:

$$q_{101} = \frac{2\pi}{a} \sqrt{1 + \left(\frac{a}{c}\right)^2} \quad (4)$$

4 Notes

- For the compiling of the software, follow the instructions on the following website https://www.rosettacommons.org/demos/latest/tutorials/install_build/install_build
- The website www.rosettacommons.org is a great resource for learning new Rosetta protocols or trouble shooting.
- If you want to calculate multiple poses in parallel, you need to set the extras=mpi option directly during the compiling of the software suite. More information at https://new.rosettacommons.org/docs/latest/rosetta_basics/MPI
- An explanation for the used options, and also more options and in-depth information on symmetry files can be found at https://www.rosettacommons.org/manuals/archive/rosetta3.1_user_guide/symmetry.html
- These are just a small collection of all possible options. A comprehensive list can be found at https://new.rosettacommons.org/docs/latest/application_documentation/structure_prediction/relax
- More in-depth information about the packer and rotamer packing options can be found at https://new.rosettacommons.org/docs/latest/rosetta_basics/options/packing-options and https://www.rosettacommons.org/demos/latest/tutorials/Optimizing_Sidechains_The_Packer/Optimizing_Sidechains_The_Packer

7. If you are using a different operating system and compiled Rosetta accordingly, you should find the protocol in the same place but for example ending on macgccrelease.
8. For a list of all commands check “Command for controlling sequence identity” on https://new.rosettacommons.org/docs/latest/rosetta_basics/file_types/resfiles
9. For the positive supercharging of ferritin, the “METHOD_WEIGHTS ref” values for arginine were –0.3, –0.4, –0.5, –0.6, –0.7, –0.8, –0.9, –1.0, –1.1, –1.2, –1.3, –1.4, –1.5 and for lysine –0.15, –0.27, –0.33, –0.40, –0.46, –0.53, –0.60, –0.66, –0.73, –0.80, –0.86, –0.93, –0.99.
10. For the negative supercharging of ferritin, the “METHOD_WEIGHTS ref” values for glutamic acid were –0.3, –0.4, –0.5, –0.6, –0.7, –0.8, –0.9, –1.0, –1.1, –1.2, –1.3, –1.4, –1.5 and for aspartic acid –0.25, –0.33, –0.41, –0.50, –0.58, –0.66, –0.74, –0.83, –0.91, –0.99, –1.08, –0.16, –1.24.
11. Make sure the culture flask is closed properly with the correct lid or aluminum foil to prevent any contamination.
12. Loosening the cells from the bottom of the flask can be facilitated by pipetting up/down multiple times or placing the whole culture flask on a bench-top vortexer on a low power setting.
13. A sterile spatula or large pipette tip can be used to dislodge and break up the pellet before sonication.
14. The suspension should contain no visible aggregates after sonication. If aggregates are still present, the solution is sonicated again until no aggregates remain.
15. Insufficient digestion can hinder binding of Ftn^(pos) on the cation exchange column due to charge neutralization.
16. The ferritin nanocage is thermostable and stays assembled in solution.
17. To add the correct volume, make sure that the canula of the syringe is completely filled with the solution by carefully pressing the first few drops out of it.
18. To avoid a reaction between the precursors, make sure that both tips of the syringes are immersed in the solution and as far separated from each other as possible.
19. Be cautious with the sample to avoid mixing the fractions with different densities after the centrifugation.
20. Due to the incorporation of the nanomaterial, the ratio of the absorbance between 260 and 280 nm shown in the NanoDrop should shift to higher values than the native protein.

21. Amicon Ultra-0.5 Ultracel-30 Membrane, 30 kDa for small volumes or Amicon Ultra-15 centrifugal filter units, 30 kDa Molecular weight cut-off (Merck Millipore) for larger volumes.
22. Each centrifugal filter unit is used for one protein variant exclusively to avoid cross contamination
23. Protein stocks are added to the initial droplet by pushing the pipette to the first stop, forming a droplet at the pipette tip, which is then gently deposited on the already present drop on the cover slip, merging the two together.
24. Glutaraldehyde diffuses slowly into the crystallization drop.
25. This setup ensures a gentle crosslinking process.
26. The protein crystals should not be too small, at least 50 µm upwards
27. Fixated crystals must be washed at least five times with water to remove residual cross-linking agent.
28. The reflection intensities of the diffraction rings can be reliably quantified and put in relation to each other.
29. Rotation ensures that ensemble diffraction from the sample is recorded.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) with project ID 401323995 and by the Cluster of Excellence “CUI: Advanced Imaging of Matter” of the Deutsche Forschungsgemeinschaft (DFG)-EXC 2056-project ID 390715994.

References

1. Ha M, Kim JH, You M, Li Q, Fan C, Nam JM (2019) Multicomponent plasmonic nanoparticles: from heterostructured nanoparticles to colloidal composite nanostructures. *Chem Rev* 119:12208–12278
2. Silvi S, Credi A (2015) Luminescent sensors based on quantum dot-molecule conjugates. *Chem Soc Rev* 44:4275–4289
3. Hildebrandt N, Spillmann CM, Algar WR, Pons T, Stewart MH, Oh E, Susumu K, Diaz SA, Delechanty JB, Medintz IL (2017) Energy transfer with semiconductor quantum dot bioconjugates: a versatile platform for biosensing, energy harvesting, and other developing applications. *Chem Rev* 117:536–711
4. Rainò G, Becker MA, Bodnarchuk MI, Mahrt RF, Kovalenko MV, Stöferle T (2018) Superfluorescence from lead halide perovskite quantum dot superlattices. *Nature* 563:671–675
5. Li J, Wang Y, Zhou T, Zhang H, Sun X, Tang J, Zhang L, Al-Enizi AM, Yang Z, Zheng G (2015) Nanoparticle superlattices as efficient bifunctional electrocatalysts for water splitting. *J Am Chem Soc* 137:14305–14312
6. Boles MA, Engel M, Talapin DV (2016) Self-assembly of colloidal nanocrystals: from intricate structures to functional materials. *Chem Rev* 116:11220–11289
7. Künzle M, Eckert T, Beck T (2016) Binary protein crystals for the assembly of inorganic nanoparticle superlattices. *J Am Chem Soc* 138:12731–12734

8. Lach M, Künzle M, Beck T (2017) Free-standing metal oxide nanoparticle superlattices constructed with engineered protein containers show in crystallo catalytic activity. *Chem Eur J* 23:17482–17486
9. Künzle M, Mangler J, Lach M, Beck T (2018) Peptide-directed encapsulation of inorganic nanoparticles into protein containers. *Nanoscale* 10:22917–22926
10. Künzle M, Lach M, Beck T (2018) Crystalline protein scaffolds as a defined environment for the synthesis of bioinorganic materials. *Dalton Trans* 47:10382–10387
11. Künzle M, Lach M, Beck T (2019) Multi-component self-assembly of proteins and inorganic particles: from discrete structures to biomimetic materials. *Isr J Chem* 59:906–912
12. Künzle M, Lach M, Budiarta M, Beck T (2019) Higher-order structures based on molecular interactions for the formation of natural and artificial biomaterials. *ChemBioChem* 20: 1637–1641
13. Lach M, Künzle M, Beck T (2019) Proteins as sustainable building blocks for the next generation of bioinorganic nanomaterials. *Biochemistry* 58:140–141
14. Miklos AE, Kluwe C, Der BS, Pai S, Sircar A, Hughes RA, Berrondo M, Xu J, Codrea V, Buckley PE, Calm AM, Welsh HS, Warner CR, Zacharko MA, Carney JP, Gray JJ, Georgiou G, Kuhlman B, Ellington AD (2012) Structure-based design of supercharged, highly thermostable antibodies. *Chem Biol* 19:449–455
15. Beck T, Tetter S, Künzle M, Hilvert D (2015) Construction of matryoshka-type structures from supercharged protein nanocages. *Angew Chem Int Ed* 54:937–940
16. <https://www.chem-agilent.com/pdf/strata/200518.pdf>
17. Nelson MD, Fitch DH (2011) Overlap extension PCR: an efficient method for transgene construction. *Methods Mol Biol* 772: 459–470
18. Bradford MM (1976) A rapid and sensitive method for the quantitation microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 254:248–254
19. Salem M, Mauguin Y, Prangé T (2010) Revisiting glutaraldehyde cross-linking: the case of the Arg-Lys intermolecular doublet. *Acta Crystallogr Sect F* 66:225–228



Chapter 22

Production and Purification of Virus-Like Particles by Transient Expression in Plants

Lygie Esquirol, Donna McNeale, Micol Venturi, and Frank Sainsbury

Abstract

Transient expression in plants has become a useful production system for virus-like particle (VLP) expression. High yields and flexible approaches to assembling complex VLPs, combine with ease of scale-up and inexpensive reagents to provide an attractive method for recombinant protein expression in general. Plants have demonstrated excellent capacity for the assembly and production of protein cages for use in vaccine design and nanotechnology. Furthermore, numerous virus structures have now been determined using plant-expressed VLPs, showing the utility of this approach in structural virology. Transient protein expression in plants uses common microbiology techniques, leading to a straightforward transformation procedure that does not result in stable transgenesis. In this chapter, we aim to provide a generic protocol for transient expression of VLPs in *Nicotiana benthamiana* using soil-free plant cultivation and a simple vacuum infiltration procedure, along with methodology for purifying VLPs from plant leaves.

Key words Virus-like particle, Transient expression, Plant-based expression, Density gradient, Vacuum infiltration

1 Introduction

Plant-based transient expression has emerged as a production system for virus-like particles (VLPs) that is flexible, scalable and commercially viable [1, 2]. The technique is based on the transcriptional competency of Agrobacteria-derived transfer DNA that allows for the simultaneous transformation of the somatic tissues of mature plants. During the process known as “agroinfiltration,” Agrobacteria is forced under pressure into the interstitial spaces of fully grown leaves where it has access to every leaf cell. This results in expression that is typically 100–10,000 times higher than is possible by stable transformation. Using state-of-the-art expression vectors [3], some of which are freely available [4], transient expression has seen most success in the plant *Nicotiana benthamiana*. In combination with transient expression, *N. benthamiana* has become a commercial protein production host [1] as well as a

laboratory model for plant pathology [5], metabolic engineering [6], and synthetic biology [7]. This recombinant protein expression approach has also proven to be particularly useful for the expression of VLPs [2, 8].

VLPs are examples of natural protein cages derived from virus capsids that are engineered, largely via heterologous expression, to be free of replicating nucleic acid. Their expression in plants has attracted interest for the development of vaccines [9–12] and as tools in nanotechnology [13]. As protein cages in nanotechnology, plant-made VLPs have been used to present proteins [14], encapsulate proteins [15, 16] or inorganic materials [17], and to deliver cargos to mammalian cells [15]. As a reverse genetics tool in fundamental virology, transient expression of VLPs in plants enables decoupling of the effects of coat protein variants from genome encapsidation [18, 19] and from assembly [19, 20]. As viable virus is not required, deeper insights into the molecular mechanisms of assembly processes can be obtained. In addition, the structural fidelity of VLPs and VLP assembly in plants during transient expression has enabled capsid structure determination of plant viruses [18–21], mammalian viruses [11, 15], fish viruses [12], and insect viruses [22].

As a technology for the expression of a wide variety of VLPs as protein cages, transient expression in plants is quite mature (*see* Fig. 1). The technique is straightforward and does not require expensive consumables or reagents to maintain growth of the expression host. The ease with which transient expression facilitates co-expression of proteins, simply by mixing agrobacterial cultures for infiltration, has led to the expression of more complex VLPs and innovative expression strategies. For example, Bluetongue virus VLPs required modulation of expression levels among the four co-expressed structural proteins to maximize yield [23]. Optimal expression and assembly of Cowpea mosaic virus VLPs was achieved by fusing the two capsid proteins as a polyprotein and co-expressing with the cognate protease, leading to better yields than co-expression of individual capsid proteins [24]. Seeing the potential for plant-based transient expression of VLPs, in 2015 Peyret identified a need for generic protocols for their purification at laboratory scale [25]. Although not suited to industrial processes due to the use of ultracentrifugation, such protocols help advance the VLP and protein cage field in basic virology and nanotechnology.

This chapter aims to provide a generalizable approach enabling the laboratory-scale transient expression and purification of VLPs from plants. Protocols for cloning and engineering constructs for transient protein expression in plants can be found elsewhere [26, 27]. The current protocol starts with growth conditions for soil-free cultivation of *N. benthamiana* using a small-footprint growth cabinet, which ensures better consistency in plant growth

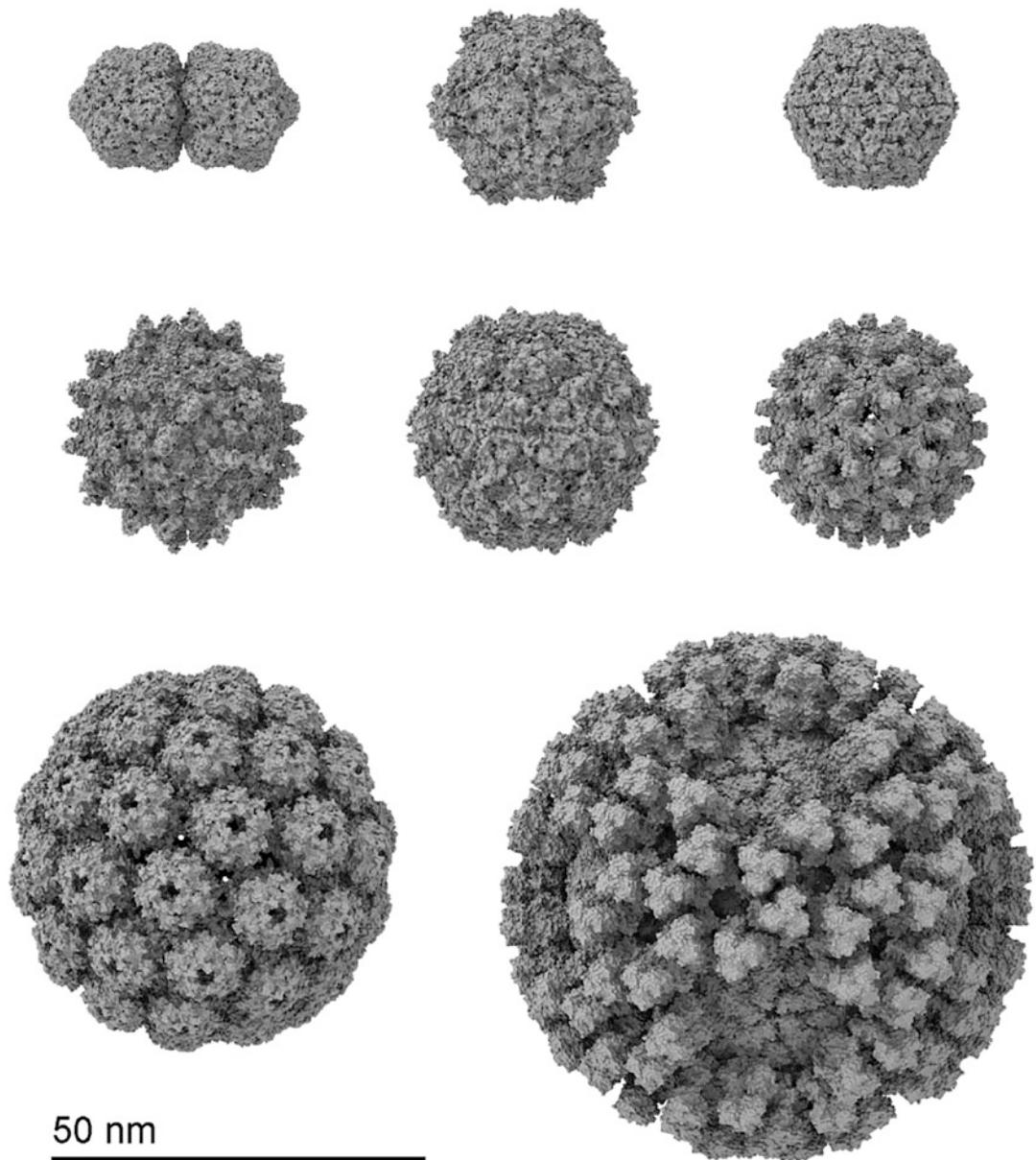


Fig. 1 Diversity of VLPs produced in plants. Examples of VLP structures made by transient expression in plants. Top from left to right: Ageratum yellow vein virus [19] (PDB 6F2S), Cowpea mosaic virus [18] (PDB 5A33), poliovirus [11]. Middle from left to right: Pepper cryptic virus 1 [21] (PDB 7NCR), barley yellow dwarf virus [20] (PDB 6SCL), Hepatitis B core [14]. Bottom from left to right: Murine polyomavirus (MPyV), Bluetongue virus core. PDB code indicates new atomic coordinates obtained from plant-made VLPs

and avoids specialized decontamination procedures of soil waste. We then describe a process for laboratory scale vacuum infiltration that uses cost effective off-the-shelf parts. The purification of VLPs from plant tissue involves tissue disruption, clarification, and isolation of VLPs via ultracentrifugation through a density medium.

Further purification is often needed and, while this has traditionally involved subsequent ultracentrifugation steps, we describe a size-exclusion chromatography (SEC) process that provides buffer exchange and lends itself to more selective recovery of VLPs as well as initial analytical assessment by coupled dynamic light scattering (DLS). We also describe the minimal analytical requirements for confirming the presence of VLPs, polyacrylamide gel electrophoresis (PAGE) and transmission electron microscopy (TEM).

2 Materials

2.1 Lab Equipment

1. Growth chamber.
2. Vacuum Pump (≥ 20 L/min to ≤ 0.09 MPa).
3. Desiccator with pressure gauge (≥ 20 cm internal height).
4. 1 L plastic beakers.
5. Waring Blender.
6. Low speed centrifuge.
7. Ultracentrifuge (e.g., ThermoFisher Sorvall WX+ Ultracentrifuge Series).
8. Fast protein liquid chromatography (FPLC) system (e.g., Cytiva AKTA or Bio-Rad NGC).
9. Dynamic light scattering instrument.
10. SDS-PAGE gel electrophoresis tank and power pack.
11. Gel documentation system.
12. Transmission electron microscope.

2.2 Reagents and Consumables

1. Peat pellets.
2. Rockwool cubes (Grodan Rockwool Delta 4G 42/40).
3. Liquid fertilizer (Example approximate nutrient levels: Nitrogen, 229 ppm; Potassium, 307 ppm; Phosphorus, 83 ppm).
4. Infiltration buffer 10 mM 3-(N-morpholino)propanesulfonic acid (MES) at pH 5.6, 1 mM MgCl₂, 100 µm acetosyringone (*see Note 1*).
5. Miraclot (Merck-Millipore).
6. Polyclear thin-wall ultracentrifugation tubes (e.g., Beckman Coulter ultra-clear centrifuge tubes).
7. 3 and 5 mL syringes, 18G × 38 mm and 21G × 32 mm needles.
8. Iodixanol (OptiPrep; Merck-Sigma).
9. PD MidiTrap G-25 columns (Cytiva).
10. Bradford reagent (BioRad).

11. Bovine serum albumin (BSA).
12. SDS-PAGE sample buffer: 250 mM Tris–HCl (pH 6.8), 10% w/v SDS, 0.25% w/v bromophenol blue, 50% v/v glycerol, and 0.5 M DTT yields a 5× buffer.
13. Protein gels.
14. SDS-PAGE running buffer (250 mM Tris–HCl, 190 mM glycine, 1% w/v SDS).
15. Gel code BlueSafe protein stain (Thermofisher).
16. HiPrep 16/60 Sephadex S-500 HR SEC column (Cytiva).
17. Electron microscopy grids (copper, 200 mesh, carbon-coated; ProSciTech).
18. Uranylless EM stain (ProSciTech).

3 Methods

3.1 Plant Growth

Access to a glasshouse facility may be limiting for many labs that would otherwise be interested in transient plant-based expression in plants as an alternative eukaryotic protein production host. The following section provides details on the cultivation of *N. benthamiana* in a growth chamber. The use of soil can lead to problems with consistency and sterility; as such, an alternative method based on peat pellets and rockwool cubes is provided here. We have found that a long-day light regime without humidity regulation (Table 1) results in faster and more regular growth of plants than an outdoor greenhouse. It should be noted that high humidity results in plants that are very difficult to infiltrate by syringe. However, this does not impact the effectiveness of vacuum infiltration as described below.

1. Immerse peat pellets in water until they swell, approximately 30 s in hot water.
2. Gently spread 20–30 seeds on the surface of one pre-wetted peat pellet with the top cut off. Place in a tray and cover with a lid to keep high humidity during germination.
3. After 7 days of incubation (*see* Fig. 2a), carefully move individual sprouts to new peat pellets using tweezers and cover with a lid (*see* Fig. 2b) for a further 14 days.
4. To prepare rockwool cubes, soak them in a liquid fertilizer solution.
5. Once the plants have four true leaves, insert the peat pellet into a pre-wetted rockwool cube (*see* Fig. 2c).

Table 1
Typical growth conditions for *N. benthamiana* in a programmable growth chamber

Schedule	Temp (°C)	Light intensity
12 am	23	0
5 am	24	1
6 am	25	3 (~450 µmole/m ² /s)
8 pm	24	1
9 pm	23	0

6. Water approximately once a week (depending on humidity) by immersing the cube in water, taking care not to submerge the plant, then gently squeeze to remove excess water.
7. Plants are ready for infiltration 30–35 days from sowing (*see* Fig. 2d and Note 2).

3.2 Infiltration

The following protocol is for vacuum infiltration, which may be performed at a wide scale range from a single plant to as many plants that fit in the growth chamber, with appropriate scale-up of the Agrobacterial culture volume. The transformation is complete in a matter of minutes, with every leaf completely infiltrated. The suggested volumes below enable the preparation of the bacterial suspension and vacuum infiltration procedure for 1–2 L of suspension per culture, enough for 5–10 plants.

1. Agrobacterium should be grown from a single colony with antibiotics appropriate for the strain and binary vector [28] in 13 mL LB in a 50 mL falcon tube for 2 days at 28 °C, shaking at 220 rpm.
2. Pour 13 mL of the starter culture into 187 mL LB in a 1 L Erlenmeyer flask with the same antibiotics and grow overnight at 28 °C, shaking at 220 rpm.
3. Pellet the bacteria by centrifugation at 2222 g for 22 min at 22 °C.
4. Resuspend the pellet gently in a small volume using a Pasteur pipette with infiltration buffer (*see* Note 1). Once the pellet is completely dispersed, add infiltration buffer until a final OD₆₀₀ of 0.4 is reached. This will produce 1–2 L of Agrobacterial suspension (*see* Note 3).
5. Incubate at room temperature for 2–3 h to allow for the activation of Agrobacterium virulence genes.
6. Place the plants upside down into 1 L beakers containing the Agrobacterial suspension, such that all the leaves are

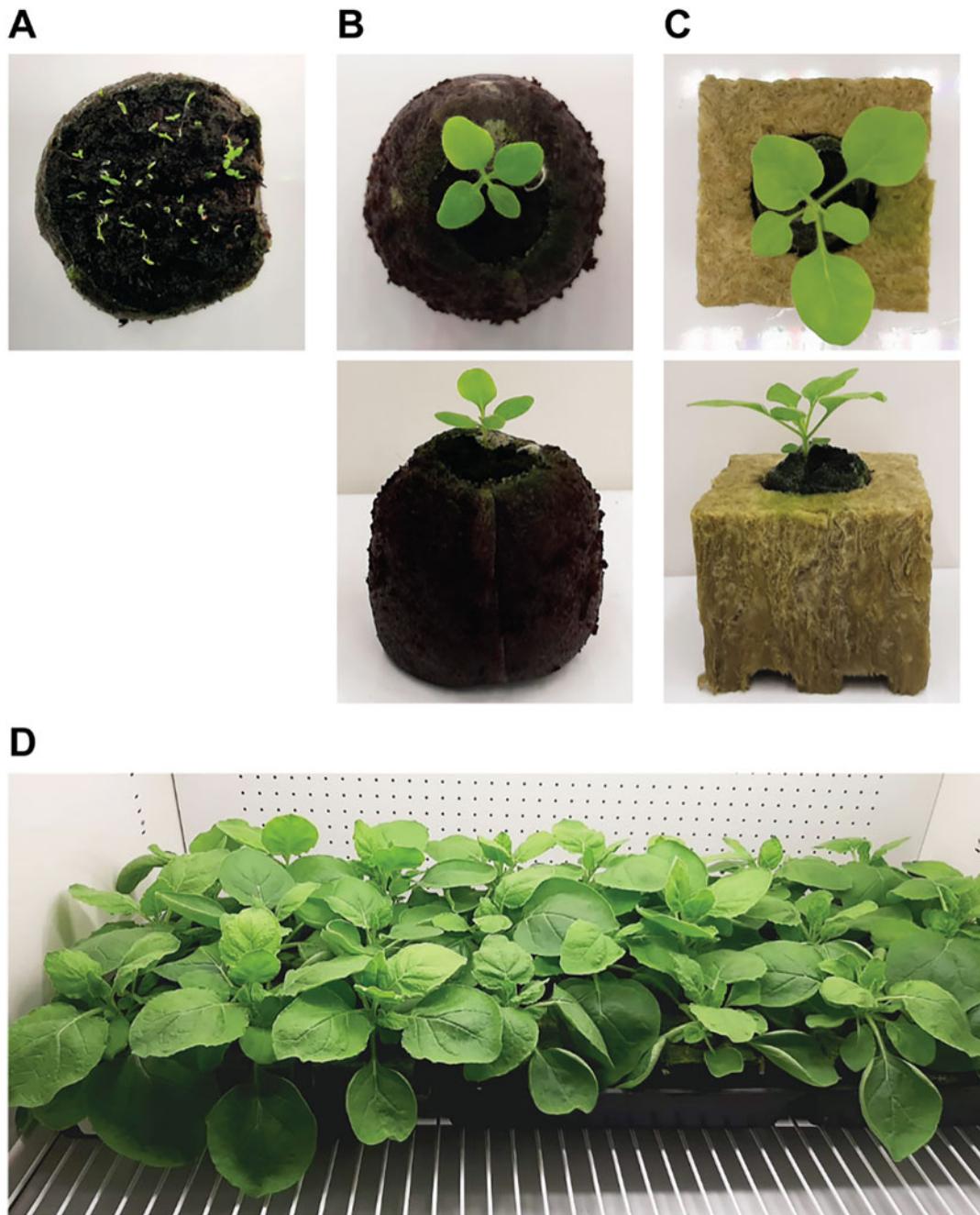


Fig. 2 Key stages of *N. benthamiana* cultivation in the growth chamber. (a) Germinating seedlings at 5 days following sowing onto a cut peat pellet (~35 mm diameter). (b) Seven to eight-day-old sprouts are transferred to individual peat pellets (image shows seedlings at 14 days-old). (c) On day 21, the peat pellets are placed into 75 mm rockwool cubes. (d) Plants are ready for infiltration up to 35 days old (image shows plants at 32 days old)

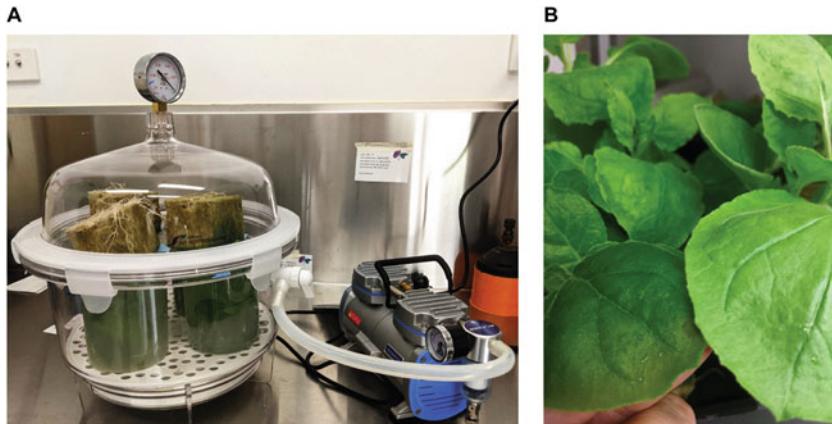


Fig. 3 Infiltration apparatus. (a) Desiccator and vacuum pump set up for the infiltration of four plants simultaneously with all leaves submerged inside 1 L plastic beakers. (b) Saturation of the leaf is visible in infiltrated plants (left) compared to a non-infiltrated plant (right)

submerged. Up to four 1 L beakers will fit inside a desiccator with an internal diameter of ~30 cm (*see Fig. 3*).

7. Connect and run a vacuum pump until the vacuum reaches 0.09 MPa (90 mbar) and air is bubbling out of the leaves.
8. Hold the pressure for 30 s before opening the valve to release the vacuum. The infiltration of the bacterial suspension into the leaves can be observed.
9. Remove the plants from the bacterial suspension and verify that the infiltration is complete for all leaves. After watering the rockwool cubes as above, put the plants back into the growth chamber.

3.3 Purification I: Harvest, Lysis and Clarification, Differential Centrifugation

This section will describe the collection and disruption of the infiltrated leaves, clarification of the lysate, and initial isolation of VLPs through a one-step ultracentrifugation protocol [29]. Separation of VLPs from the lysate makes use of a high-density media and simple rate-zonal centrifugation using a gradient. In some cases, a discrete band is formed, allowing the collection of purified VLPs [15, 16]. However, in other cases, a simple cushion of a single density can be more effective to enrich VLPs before purifying them with a second step, for example, a second ultracentrifugation step [25] or via SEC as described below. The density gradient can be constructed from various compounds. Here we have used iodixanol (marketed as Optiprep), which is osmotically neutral and can provide excellent separation.

3.3.1 Harvesting

1. Remove the infiltrated leaves 5–8 days post-infiltration, depending on the protein expression profile, which should be empirically determined for each recombinant protein (*see Note 4*).

2. Remove the petiole and mid vein with a scalpel before weighing leaves in batches of 20–30 g.
3. Leaves can be frozen for many months without loss of VLP yield in plastic zip-lock bags at –80 °C.

3.3.2 Tissue Disruption and Clarification

1. Prepare 50 mL of ice-cold extraction buffer suited to the VLP of interest (*see Note 5*), containing protease inhibitor cocktail, per 20–30 g of fresh or frozen leaves.
2. In a Waring blender or similar, add 50 mL to the leaf tissue and set to high speed for 2 min (*see Note 6*).
3. Filter the lysate through a double layer of Miracloth into 50 mL centrifuge tubes on ice to remove large debris.
4. Clarify the lysate by centrifugation at 20,000–25,000 g for 20 min at 4 °C.
5. Transfer the supernatant to fresh tubes.

3.3.3 Differential Centrifugation

1. Stepwise gradients of 50%, 40%, 30%, and 20% iodixanol in a buffer suited to the VLPs are prepared in polyclear ultracentrifuge tubes using 3 mL of each step carefully dispensed from an 18-gauge needle (*see Note 7*).
2. Overlay 23 mL of clarified cell lysate containing the extracted particles on top of the stepwise gradient (*see Fig. 4*).
3. Centrifuge samples at 130,000 g for 3 h at 4–16 °C.
4. Carefully remove tubes, being sure not to disturb and place in a suitable rack for sampling from the particle band (*see Fig. 4*).

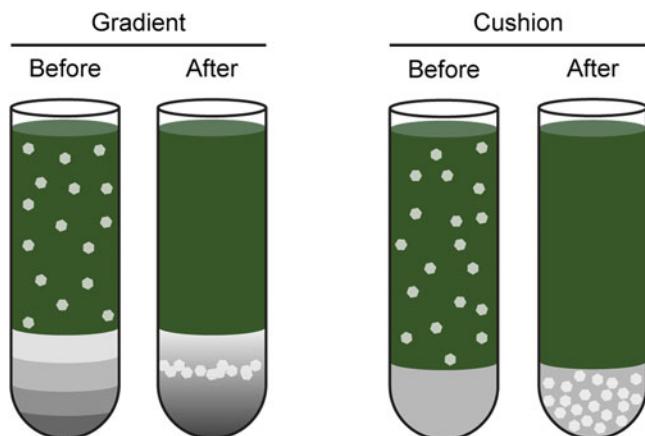


Fig. 4 Schematic diagram for differential ultracentrifugation. Gradients or cushions of a suitable density media may be prepared for the isolation or enrichment of VLPs from plant leaf lysates

5. Recover the VLPs by puncturing the side of the tube approximately 5 mm below the visible particle band using a 21-gauge needle (*see Note 8*).
6. The side puncture technique may also be used to recover enriched VLPs from a density cushion. However, due to the diffuse distribution of VLPs within a cushion, a second ultracentrifugation step is recommended, particularly in cases where yield is low.

3.3.4 Desalting

A desalting step before analysis or further purification helps to remove iodixanol effectively. The gravity flow PD MidiTrap G-25 columns are rapid and convenient, requiring no additional equipment (*see Note 9*).

1. Remove the storage buffer from the PD-10 column by decanting.
2. Equilibrate with VLP buffer by flowing 3×5 mL buffer through the column.
3. Apply 1 mL of sample; top up to 1 mL if necessary.
4. Apply 1.5 mL elution buffer and collect eluate.

3.4 Purification II: Size Exclusion Chromatography and Dynamic Light Scattering

SEC is an effective method to separate VLPs from possible aggregates (*see Fig. 5*). In addition, inline absorbance measurements can provide an indication of protein concentration and, in combination with light scattering, information on the size and integrity of the VLPs. To achieve this, SEC is commonly coupled to multi-angle light scattering (SEC-MALS) to give information on the molar mass and radius of gyration, usually as an analytical approach. Here we describe offline measurements of hydrodynamic radius in individual fractions collected from an SEC column as an alternative method more suited to preparative rather than analytical SEC.

3.4.1 SEC

1. For the separation of VLPs, a HiPrep 16/60 Sephadryl S-500 HR column (GE Healthcare) can be used. Prior to connecting the column, ensure the system is free of air and set maximum Δ column pressure alarm to 0.15 MPa (or as appropriate for the column).
2. Start pump and set flow rate to <1 mL/min. Connect column drop-by-drop to avoid introducing air and ensure wavelength is monitored at 280 nm to measure protein absorbance.
3. Equilibrate the column with 2 column volumes (CV) of filtered and degassed VLP buffer.
4. Filter sample through a 0.22 μ m filter or centrifuge at 16,000 g for 5 min. Inject sample (<5 mL) into an appropriately sized loop on the FPLC system and check that flow path configuration includes the sample loop.

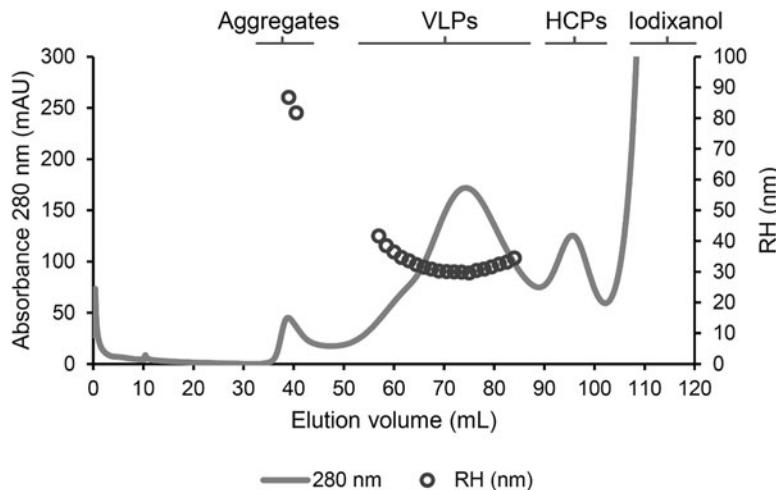


Fig. 5 SEC polishing coupled with dynamic light scattering. The SEC chromatogram plotted with the hydrodynamic radius measured offline by DLS for MPyV VLPs purified from plants. Three distinct peaks are evident before the iodixanol elution corresponding to aggregates, VLPs, and host cell protein (HCP) contaminants. Hydrodynamic radius as determined by DLS was taken for aggregate and VLP peak fractions

5. Elute with 1 CV of buffer at 0.5 mL/min. Flow rate may need to be adjusted depending on sample viscosity and operating temperature, or to improve resolution.
6. Collect fractions from the elution peak. Fraction size depends on the sample, but generally, 1–5 mL is sufficient.
7. Continue to pass buffer through column until UV baseline has stabilized to wash out residual iodixanol before loading subsequent samples (*see Note 10*).

3.4.2 DLS

1. A Dynapro plate reader (Wyatt Technology) can be used to measure all of the SEC fractions at once.
2. Load samples, 20–30 µL at a minimum concentration of 100 µg/mL, into 384-well polystyrene plates.
3. If measuring a large number of samples, it is advisable to overlay the sample with 5 µL of silicone oil to prevent evaporation.
4. Centrifuge the plate (1000 g, 5 min) to remove microscopic air bubbles.
5. Measure samples in duplicate with 10 acquisitions per well at 25 °C.
6. DYNAMICS software (Wyatt Technology) is used to perform a cumulants analysis, which provides information on hydrodynamic radius and polydispersity (*see Fig. 5*).

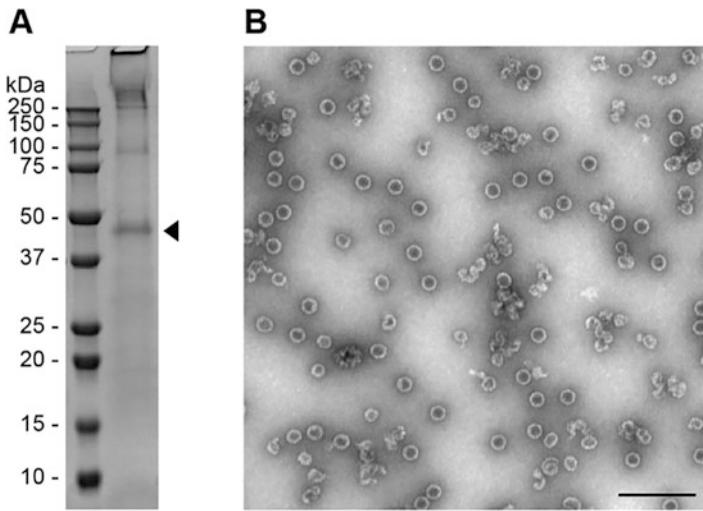


Fig. 6 Compositional analysis of VLPs derived from plants. (a) SDS-PAGE separation of the coat protein, VP1, of MPyV VLPs directly from an iodixanol gradient showing 47 kDa monomers (arrowhead) along with high molecular weight multimers. (b) TEM of MPyV VLPs; scale bar = 200 nm

3.5 Compositional Analysis

After confirming the estimation of VLP concentration, one can use various methods to characterize the particles further. The Bradford assay is not affected by residual iodixanol, so it is useful for quantification of desalted samples. It is also not affected by light scattering effects that can be misinterpreted as absorbance when analysing by UV/vis. This section briefly describes the microplate format Bradford assay followed by the confirmation of the presence of expected capsid proteins by SDS-PAGE and of VLPs assembled from those proteins by TEM (*see Fig. 6*). These should be taken as a minimum requirement for the demonstration of VLP production and purification.

3.5.1 Concentration Determination by Bradford Assay

1. Prepare a standard curve (between 0 and 1.5 mg/mL) using BSA stock solution (*see Note 11*).
2. Apply 5 µL of the standards and VLP protein samples in triplicate to a transparent 96-well plate.
3. Add 200 µL of the Coomassie reagent pre-warmed to room temperature to each well.
4. Incubate at room temperature for at least 5 min, but no longer than 60 min.
5. Measure absorbance at 595 and 470 nm.
6. Plot ratio of 595/470 as a function of the protein concentration.
7. Based on the polynomial fit to the standard curve, calculate the protein concentration in the VLP samples.

3.5.2 SDS-PAGE

1. Prepare samples with reducing SDS-PAGE sample buffer.
2. Denature the samples by heating at 95 °C for 5 min before cooling and briefly centrifuging to collect samples before gel loading.
3. Set up an electrophoresis tank with an Any kD TGX gel and fill the inner chamber with SDS-PAGE running buffer to inspect for leaks.
4. Load the samples onto the gel using an appropriate volume for the well size.
5. Fill the outer chamber with SDS-PAGE running buffer and run the gel at 200 V for approximately 30 min.
6. Remove the gel from the plastic cassette and rinse the gel in water for 10 min to remove the SDS.
7. Stain with GelCode BlueSafe Protein Stain for at least 60 min.
8. Destain in water until high-contrast bands are observed (usually overnight).

3.5.3 TEM

1. Pipette 5–10 µL VLP solution at 0.1–0.25 mg/mL onto 200-mesh carbon-coated copper grids and let settle for 1–2 min.
2. Wash the grids by placing them upside down onto a drop of water for 30 s.
3. Repeat the wash.
4. Apply negative stain to the grids by placing them upside down onto a drop of Uranyless EM Stain for 60 s.
5. Blot the excess stain with Whatman filter paper and leave to air dry for 2–5 min before placing it in a grid box.
6. Image using a Hitachi HT7700 or similar TEM at 80 kV (*see* Fig. 6b).

4 Notes

1. Acetosyringone (Sigma: 3',5'-Dimethoxy-4'-hydroxyacetophenone) should be added fresh immediately before use of the infiltration buffer. It is convenient to maintain filter-sterilized stock solutions of 0.1 M MES pH 5.6 and 0.1 M MgCl₂ both at room temperature; 0.1 M acetosyringone aliquoted at –20 °C.
2. Do not wait beyond 5-week-old plants (or when plant start to develop flowers) as the leaves tend to turn yellow and exhibit diffuse chlorosis and expression levels will be negatively affected.

3. For co-expression of two proteins, each individual agrobacterium strain should be grown and resuspended to an OD₆₀₀ of 0.4 independently. Bacterial suspensions can then be mixed prior to infiltration. For co-expression of more than two proteins, strains should be resuspended to an OD₆₀₀ that will allow for each individual strain to be at the equivalent of OD₆₀₀ ≥ 0.2 upon mixing.
4. The time of incubation should be empirically determined for each construct. Some constructs will lead to chlorosis and possibly eventual necrosis of the infiltrated tissue [16]. In these cases, it can help downstream processes to harvest before tissue death.
5. Buffers suited to non-enveloped VLPs need to be optimized. Depending on the stability and chemical properties of the VLP, the main considerations are pH and salt concentration. Typically, these values are around physiological levels of salt and pH, although there are exceptions for engineered protein cages. Traditionally, plant virus protocols make use of phosphate buffers. However, these can be incompatible with some uses of protein cages such as enzyme-containing nanoreactors, therefore, the general recommendation is for an appropriate Good buffer.
6. Following blending, the appearance of the lysate should be smooth and should contain no visible particles. If not, blend for a further minute until a smooth liquid consistency is reached. If ice-cold buffer is used, the lysate will not overheat.
7. It is possible to lay the gradient steps one on top of the other, starting with the highest density. However, an easier method is to start with the lowest density and, by using a long, blunt needle to get as close to the bottom of the tube as possible, add the next steps in ascending order to build up the gradient.
8. If a band cannot be seen or one is trying to optimize the purification of a novel VLP, or for recovery from a cushion, it is recommended to pierce a hole at the bottom of the tube and fractionate every 200 µL followed by SDS-PAGE to determine the position of VLPs on the gradient.
9. The removal of iodixanol via this method allows for analyses such as SDS-PAGE and TEM, but residual iodixanol after this step will interfere with DLS and UV-vis measurements due to viscosity and strong absorbance at 280 nm, respectively.
10. Although a desalting step is added, expect to have a prominent peak corresponding to iodixanol eluting after the 1st column volume. This will give a very high absorbance signal, possibly saturating the UV detector >4000 mAU for at least 2–3 CV.

11. Protein sample concentration for measurement should always be within the standard curve range. If the concentration is >1 mg/mL, dilute the sample and repeat the measurement.

Acknowledgments

This work was performed on the traditional lands of the Yugabarbul, Yuggera, Jagera, and Turrbal peoples. MV and DM acknowledge support in the form of Commonwealth Scientific and Industrial Research Organisation (CSIRO) Synthetic Biology Future Science Platform PhD Top-up Scholarship. FS acknowledges support from a CSIRO Synthetic Biology Future Science Platform Fellowship.

References

1. Lomonosoff GP, D'Aoust MA (2016) Plant-produced biopharmaceuticals: a case of technical developments driving clinical deployment. *Science* 353:1237–1240
2. Sainsbury F (2020) Innovation in plant-based transient protein expression for infectious disease prevention and preparedness. *Curr Opin Biotechnol* 61:110–115
3. Peyret H, Lomonosoff GP (2015) When plant virology met Agrobacterium: the rise of the deconstructed clones. *Plant Biotechnol J* 13: 1121–1135
4. Peyret H, Brown JKM, Lomonosoff GP (2019) Improving plant transient expression through the rational design of synthetic 5' and 3' untranslated regions. *Plant Methods* 15:108
5. Bally J, Jung H, Mortimer C, Naim F, Philips JG, Hellens R, Bombaréy A, Goodin MM, Waterhouse PM (2018) The rise and rise of *Nicotiana benthamiana*: a Plant for all Reasons. *Annu Rev Phytopathol* 56:405–426
6. Reed J, Stephenson MJ, Miettinen K, Brouwer B, Leveau A, Brett P, Goss RJM, Goossens A, O'Connell MA, Osbourn A (2017) A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab Eng* 42: 185–193
7. Sainsbury F, Lomonosoff GP (2014) Transient expressions of synthetic biology in plants. *Curr Opin Plant Biol* 19:1–7
8. Marsian J, Lomonosoff GP (2016) Molecular pharming—VLPs made in plants. *Curr Opin Biotechnol* 37:201–206
9. Chen Q, Lai H (2013) Plant-derived virus-like particles as vaccines. *Hum Vaccin Immunother* 9:26–49
10. Naupu PN, van Zyl AR, Rybicki EP, Hitzeroth II (2020) Immunogenicity of plant-produced human papillomavirus (HPV) virus-like particles (VLPs). *Vaccine* 8:740
11. Marsian J, Fox H, Bahar MW, Koteka A, Fry EE, Stuart DI, Macadam AJ, Rowlands DJ, Lomonosoff GP (2017) Plant-made polio type 3 stabilized VLPs—a candidate synthetic polio vaccine. *Nat Commun* 8:245
12. Marsian J, Hurdiss DL, Ranson NA, Ritala A, Paley R, Cano I, Lomonosoff GP (2019) Plant-made nervous necrosis virus-like particles protect fish against disease. *Front Plant Sci* 10: 880
13. Steele JFC, Peyret H, Saunders K, Castells-Graells R, Marsian J, Meshcheriakova Y, Lomonosoff GP (2017) Synthetic plant virology for nanobiotechnology and nanomedicine. *WIREs Nanomed Nanobiotechnol* 9:e1447
14. Peyret H, Gehin A, Thuenemann EC, Blond D, El Turabi A, Beales L, Clarke D, Gilbert RJC, Fry EE, Stuart DI, Holmes K, Stonehouse NJ, Whelan M, Rosenberg W, Lomonosoff GP, Rowlands DJ (2015) Tandem fusion of hepatitis B core antigen allows assembly of virus-like particles in bacteria and plants with enhanced capacity to accommodate foreign proteins. *PLoS One* 10:e0120751
15. Brillaire L, Jutras PV, Dashti N, Thuenemann EC, Morgan G, Lomonosoff GP, Landsberg MJ, Sainsbury F (2017) Engineering recombinant virus-like nanoparticles from plants for cellular delivery. *ACS Nano* 11:3476–3484
16. Catrice EVB, Sainsbury F (2015) Assembly and purification of polyomavirus-like particles from plants. *Mol Biotechnol* 57:904–913
17. Aljabali AAA, Sainsbury F, Lomonosoff GP, Evans DJ (2010) Cowpea mosaic virus

- unmodified empty viruslike particles loaded with metal and metal oxide. *Small* 6:818–821
- 18. Hesketh EL, Meshcheriakova Y, Dent KC, Saxena P, Thompson RF, Cockburn JJ, Lomonossoff GP, Ranson NA (2015) Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nat Commun* 6:10113
 - 19. Hesketh EL, Saunders K, Fisher C, Potze J, Stanley J, Lomonossoff GP, Ranson NA (2018) The 3.3 Å structure of a plant gemini-virus using cryo-EM. *Nat Commun* 9:2369
 - 20. Byrne MJ, Steele JFC, Hesketh EL, Walden M, Thompson RF, Lomonossoff GP, Ranson NA (2019) Combining transient expression and cryo-EM to obtain high-resolution structures of luteovirid particles. *Structure* 27:1761–1770. e1763
 - 21. Byrne M, Kashyap A, Esquirol L, Ranson N, Sainsbury F (2021) The structure of a plant-specific partitivirus capsid reveals a unique coat protein domain architecture with an intrinsically disordered protrusion. *Commun Biol* 4: 1155
 - 22. Castells-Graells R, Ribeiro JRS, Domitrovic T, Hesketh EL, Scarff CA, Johnson JE, Ranson NA, Lawson DM, Lomonossoff GP (2021) Plant-expressed virus-like particles reveal the intricate maturation process of a eukaryotic virus. *Commun Biol* 4:619
 - 23. Thuenemann EC, Meyers AE, Verwey J, Rybicki EP, Lomonossoff GP (2013) A method for rapid production of heteromultimeric protein complexes in plants: assembly of protective bluetongue virus-like particles. *Plant Biotechnol J* 11:839–846
 - 24. Saunders K, Sainsbury F, Lomonossoff GP (2009) Efficient generation of cowpea mosaic-virus empty virus-like particles by the proteolytic processing of precursors in insect cells and plants. *Virology* 393:329–337
 - 25. Peyret H (2015) A protocol for the gentle purification of virus-like particles produced in plants. *J Virol Methods* 225:59–63
 - 26. Sainsbury F, Saxena P, Aljabali AAA, Saunders K, Evans DJ, Lomonossoff GP (2014) Genetic engineering and characterization of cowpea mosaic virus empty virus-like particles. In: Lin B, Ratna B (eds) *Virus hybrids as nanomaterials: methods and protocols*. Humana Press, Totowa, pp 139–153
 - 27. Saxena P, Thuenemann EC, Sainsbury F, Lomonossoff GP (2016) Virus-derived vectors for the expression of multiple proteins in plants. In: MacDonald J, Kolotilin I, Menassa R (eds) *Recombinant proteins from plants: methods and protocols*. Springer New York, New York, pp 39–54
 - 28. Lee L-Y, Gelvin SB (2008) T-DNA binary vectors and systems. *Plant Physiol* 146:325–332
 - 29. Dashti NH, Sainsbury F (2020) Virus-derived nanoparticles. In: Gerrard JA, Domigan LJ (eds) *Protein nanotechnology: protocols, instrumentation, and applications*. Springer US, New York, pp 149–162



Chapter 23

Laboratory Scale Production of Complex Proteins Using Charge Complimentary Nanoenvironments

Girish Vallerinteavide Mavelli, Samira Sadeghi, and Chester Lee Drum

Abstract

Protein refolding is a crucial procedure in bacterial recombinant expression. Aggregation and misfolding are the two challenges that can affect the overall yield and specific activity of the folded proteins. We demonstrated the *in vitro* use of nanoscale “thermostable exoshells” (tES) to encapsulate, fold and release diverse protein substrates. With tES, the soluble yield, functional yield, and specific activity increased from 2-fold to >100-fold when compared to folding in its absence. On average, the soluble yield was determined to be 6.5 mg/100 mg of tES for a set of 12 diverse substrates evaluated. The electrostatic charge complementation between the tES interior and the protein substrate was considered as the primary determinant for functional folding. We thus describe a useful and simple method for *in vitro* folding that has been evaluated and implemented in our laboratory.

Key words Protein expression, Protein refolding, Inclusion bodies, *In vitro* folding, Thermostable exoshells, Protein nanoparticles, tES

1 Introduction

Heterologous protein expression in *E. coli* is usually accompanied by the formation of inclusion bodies [1, 2]. These are aggregates of the protein-of-interest (POI) and exist in non-native conformations. Recovery of bioactive soluble proteins from them requires the use of various refolding techniques [3, 4]. These methodologies are often time-consuming and further challenged by aggregation and misfolding, leading to low folded yields and a high cost of production [5, 6]. Moreover, the applicability of chaperones for *in vitro* protein production is restricted due to difficulties concerning their large-scale production, *in vitro* inactivation, and limited success rates with only a small number of protein substrates

folded [7–9]. It is noteworthy to mention that of all the in vitro folding protocols (RefoldDB), only 0.6% have used chaperones for protein folding [10]. To overcome these issues, refolding techniques need to undergo extensive development.

Recently, we have developed and extensively characterized the “thermostable ExoShells” or tES as a means to provide a simpler solution for nanoencapsulation of peptides or proteins [11–15]. Engineered from thermophilic *Archaeoglobus fulgidus* ferritin, they are 12 nm assembled protein nanoparticles with an 8 nm aqueous cavity, a volume that can theoretically accommodate molecules up to a volume of ~306 nm³ [16]. tES are characterized by point mutations that allow for pH titratable assembly and disassembly [tES-F116H] and net positive [tES-F116H(+)], negative [tES-F116H(–)] and neutral [tES-F116H(+/–)] interior charges [11]. All the three variants of tES can be expressed in high yields (≥ 400 mg/L culture) and can be purified ($\geq 95\%$) using simple methods. A previous study suggested that POIs covalently tethered to the tES interior can be folded and protected from external denaturants, though with low yields [11]. These encapsulations were promoted by complementary electrostatic interactions between the substrate protein and the tES interior [17, 18]. As these proteins are closely packed within the cavity via electrostatic stabilizations, we assume the subsequently reduced mobility prevents their aggregation and promotes structural and functional stability during folding. We thus tested the hypothesis that tES in the absence of a covalent linkage could encapsulate and promote protein folding (see Fig. 1). In detail, we studied the folding of phylogenetically diverse POIs varying in size, charge, cysteine density, and number of disulfides with the sample set including both monomeric and multimeric proteins [12] (see Fig. 2).

In general, we have developed a simple protocol for in vitro protein folding that can be used as a common laboratory practice (see Fig. 1). The protocol explains the recombinant expression of tES and denatured POIs, its purification followed by non-covalent nanoencapsulation of the POI within the tES cavity. Post disassembly by mild pH titration, the tES subunits are incubated with the denatured POI after which the tES assembly is initiated by increasing the pH of the mixture to 8.0. The encapsulated POI is dialyzed against a refolding buffer and the final folded protein is obtained through shell disassembly and chromatographic separation. We presume the yields obtained by this method can be sufficient for various physicochemical or structural-functional characterization of otherwise difficult to produce bioactive proteins.

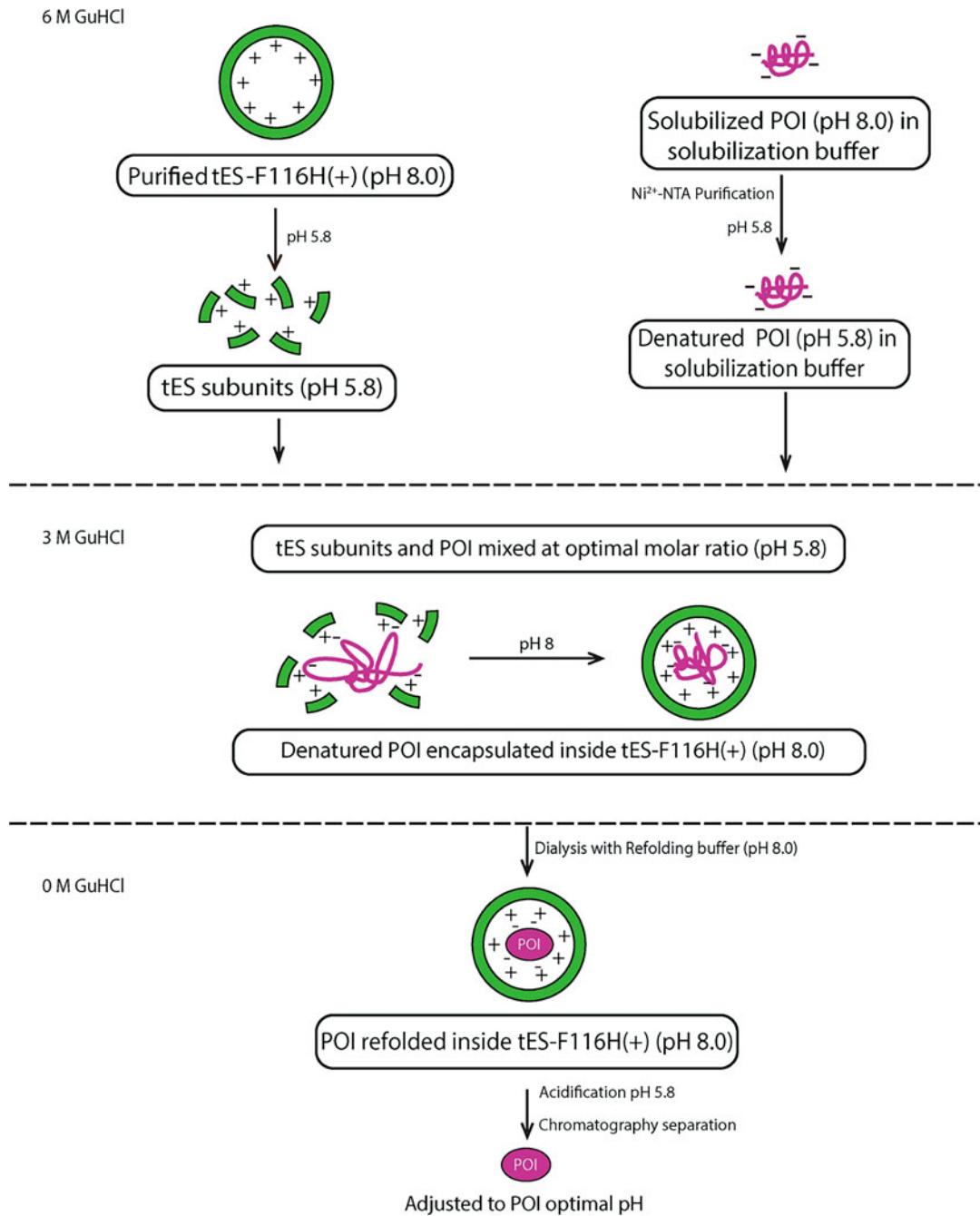


Fig. 1 Illustration of tES nanoencapsulation and folding protocol depicting the charge complementarity based association and folding

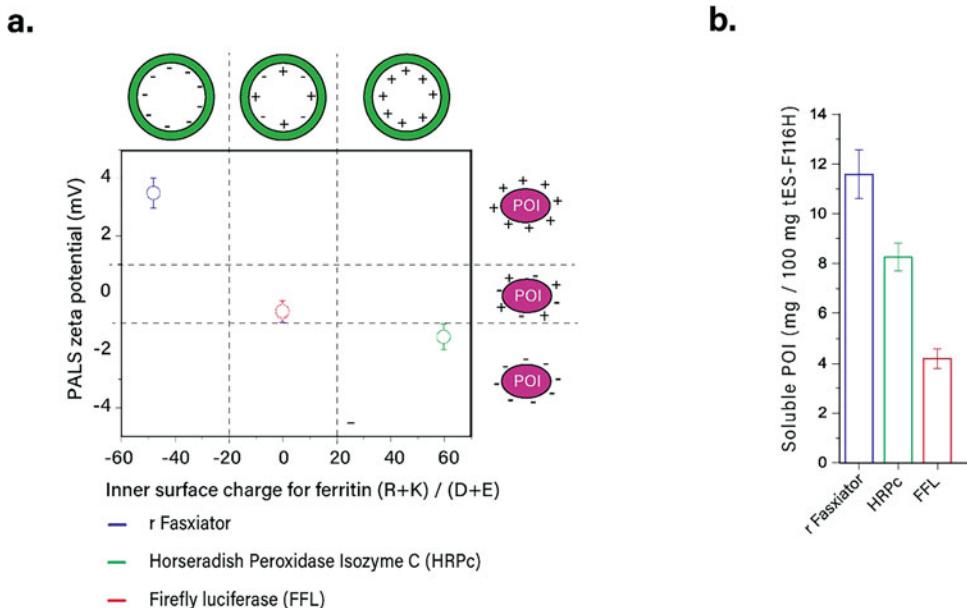


Fig. 2 In vitro folding of representative POIs using tES variants. **(a)** Complementary charge effects between the interior cavity of tES variants and the measured zeta potential of POI predict optimal folding. **(b)** Soluble yields of folded POI through nanoencapsulation expressed as mg POI/100 mg of tES-F116H. (Adapted from Nat. Commun. 12, 5720, 2021)

2 Materials

Prepare all buffers, antibiotics (kanamycin and ampicillin), inducers (isopropylthio- β -galactoside, IPTG; L-arabinose), and other reagents with MilliQ ultrapure water. Sterilize all the buffers and reagents (antibiotics, inducers, and others) with 0.22 μ m filters. Store buffers at room temperature (RT) and aliquots of kanamycin (50 mg/mL), ampicillin (100 mg/mL), and IPTG (1 M) stocks at -20 °C. Prepare L-arabinose stocks (1%) freshly before use.

2.1 Bacterial Expression

1. Agar plate: Luria-Bertani (LB) agar. Suspend 40 g of LB agar in 1 L of water and autoclave. Add kanamycin or ampicillin to final concentrations of 50 μ g/mL and 100 μ g/mL, respectively when the media is warm. Pour 15 mL of the warm agar media (~40 °C) into sterile Petri dishes (90 \times 15 mm) (*see Note 1*). Store the prepared plates at 4 °C.
2. Growth media: Terrific Broth (TB) and LB media. For TB starter cultures, dissolve 4.7 g of TB powder and 0.8 mL of glycerol to a final volume of 100 mL with water. For TB flask cultures, dissolve 47 g of TB powder and 8 mL of glycerol to a final volume of 1 L with water. For LB starter culture, dissolve

2.5 g of LB broth powder in 100 mL water. Similarly, for LB flask culture, dissolve 25 g of LB broth powder in 1 L water. Autoclave all the media and store the media at 4 °C (*see Note 2*).

2.2 Isolation of Proteins

1. tES Lysis buffer: 50 mM Tris–HCl, 150 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, pH 8.0. Weigh 6.057 g of Tris–HCl salt into 1 L beaker. Add 900 mL of water and mix on a magnetic stirrer until all the salts are dissolved. To this, add 8.8 g of NaCl, 0.34 g of disodium salt of EDTA, and 1 mL of Triton X-100. Mix and adjust pH to 8.0 with 1 M HCl. Make up the volume to 1 L using water, filter, and store at RT (*see Note 3*).
2. POI lysis buffer: 50 mM Tris–HCl, 200 mM NaCl, 5 mM DTT, 10 mM EDTA, 1.5% Triton X-100, pH 8.0. Weigh 6.057 g of Tris–HCl salt into 1 L beaker. Add 900 mL of water and mix on a magnetic stirrer until all the salts are dissolved. To this, add 11.69 g of NaCl, 3.4 g of disodium salt of EDTA, and 15 mL of Triton X-100. Mix and adjust pH to 8.0 with 1 M HCl. Make up the volume to 1 L, filter, and store at RT (*see Note 3*). Just before use, add and mix DTT to make a final concentration of 5 mM (*see Note 4*).
3. Wash Buffer 1: 50 mM Tris–HCl, 200 mM NaCl, 5 mM DTT, 10 mM EDTA, 2 M urea, 1% Triton X-100, pH 8.0. Weigh 6.057 g of Tris–HCl salt into 1 L beaker. Add 900 mL of water and mix on a magnetic stirrer until all the salts are dissolved. To this, add 11.69 g of NaCl, 3.4 g of disodium salt of EDTA, and 10 mL of Triton X-100. Mix and adjust pH to 8.0 with 1 M HCl. Make up the volume to 1 L, filter, and store at RT (*see Note 3*). Just before use, add and mix DTT and urea to make final concentrations of 5 mM and 2 M, respectively (*see Note 5*).
4. Wash buffer 2: 50 mM Tris–HCl, 200 mM NaCl, 5 mM DTT, 5 mM EDTA, pH 8.0. Weigh 6.057 g of Tris–HCl salt into 1 L beaker. Add 900 mL of water and mix on a magnetic stirrer until all the salts are dissolved. To this, add 11.69 g of NaCl and 1.68 g of disodium salt of EDTA. Mix and adjust pH with 1 M HCl to 8.0. Make up the volume to 1 L, filter, and store at RT (*see Note 3*). Just before use, add and mix DTT to make a final concentration of 5 mM (*see Note 4*).
5. Solubilization buffer: 50 mM Tris–HCl, 200 mM NaCl, 2 mM β-mercaptoethanol (BME), 6 M guanidine hydrochloride (GuHCl), pH 8.0. Weigh and dissolve 6.057 g Tris–HCl salt in minimal water. To this, add 11.69 g of NaCl, and 573.18 g of GuHCl (*see Note 6*). Mix and adjust pH to 8.0. Make up the volume to 1 L, filter, and store at RT (*see Note 3*). Just before use, add and mix BME to a final concentration of 2 mM (*see Note 7*).

2.3 Purification of Proteins

1. Size-exclusion chromatography (SEC) buffer 1: 50 mM Tris-HCl, pH 8.0. Weigh 6.057 g Tris-HCl salt into 1 L beaker. Add 900 mL of water and mix on a magnetic stirrer until all the salts are dissolved. Adjust pH to 8.0 with 1 M HCl. Make up the volume to 1 L using water, filter, and store at RT (*see Note 3*).
2. Size-exclusion chromatography (SEC) buffer 2: 50 mM Sodium acetate, pH 5.4/5.6 (*see Note 8*). Weigh and dissolve 4.1 g of sodium acetate in 900 mL of water and adjust the pH to either 5.4 or 5.6 using acetic acid. Make up the volume to 1 L, filter, and store at RT (*see Note 3*).
3. Ni²⁺-NTA affinity chromatography binding buffer: 50 mM Tris-HCl, 200 mM NaCl, 2 mM BME, and 6 M GuHCl, pH 8.0. Weigh and dissolve 6.057 g of Tris-HCl salt with water on a magnetic stirrer. To this, add 11.69 g of NaCl, and 573.18 g of GuHCl. Mix and adjust pH to 8.0 with 1 M HCl. Make up the volume to 1 L using water, filter, and store at RT (*see Notes 3 and 6*). Just before use, add and mix BME to a final concentration of 2 mM (*see Note 7*).
4. Ni²⁺-NTA affinity chromatography washing buffer: 50 mM Tris-HCl, 200 mM NaCl, 2 mM BME, 6 M GuHCl, and 20 mM imidazole, pH 8.0 (*see Note 9*). Weigh and dissolve 6.057 g of Tris-HCl salt with water on a magnetic stirrer. To this, add 11.69 g of NaCl, 1573.18 g of GuHCl, and 1.4 g of imidazole. Mix and adjust pH to 8.0 with 1 M HCl. Make up the volume to 1 L using water, filter, and store at RT (*see Notes 3 and 6*). Just before use, add and mix BME to a final concentration of 2 mM (*see Note 7*).
5. Ni²⁺-NTA affinity chromatography elution buffer: 50 mM Tris-HCl, 200 mM NaCl, 2 mM BME, 6 M GuHCl, and 250 mM imidazole, pH 8.0. Weigh and dissolve 6.057 g of Tris-HCl salt with water on a magnetic stirrer. To this, add 11.69 g of NaCl, 573.18 g of GuHCl, and 17.02 g of imidazole. Mix and adjust pH to 8.0 with 1 M HCl. Make up the volume to 1 L using water, filter, and store at RT (*see Notes 3 and 6*). Just before use, add and mix BME to a final concentration of 2 mM (*see Note 7*).

2.4 Protein Folding

1. Refolding buffer: 50 mM Tris-HCl, 200 mM NaCl, 5% glycerol, pH 8.0 (*see Note 10*). Weigh 6.057 g of Tris-HCl salt into 1 L beaker. Add 100 mL of water and mix on a magnetic stirrer until all the salts are dissolved. To this, add 11.69 g of NaCl, and 50 mL of 100% glycerol. Mix and adjust pH with 1 M HCl to 8.0. Make up the volume to 1 L using water, filter, and store at RT (*see Note 3*).

3 Methods

Conduct all experiments at RT unless otherwise stated. Perform all bacterial cloning, plating, and inoculation works under a sterile environment. Incubate all the bacterial cultures overnight at 37 °C. Use the antibiotics, kanamycin, and ampicillin for all the tES and POI clones, respectively. For agar plates, the antibiotic concentrations used are kanamycin: 50 µg/mL; ampicillin: 100 µg/mL. For broth cultures, the antibiotic concentrations used are kanamycin: 50 µg/mL; ampicillin: 50 µg/mL. TB and LB media are used for tES and POI expression, respectively. Perform all bacterial centrifugations at 13,750 × *g* for 20 min unless otherwise stated.

3.1 Preparation of tES Variant Clones

1. Clone the genes encoding tES variants (with mutation(s) for c-terminus truncation, altered charges, and pH-dependent assembly/disassembly) to pRSF1b expression vector (*see Note 11*).
2. Transform the ligation mixture to chemically competent XL1 Blue *E. coli* cells. Plate them on LB agar and incubate at 37 °C (*see Note 12*).
3. Select positive colonies, culture them in 5 mL LB broth, and isolate the plasmid DNA using a plasmid miniprep kit (*see Note 13*).
4. Sequence the isolated and purified plasmids to confirm the positive clones. Transform the positive tES variant clones into BL21(DE3) *E. coli* cells. Plate them on LB agar and incubate at 37 °C.
5. Pick individual colonies to inoculate 5 mL LB cultures. Prepare glycerol stocks with 50% glycerol and store at –80 °C for future expression works (*see Note 14*).

3.2 Preparation of h6POI (6× Histidine-Protein-of-Interest) Clones

1. Clone the genes encoding h6POI to pBAD/HisB expression vector.
2. Transform the ligation mixture to chemically competent XL1 Blue *E. coli* cells. Plate them (*see Note 12*) on LB agar and incubate at 37 °C.
3. Select positive colonies, culture them in 5 mL LB broth, and isolate the plasmid DNA using a plasmid miniprep kit (*see Note 13*).
4. Sequence the isolated and purified plasmid to confirm the positive clones. Transform the positive h6POI clones into BL21(DE3) *E. coli* cells. Plate them on LB agar at 37 °C.

- Pick individual colonies to inoculate 5 mL LB cultures. Prepare glycerol stocks with 50% glycerol and store at –80 °C for future expression works (*see Note 14*).

3.3 Expression of tES Variants

- Inoculate the TB starter culture using the glycerol stocks of tES clones and grow the cultures at 37 °C (*see Note 15*).
- Inoculate 1 L flask with 25 mL of fully-grown starter culture and allow to grow until the absorbance (OD600) of 0.4–0.5 reaches.
- Induce the protein expression by adding IPTG at a final concentration of 0.4 mM. Continue to grow the culture for another 5 h and pellet down the bacterial cells by centrifugation at the end of the culture (*see Note 16*).

3.4 Expression of h6POI

- Inoculate the LB starter culture using the glycerol stocks of h6POI clones and grow the cultures at 37 °C (*see Note 17*).
- Inoculate 1 L flask with 25 mL of fully-grown starter culture and allow them to grow until the absorbance (OD600) of 0.4–0.5 is reached.
- Induce the protein expression by adding L-arabinose at a final concentration of 0.1% from the freshly prepared 1% stock. Continue to grow the culture for another 5 h and pellet down the bacterial cells by centrifugation at the end of culture (*see Notes 16 and 18*).

3.5 Isolation and Purification of tES Variants

- Resuspend the cell pellet in lysis buffer (refer to Subheading [2.2](#), item 1), sonicate, and centrifuge to separate the cell debris (*see Note 19*).
- Heat the supernatant at 70 °C water bath for 10 min and centrifuge to separate the precipitate (*see Note 20*).
- Run the supernatant on SEC (refer to Subheading [2.3](#), item 1) to purify the tES proteins (*see Note 21*).
- Concentrate, quantify the proteins in nanodrop, and store them as aliquots at –20 °C (*see Note 22*).
- Alternatively, buffer exchange purified tES variants with 10 mM ammonium bicarbonate pH 8.0 and lyophilize to obtain a salt-free powder. The freeze-dried tES can be stored at RT for long durations and used for folding denatured POIs (*see Note 23*).

3.6 Isolation and Purification of h6POIs

- Resuspend the cell pellet in lysis buffer (refer to Subheading [2.2](#), item 2), sonicate (*see Note 19*), and centrifuge to obtain the cell pellet containing inclusion bodies.
- Wash the cell pellets with wash buffers 1 and 2 (refer to Subheading [2.2](#), items 3 and 4). Centrifuge at each step to separate

the pellets. Resuspend the cell pellets in solubilization or binding buffer (refer to Subheading 2.2, item 5 or Subheading 2.3, item 3) to solubilize the inclusion bodies for 1–2 h (*see Note 24*), followed by centrifugation.

3. Subject the supernatant to Ni²⁺-NTA affinity chromatography. Treat the supernatant with washed Ni²⁺-NTA resins and incubate at 4 °C for 30–60 min on a rocking platform for binding (*see Note 25*).
4. Centrifuge the resins at low speed ($800 \times g$) for 1 min. Aspirate the supernatant and store it at 4 °C as a flowthrough.
5. Wash the resins with 3× volume of wash buffer (refer to Subheading 2.3, item 4). Centrifuge and aspirate the supernatant. Store the supernatant at 4 °C as wash 1.
6. Repeat step 5 one more time and store the supernatant as wash 2 (*see Note 26*).
7. Elute the bound proteins from the resins using the imidazole buffer (refer to Subheading 2.3, item 5). Centrifuge and aspirate the supernatant. Store the supernatant at 4 °C as eluate 1.
8. Repeat step 7 and store the supernatant as eluate 2 (*see Note 27*).
9. Analyze all the stored fractions (flowthrough, wash 1, wash 2, eluate 1, and eluate 2) in SDS-PAGE. Quantify the eluted pure POIs. Add DTT to the protein fractions to a final concentration of 10 mM (*see Note 28*). Store them at 4 °C for immediate use.

3.7 In Vitro Folding

Use this optimized protocol to fold monomeric proteins (*see Fig. 1*). For multimeric proteins, additional steps have to be followed as mentioned:

1. Buffer exchange the purified POI to 50 mM Sodium acetate, 10 mM DTT, and 6 M guanidine hydrochloride pH 5.4/5.6 using centrifugal concentrators (select molecular weight cut-off based on POI molecular weight) (*see Note 8*).
2. Acidify purified tES variants and buffer exchange to 50 mM Sodium acetate, pH 5.4/5.6 using centrifugal concentrators (10 kDa molecular weight cut-off, MWCO) to obtain the subunits (*see Note 8*).
3. Incubate the POIs with tES subunits at RT for 1 h at pH 5.4/5.6. Use ratios of POI to tES subunits where maximum encapsulation efficiency is observed (*see Note 29*).
4. Adjust the pH of the mixture to 8.0 and set the mixture for 24 h dialysis using the refolding buffer (refer to Subheading 2.4, item 1) (*see Note 30*).

5. Purify the folded encapsulated proteins in SEC run at pH 8.0 buffer. Acidify the fractions using 1 M acetic acid followed by a second SEC (refer to Subheading 2.3, item 2) using 50 mM Sodium acetate pH 5.4/5.6 buffer to isolate the POIs from tES subunits. Use Superdex 200 10/300 (Cytiva, Marlborough, MA, USA) column for all the SEC runs. Collect the folded protein fractions and adjust to the required pH. After measuring the yield, analyze their mass in SDS gel (*see Note 31*).
6. For folding oligomeric proteins, perform steps 1–5 to obtain the structural monomers. Incubate them in an appropriate environment to facilitate the assembly to form a functional oligomer (*see Note 32*).

4 Notes

1. Prepare the plates under sterile conditions. Keep the lid open while agar solidifies to avoid condensation on the lids. Label and store the plates upside down under refrigerated conditions. Make sure not to use the antibiotic plates after a month of preparation. Before using the agar plates, check for any visible microbial contamination.
2. We observed that for tES expression, TB media provided two to threefolds higher cell density and protein yields compared to commonly used LB media. Further, we obtained yields of ~400–600 mg of purified proteins from 1 L flask cultures [13]. Antibiotics, kanamycin, and ampicillin are heat-labile and must be used after the autoclave. Thaw the frozen antibiotic stocks at room temperature and add under the sterile condition at a final concentration of 50 µg/mL of kanamycin or 50 µg/mL of ampicillin. Swirl the culture flasks to make sure they are mixed properly.
3. Concentrated HCl is 12.1 M. Prepare a 1 M HCl solution for adjusting pH. Check the pH of the buffer every time before using it for experiments. We advise not to store the buffers at RT for more than a month.
4. Select the pH of the lysis buffer based on the isoelectric point (pI) of the protein. Use DTT if POI has free cysteines or disulfide bonds. DTT is unstable in solution and oxidizes over time. Thus, care must be taken to prepare fresh buffer solutions containing DTT [19].
5. Use 2 M urea wash buffer to remove those contaminating proteins non-specifically bound to inclusion bodies. Always prepare urea buffer freshly as it can generate reactive cyanate ions on standing [20].

6. We advise using 6 M GuHCl over 8 M urea as it is one to two times more effective as a protein denaturant and can be stored for longer periods at 4 °C. Further, protein modifications like carbamylation observed with the use of urea and enhanced by temperature and storage conditions are not seen with the use of GuHCl [21]. However, care must be taken while dissolving the GuHCl salts as the buffer volume can exceed beyond the required volume if the entire salts are dissolved in one go. Hence, we advise dissolving the required amount of GuHCl gradually at 35 °C. Once all the GuHCl salts are dissolved, adjust the pH to 8.0 and make up the volume to 1 L, and store at RT.
7. DTT reacts with Nickel ions (Ni^{2+}) in the Ni^{2+} -NTA resin to form a black precipitate that strips off the bound Ni^{2+} , thus damaging the affinity resin [19]. Hence, use BME as a reducing agent in the solubilization buffer. DTT can be added to the purified fractions at the end of Ni^{2+} -NTA elution at a final concentration of 5 mM. Avoid EDTA from solubilization buffers as it can chelate and remove Ni^{2+} ions from the resin thereby affecting the binding efficiency.
8. We observed complete dissociation into subunits at pH 5.4 for tES-F116H(+/−) and tES-F116H(−) and at pH 5.6. for tES-F116H(+). Note that the pI of tES variants vary with the positive shell having a theoretical pI of 5.46, the neutral shell having a pI of 4.9, and the negative shell having a pI of 4.77.
9. We advise using ~20 mM imidazole in the wash buffers to remove contaminant proteins, the efficacy of this can be analyzed through SDS-PAGE. Prepare fresh imidazole buffers if the solution turns yellow during long-term storage.
10. We advise modifying the refolding buffer according to the POI characteristics. Use specific additives such as metal ions necessary for the structural and functional folding of POIs like enzymes, DNA binding proteins, and others.
11. tES was engineered from the *Archaeoglobus fulgidus* ferritin gene sequence (GenBank: AF_RS04235). The gene with mutations for c-terminus truncation, shell disassembly/assembly, and altered charges was synthesized from GenScript. The tES variant genes synthesized are [11]:
tES-F116H(+) – E65K, E128K, E131K, D138A, F116H, ΔC after Q164
tES-F116H(+/−) – E65Q, D138A, F116H, ΔC after Q164
tES-F116H(−) – F116H, ΔC after Q164
12. Use either sterile cell spreaders or autoclaved glass beads for plating.

13. Determine the plasmid DNA concentrations for the three tES variants using nanodrop. Label and store at -20°C .
14. Prepare the glycerol stock by gently mixing 500 μL of fully-grown culture with 500 μL of sterile 50% glycerol in a 2 mL sterile tube or cryovial. Label the glycerol stock tubes and freeze them at -80°C for prolonged storage.
15. Place the glycerol stock vial on an icebox or dry ice. Recover the bacteria from the top of the vial using a sterile pipette tip by scraping some of the frozen cultures. Try not to freeze/thaw the glycerol stock multiple times. Discard the frozen stock if reduced viability is observed and use a fresh vial. As tES is involved in multiple applications including in vitro folding and due to its high demand, we optimized its expression over various nutrient media and found that TB cultures gave higher cell density and protein yield compared to LB and others [13].
16. The bacterial pellets can be stored at -20°C for up to a month, or at -80°C for longer durations.
17. We used initially LB media to express POIs, however, a nutrient media optimization can be performed for individual POIs to achieve a higher cell density and yield. For handling glycerol stocks, refer to **Note 14**.
18. Care must be taken to provide appropriate post-induction expression conditions for the proteins. For example, p53 expression requires zinc chloride during induction followed by its incubation at 30°C .
19. We usually obtain around 5–7 g of cell pellet from TB bacterial cultures. Resuspend them in 30–40 mL of lysis buffer and sonicate in short pulses on an ice bed. The sonication parameters have to be determined based on the ultrasonic probe sonicator used. Make sure the sample does not get heated during the procedure.
20. The heating denatures and precipitates out all contaminant proteins except the thermostable tES (denaturation temperature $>90^{\circ}\text{C}$). Make sure to monitor the temperature using a thermometer. The supernatant having tES can be filtered through 0.22 μm syringe filters to remove any aggregates or particulates before SEC purification.
21. Use a superdex 200 10/300 (Cytiva, Marlborough, MA, USA) column with a separation range of 10–600 kDa to purify the tES variants. Calibrate the SEC column using protein molecular standards before running the sample. tES will elute around 9–10 mL. We have observed aggregates of tES eluting in the void volume ($\sim 8 \text{ mL}$) as a small peak. Collect fractions under the monomeric tES peak. The purity of the protein fractions can be analyzed in SDS-PAGE.

22. tES is highly stable at RT storage however we have observed some precipitation during prolonged storage as the protein is highly hydrophobic. We advise to aliquot and store them at -20°C . The concentration of the purified shell proteins can be measured on a Nanodrop as per Beer–Lambert’s equation by measuring the absorbance at 280 nm and using the molar extinction coefficient of $\epsilon_{280} = 814,080 \text{ M}^{-1} \text{ cm}^{-1}$.
23. Buffer exchanged pure tES variants can be either frozen at -80°C or flash freeze with liquid nitrogen before lyophilization. As ammonium bicarbonate is a volatile solution, after freeze-drying it does not leave any trace of salts with the protein powder. The freeze-dried tES powder is structurally stable at RT storage and after reconstitution, retains the ability to disassemble/assemble under mild pH titrations.
24. The time of solubilization can vary with POI. Do not use DTT in the solubilization buffer as it can negatively impact the Ni^{2+} -NTA resins (*discussed in Note 7*). EDTA must be avoided from the solubilization or binding buffers as it can strip the Ni^{2+} from the resins. Be sure to check the pH of the buffers before using them.
25. Commercially available Ni^{2+} -NTA resins are usually stored in 20% ethanol. Resuspend the resins in the bottle by gentle mixing. Take out required amounts and wash first with water followed by binding buffer (refer to Subheading 2.3, item 3). The incubation time has to be determined depending upon the binding affinity of the POI to the resin. The binding buffer which is the same as the solubilization buffer can have 10 mM of imidazole to prevent nonspecific interaction of contaminant proteins with resins. Centrifuge the Ni^{2+} -NTA resins at low speed ($800 \times g$) for 1–2 min. Carefully, aspirate the supernatant (flowthrough) without disturbing the resin bed and store an aliquot at 4°C for SDS-PAGE analysis. Protein precipitation observed if any, during the binding phase can be avoided using Triton-X 100/Tween 20 (up to 1%) or glycerol (up to 10%) in the protein buffers [22].
26. Increase wash steps if contaminating proteins are high in the final eluate. Alternatively, use up to 50 mM imidazole in the wash buffers to remove the contaminants without affecting POI binding to the resin.
27. If the POI is still bound to the resin, either increase the number of elution steps or imidazole concentration (up to 500 mM) or decrease the pH of the buffer. After the purification is complete, wash the resin with water to remove the traces of elution buffer followed by 0.5 M NaOH to remove any bound proteins. Wash with water to remove NaOH and store them at 4°C in 20% ethanol.

28. The concentration of the purified POIs can be measured on a Nanodrop as per Beer–Lambert’s equation by measuring the absorbance at 280 nm and using their specific molar extinction coefficient. Use GuHCl-free fractions for SDS-PAGE analysis as SDS cause precipitation while reacting with GuHCl [19]. Review and troubleshoot the POI purification based on the SDS-PAGE analysis:
- (a) Low recovery of h6POI after final elution
 - (i) Check the expression level.
 - (ii) Check the wash steps, reduce the number of steps as well as the imidazole concentration in wash buffers.
 - (iii) Make sure the POI is not degraded or the His tag is not cleaved. Use protease inhibitor cocktails in the lysis buffer and perform all purification steps at 4 °C.
 - (iv) Make sure the POI does not have a high affinity to the resin. If so, follow **Note 27**.
 - (v) Use high salt concentrations of up to 500 mM NaCl in the buffers to disrupt any nonspecific interactions between negatively charged proteins and Ni²⁺-NTA resins (resins have a net positive charge).
 - (b) POI observed in the flowthrough and wash fractions
 - (i) Make sure the resins are not overloaded with POIs. Check the concentration of POIs before loading. Either use less protein or more resins in such cases.
 - (ii) Check the resins for their binding efficiency. In such cases, use either freshly charged Ni²⁺-NTA resins or check the pH of the binding buffer.
 - (c) Contaminant proteins observed in the eluate [22]
 - (i) Use BME at concentrations up to 10 mM in the binding, wash, and elution buffer to break any non-specific disulfides formed between POI and contaminant proteins.
 - (ii) Use of Triton-X 100 or Tween 20 up to 1% in the buffers or glycerol up to 10% can disrupt any non-specific hydrophobic interactions.
 - (iii) Increase the wash steps or imidazole concentration in wash buffers as mentioned in **Note 26**.
 - (iv) Follow a second Ni²⁺-NTA purification with a fresh batch of resin after buffer exchange of eluate with binding buffer.
 - (v) Try a step-gradient elution with either increasing imidazole concentration or decreasing the pH of the elution buffer.

29. Determine the optimum tES_{subunits}:POI ratio to maximize the encapsulation and folding efficiency by titrating fixed concentrations of tES_{subunits} with varying concentrations of denatured POI. After removal of denaturants (*see* Fig. 1), the soluble fraction can be assayed to quantify them while the monomer is still inside or requires release from the shell. As the nanoencapsulation is based on charge complementarity, we observed efficient folding of negatively charged POIs within tES-F116H (+) and vice versa (*see* Fig. 2). For encapsulation optimizations, use dialysis cassettes (0.5 mL) with membrane MWCO selected based on the POI and tES_{subunit} molecular size. Use of solubilized inclusion body suspensions without purification can also be tested for in vitro folding, provided the suspension is >80% pure with POI.
30. We used dialysis cassettes (3 mL) with the membrane MWCO selected based on the POI and tES_{subunit} molecular size. Follow the instructions in using the dialysis cassettes. Observe for any precipitation. Usually, POIs with high cysteine density can precipitate. In that case, perform step dialysis where the DTT and GuHCl concentrations can be reduced gradually. Adjust pH to 8.0 using 1 M NaOH. Immerse the dialysis cassettes in a beaker containing a large volume of the refolding buffer (usually 1 L buffer for 1–2 mL sample) and gently stir at the desired temperature. Consider the thermal stability of the POI while determining the dialysis temperature. Change the dialysis buffer as necessary. Finally, remove the cassette from the buffer and carefully take out the sample based on the user manual.
31. If POI is partially released from tES after final acidification, high salt concentrations of up to 200 mM NaCl can be used with the release buffers (50 mM Sodium acetate pH 5.4/5.6). Further, the buffer for separating the POI from subunits can have mild NaCl concentrations of 75–100 mM to prevent any secondary interactions between the POI and the SEC resin matrix. To isolate POIs from tES subunits sharing a similar molecular mass of 19 kDa, use different chromatographic techniques like ion exchange or hydrophobic interaction chromatography based on the protein characteristics like isoelectric point or hydrophobicity, respectively.
32. The incubation condition and buffer composition for the oligomerization of structural monomers into multimeric proteins will vary depending on the POI. Make sure the oligomerization buffer contains all additives necessary for the structural and functional stability of the oligomer. For instance, POIs like alkaline phosphatase (dimer) or p53 (tetramer) require its monomers to be incubated at 4 °C for 1–2 h whereas for trimeric Omp2a a bacterial porin, the folded monomers need to be incubated at 37 °C for 9 days to facilitate multimerization.

Acknowledgments

This research was funded by NMRC-Clinician Scientist Award, CSAINV17nov012, and A*STAR-AME, A2083c0055. We thank Kong Shik Nie and Siddhesh Sujit Vaidya for their recent contributions to the study.

References

1. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK (2015) Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microp Cell Factories* 14:1–10
2. Ramón A, Señorale M, Marín M (2014) Inclusion bodies: not that bad.... *Front Microbiol* 5:56
3. Yamaguchi H, Miyazaki M (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* 4:235–251
4. Alibolandi M, Mirzahoseini H (2011) Chemical assistance in refolding of bacterial inclusion bodies. *Biochem Res Int* 2011:631607
5. Humer D, Spadiut O (2018) Wanted: more monitoring and control during inclusion body processing. *World J Microbiol Biotechnol* 34: 1–9
6. Eiberle MK, Jungbauer A (2010) Technical refolding of proteins: do we have freedom to operate? *Biotechnol J* 5:547–559
7. Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* 16:574–581
8. Jhamb K, Jawed A, Sahoo DK (2008) Immobilized chaperones: a productive alternative to refolding of bacterial inclusion body proteins. *Process Biochem* 43:587–597
9. Ma FH, Li C, Liu Y, Shi L (2020) Mimicking molecular chaperones to regulate protein folding. *Adv Mater* 32:1805945
10. Mizutani H, Sugawara H, Buckle AM, Sangawa T, Miyazono K-i, Ohtsuka J, Nagata K, Shojima T, Nosaki S, Xu Y (2018) REFOLDdb: a new and sustainable gateway to experimental protocols for protein refolding. *BMC Struct Biol* 17:1–8
11. Deshpande S, Masurkar ND, Girish VM, Desai M, Chakraborty G, Chan JM, Drum CL (2017) Thermostable exoshells fold and stabilize recombinant proteins. *Nat Commun* 8:1–8
12. Sadeghi S, Deshpande S, Girish VM, Aksoyoglu A, Bafna J, Winterhalter M, Kini RM, Lane DP, Drum CL (2021) A general approach to protein folding using thermostable exoshells. *Nat Commun* 12:1–15
13. Girish VM, Sadeghi S, Vaidya SS, Kong SN, Drum CL (2021) Nanoencapsulation as a general solution for lyophilization of labile substrates. *Pharmaceutics* 13:1790
14. Sadeghi S, Masurkar ND, Girish VM, Deshpande S, Kok Yong Tan W, Yee S, Kang S-A, Lim Y-P, Kai Hua Chow E, Drum CL (2022) Bioorthogonal catalysis for treatment of solid tumors using thermostable, self-assembling, single enzyme nanoparticles and natural product conversion with indole-3-acetic acid. *ACS Nano* 16:10292–10301
15. Sadeghi S, Vallerinteavide Mavelli G, Vaidya SS, Drum CL (2022) Gastrointestinal tract stabilized protein delivery using disulfide thermostable exoshell system. *Int J Mol Sci* 23:9856
16. Voss NR, Gerstein M (2010) 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res* 38(suppl_2):W555–W562
17. Pulsipher KW, Bulos JA, Villegas JA, Saven JG, Dmochowski JJ (2018) A protein–protein host–guest complex: thermostable ferritin encapsulating positively supercharged green fluorescent protein. *Protein Sci* 27:1755–1766
18. Chakraborti S, Lin T-Y, Glatt S, Heddle JG (2020) Enzyme encapsulation by protein cages. *RSC Adv* 10:13293–13301
19. Palmer I, Wingfield PT (2012) Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. *Current Protoc Protein Sci* 70:6.3. 1–6.3. 20
20. Sun S, Zhou J-Y, Yang W, Zhang H (2014) Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. *Anal Biochem* 446:76–81
21. Wingfield PT (1995) Use of protein folding reagents. *Curr Protoc Protein Sci* (1):A. 3A. 1–A. 3A. 4
22. Bornhorst JA, Falke JJ (2000) [16] Purification of proteins using polyhistidine affinity tags. *Methods Enzymol* 326:245–254

INDEX

A

- Accelerating voltage 175, 178, 180
Acid-dissociated *Escherichia coli* ferritin (EcFtnA) 212–214, 216
Affinity chromatography 17, 76, 159, 408, 411
Alcohol dehydrogenase D (AdhD) 114, 119
Aldehyde-modified carbon grid 100, 104
Analytical ultracentrifugation sedimentation analyses (AUC) 74
Anisotropic network model (ANM) 308, 311–313
Antigen and adjuvant conjugation 322–325, 331
Antigen presenting cells (APCs) 321, 322
Anti-tumor immunity 330
Artificial protein 89, 95
Artificial protein cage 50

B

- Bacterial fermentation 38
Beamline setup 212, 213
 β -propiolactone (β PL) 258, 259
Binary protein crystal 381

C

- Cage like micellar structure 97, 100, 104
Cancer 70, 322, 350
Cancer vaccine 321–332
Catalytic material 111
Cellular uptake 339
Charge complimentary nanoenvironments 403–417
Charge transport (CT) 14, 241–253
Chemical conjugation 259, 262, 263
Chimera 4, 6, 9, 31, 71, 72, 75, 190, 197, 206
Circular dichroism spectroscopy 23–26, 104
Cloning 5, 7–11, 32, 34, 51, 91, 114, 158–162, 388, 409
Coarse-grained model 307–317
Coiled-coil protein origami (CCPO) 3–47
Colony PCR 10, 11, 13, 162
Complex protein 403–417

- Computational design 362, 366–369
Confocal microscopy 339
Contrast agent 175, 350, 351, 356, 358, 359
Contrast transfer function (CTF) 180, 182–184, 186, 190, 196, 203
Controlled release 147, 148
Cowpea mosaic virus 258, 388, 389
Cryo-electron microscopy (cryo-EM) 62, 63, 173–177, 180–182, 187, 188, 190, 193, 195, 196, 200, 203, 308
Cryo-TEM 62
Crystalline biohybrid materials 361–385
CTL lysis assay 325
Cytochrome b_{562} 95–103, 105
Cytotoxic T cells 322, 329

D

- Defocus 176, 179, 184, 190, 195–197, 199, 206
Denaturation 17, 24–27, 36, 44, 81, 86, 123, 151, 154, 228, 374, 376, 414
De novo protein 80–82, 84–86, 89
Density gradient 393
Desalting 99, 103, 104, 125, 323–325, 340, 352, 396, 400
Detector 15, 23, 25, 42, 43, 45, 74, 82, 87, 174, 177–180, 201, 203, 211–214, 221, 229, 235, 381, 400
Dialysis 40, 43, 65, 103, 114–116, 138, 140, 141, 143, 149, 150, 153, 154, 165, 411, 417
Differential centrifugation 394–396
Disassembly/reassembly method 335, 337
Disulfide bond 71, 112, 116, 118, 148, 153, 412
Doxorubicin hydrochloride (DOX) 335, 337, 338, 340, 342–346
Droplet 224–226, 229, 234, 380, 385
Drug carrier 79, 335
Drug delivery 136, 147, 157, 219, 257, 273, 307, 338
Drug encapsulation 158
Drug-protein complex 343–344

- Dual surface modification 148
 Dynamic light scattering (DLS) 21, 54, 60,
 61, 74, 77, 104, 106, 166, 220, 222, 224–225,
 229, 232, 258, 265, 325, 337, 340, 342, 343,
 345, 390, 397, 400
- E**
- EGaIn-method 251
 Eigenmodes 309, 313–317
 Electron dose 179
 Electrophoresis 10, 11, 59,
 74, 118, 124, 127, 128, 132, 138, 154, 159, 258,
 263, 266–267, 390, 399
 Emulsion stabilizer 219
 Encapsulin 49, 157–168
 Energy minimization 277, 283,
 285, 298, 367
 Engineered hemoprotein 95–106
 Enzyme encapsulation 111, 113, 114
 E2 protein cage 222, 226, 228,
 229, 234, 274
- F**
- Fabrication of micropore based junction 248–251
 Fabrication of electrode 243
 Fabrication of ferritin monolayer 246
 Fast protein liquid chromatography (FPLC) 15,
 19, 20, 39, 52, 56, 57, 159, 160, 163, 167, 260,
 265, 362, 373, 390, 396
 Ferritin assembly reaction 216
 Ferritin cage 69–77, 122,
 126, 130, 132, 135–144, 336, 350, 351, 353,
 356, 362, 367, 368, 377–379
 Ferritin metal conjugate 137, 140–142
 Ferrozine assay 123
 FGG tripeptide 97
 Fluorophore conjugation 259
 Force field 276, 281, 282, 295
 Formalin 258, 259, 262
- G**
- Gaussian network model (GNM) 308, 311
 Genetic modification 111
 Glycoprotein 100 (gp100) 322–326, 329, 330
 Gold coordination 49–66
 Golden gate assembly (GGA) 10–13, 33, 34
 Grid preparation 63, 129, 178
 GROMACS 274, 279–295
- H**
- Hanging drop 72, 75, 380
 Heme-azobenzene conjugate 96
 Heterogeneous protein expression 177
- Hexameric tyrosine 96
 Hexameric tyrosine coordinated heme protein
 (HTHP) 96–103, 105
 Highly ordered nanoparticles 379
- I**
- Image processing 176, 177, 180–200
 Immobilization 137, 139, 140
 Immune response 321–332
 Inclusion bodies 403, 410–412, 417
 Inductively coupled plasma mass spectrometer
 (ICP-MS) 132, 352, 353, 356, 359
In planta infection assay 258
 Interferon-gamma (IFN- γ) ELISpot 323–324
 In vitro protein folding 404
 Ion exchange chromatography 20–21,
 39, 56, 131, 373, 375
 Iron accumulation 121–132
 Iron loading 122, 123, 351, 352, 378
 Iron mineralization 126, 131
 Iron staining 124, 127
- L**
- Linker 8, 10, 32, 38,
 39, 41, 80, 81, 84–86, 96, 98, 99, 103, 105, 106,
 112–114, 117, 118, 158, 245, 324, 325, 331
 Liquid-liquid interface 219–222, 226, 227
 Lysis of bacterial cells 16–18
- M**
- Magnetic resonance imaging (MRI) 123,
 349–351, 353, 355–357, 359
 Magnification 178, 179, 188,
 195, 201, 204, 265
 Maleimide 96, 97, 106, 147, 148, 154
 Metal complex reaction 336
 Metal-fixation 135–144
 Metal oxide nanoparticle 362, 364–365,
 376–379
 Mineral core 123, 129, 130
 Molecular dynamics (MD) simulation 273–276,
 278–303
 Molecular size filter effect 148
 Motion correction 180–182, 202
 Multiscale virtual particle based anisotropic network
 model (MVP-ANM) 308, 309, 311–316
- N**
- NAMD 274, 295–303
 Nanocage 4, 69–71, 73,
 75, 79, 122, 123, 126–129, 131, 147–149, 151,
 152, 157, 164, 279, 336, 338, 346, 350, 366,
 376, 384

- Nanoencapsulation.....404–406, 417
 Nanomedicine307, 360
 Nanoparticle synthesis136, 273, 377, 378
 Nanoreactor123, 157, 400
 Nanostructure4, 50, 79–91, 343
 Ni^{2+} -NTA affinity chromatography408, 411
 Non-denaturing polyacrylamide gel electrophoresis
 (native-PAGE)72, 74, 76,
 77, 123, 124, 127
 Normal mode analysis (NMA)308, 309, 313
 Nuclear magnetic resonance (NMR)174, 177,
 221, 275, 350–355, 359
- O**
- Oil-water interface220, 226, 227
 Ordered array112, 113
- P**
- Particle analyzer160, 166
 P22 bacteriophage111
 Photon82, 87, 212, 216
 Photoresist242, 243, 247, 249
 Pickering emulsion219, 222–224, 234
 Plant-based expression391
 Plant growth388
 Plant virus257, 400
 Potato virus X (PVX)257–269
 Protein array112
 Protein cage49–66, 70, 72,
 79–91, 111, 112, 126, 128, 129, 135, 143, 147–
 154, 157, 173–207, 211–217, 219–224, 226,
 227, 229, 232–234, 273–302, 307, 350, 355,
 356, 361, 362, 366, 376, 378, 379, 388, 400
 Protein cages relaxivity349–360
 Protein complex49, 79, 87, 88
 Protein crystallization75, 362, 365, 379
 Protein design72, 76, 89
 Protein 3D reconstruction188–192
 Protein expression55–57, 71,
 73, 81, 84, 86, 91, 388, 393, 403, 410
 Protein interface redesign72–73
 Protein macromolecular framework (PMF)111–119
 Protein nanobuilding block (PN-Block)79–91
 Protein nanoparticle137, 148, 321,
 322, 325, 404
 Protein refolding87, 403, 404
 Protonation state276, 279, 281, 282, 295
 Prussian blue124, 127, 128, 132
- R**
- Radian distribution function278, 293–294
 Radiation damage28, 175,
 179, 214, 217
 Radius of gyration30, 88, 91,
 214, 215, 229, 278, 396
 Reassembly design69–77
 Refolding17, 27, 44, 81,
 86, 87, 216, 403, 404, 408, 411, 413, 417
 Relaxation measurement353–357
 RELION176, 177, 180, 181,
 183–193, 195, 196, 199–201, 203–207
 Resazurin (Alamar Blue) assay61–62
 Reverse transcription polymerase chain reaction
 (RT-PCR)258, 261, 267
 Reversible assembly123, 212, 215
 Rigid-body modelling88–90
 RMSF278, 292, 293
 Root mean square deviation (RMSD)277,
 278, 292
 Rosetta71, 73, 76, 159,
 161, 362, 366–371, 383, 384
- S**
- Scattering25, 82, 87,
 88, 91, 119, 211–216, 220, 221, 225, 229, 230,
 232, 233, 275, 380–383, 396, 398
 Self-assembled monolayer (SAM)243,
 245–248, 251–253
 Self-assembly57–58, 73, 80,
 81, 84, 122, 123, 136, 242, 274, 279, 361
 Single particle analysis (SPA)174–176,
 178–181
 Size exclusion chromatography (SEC)11, 16,
 21, 23–25, 40, 41, 46, 52, 53, 55–59, 65, 82, 85,
 87, 96, 160, 164, 165, 258, 265, 337, 342, 343,
 345, 373, 376, 377, 379, 380, 390, 391, 393,
 396, 397, 408, 410, 412, 414, 417
 Small angle X-ray scattering (SAXS)23, 25–31,
 43, 46, 47, 82, 87–91, 113, 211, 212, 220, 221,
 223, 227–233, 362, 380–383
 Solvation276, 282–283
 Spring parameter311–313
 SPring-8 beamline213, 216
 Stokes diameter152
 Stopped-flow mixing211–214, 216
 Strep-tag affinity chromatography15, 19–20
 Sub-tomogram averaging (STA)174, 175
 Sucrose gradient centrifugation378
 Superlattice361, 362, 379
 Superstructure380
 Supramolecular assembly97
 Surface-charged variant366, 371
 Surface functionalization158
 Surface pore148, 151
 Synthetic heme96, 99, 100, 103, 105
 Synthetic metal complex136, 137

T

- Targeting drug delivery 158
Template-stripping (TS) 242, 243, 245, 248, 251
Thermostable exoshells (tES) 404
Time-resolved small-angle X-ray scattering (TR-SAXS) 211, 212, 215, 216
TIP60 protein 80, 148, 150–154
Tissue disruption 389
Tobacco mild green mosaic virus 258
Tomography 174, 175
Trajectory 277, 278, 290, 292, 295, 302, 368
Transformation 12, 28, 34, 35, 88, 91, 114–116, 118, 161, 166, 167, 204, 220, 225, 232, 312, 372, 387, 392
Transient expression 387–389
Transmission electron microscopy (TEM) 59, 60, 74, 100, 104, 123, 125, 129, 130, 132, 166, 258, 265, 337, 340, 342, 343, 345, 390, 398–400
TRAP-protein cage 51, 53, 55, 57–62, 65, 66
Trypan blue stain assay 60

- Tryptophan fluorescence (TF) 220–222, 226, 227
T1/T2 relaxation 350, 353–356, 359
Tumor-associated antigens (TAAs) 329, 332
Tumor challenge 322, 328, 330
Tunnel junction 241

U

- UV irradiation 257, 258, 261, 265

V

- Vacuum infiltration 389, 391, 392
Vault cage 307–317
Virus inactivation 257–269
Virus-like particle (VLP) 111–119, 147, 387–390, 393, 395–400

W

- Water protein relaxation 353–355

Z

- Zeta potential 337, 345, 406