Title

Design and Functionalization of Natural and Synthetic Protein Nanocages

Author

Miller, Justin Evan

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Design and Functionalization of Natural and Synthetic Protein Nanocages

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular Biology

by

Justin Evan Miller

2023

ABSTRACT OF THE DISSERTATION

Design and Functionalization of Natural and Synthetic Protein Nanocages

by

Justin Evan Miller

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2023

Professor Todd O. Yeates, Chair

Biological containers known as protein cages are widespread in nature—displaying unique functionality via their ability to compartmentalize molecules in an internal lumen volume. Composed of repeated copies of a small number of protein subunits arranged symmetrically, protein cages are relatively simple in composition, and yet complex in structure. In nature these assemblies function in transport and protection, storage of important molecular cargo, and sequestration of metabolic reactions. These demonstrations of incredible functionality have inspired researchers seeking to engineer complex protein architectures with diverse functionality to direct their attention towards symmetric protein cages. Resulting from these efforts, a modest number of protein engineering studies have succeeded in replicating these structures *de novo* in hopes of applying them to the fields of medicine and biotechnology; however, developing applications for these assemblies that outperform existing technologies remains a challenging exercise. The primary aims of my thesis work are the design and development of applications for symmetric protein cages with a particular focus on mimicking natural functionalities not yet

recreated via artificial protein design methods. We note the limitations of prevalent protein design methods, and technological advances that may revolutionize the field.

The first chapter of this thesis is a brief introduction to protein cages as platforms for engineering, as well as efforts to design them *de novo* in the laboratory.

In the second chapter of this thesis we sought to advance on existing methods for protein cage engineering by designing new cages with novel interface construction. These cages incorporate a grafted interface onto each of two components making up the cages. In this work we describe challenges associated with direct rigid linkage of protein components via alpha helical extension and discuss whether the technique is a viable strategy for future cage design efforts.

Next we attempt a novel approach to addressing a long-standing challenge in the field of protein design: engineering structurally responsive protein cages. The natural world holds numerous examples of protein cages that can assemble and disassemble in response to specific stimuli, but recreating this effect in the laboratory has proven challenging. In chapter three we present work on the design of protein cages that disassemble in response to a specifically chosen target protease. This work related to protein cage disassembly has implications in therapeutic design, as proteases are common therapeutic targets, and developing systems responsive to diseases biomarkers could be of great utility.

In the next chapter we expand our focus on cage disassembly to other types of triggers for cage disassembly. In this work, we devise a system utilizing antibody-mimetic molecules called DARPins that allows for the design of protein cages that disassemble in response to any protein of interest. The work in this chapter focuses on a natural cage assembly (sulfur oxygenase) as a platform for development of this effect.

The fifth chapter of this thesis covers work related to designing protein-cage based nanoparticles to be used as immunosorbents for patients on dialysis. The work addresses a clinical need for methods to remove the protein beta-2 microglobulin from the blood, as high concentrations are common in patients on long-term hemodialysis therapy, but can be toxic. We describe a protein cage that displays nanobodies on its exterior surface to generate beta-2 microglobulin adsorbent cages (BACs).

In conclusion, the research performed in this dissertation work seeks to advance on existing methods for protein cage engineering by designing new cages with novel interface construction. I then used previously described artificial protein cages to demonstrate triggerable protein cages with dynamic assembly properties that are responsive to molecular stimuli, as well as developing of an immunosorbent nanoparticle with clinical applications.

The dissertation of Justin Evan Miller is approved.

Yvonne Y. Chen

Pascal F. Egea

William M. Gelbart

Feng Guo

Lin Jiang

Todd O. Yeates, Committee Chair

University of California, Los Angeles

2023

This dissertation is dedicated to my family and friends who pushed me forward and made this possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

APPENDIX TWO

LIST OF TABLES

CHAPTER TWO

CHAPTER THREE

ACKNOWLEDGEMENTS

Without the support of a huge number of people I would never have been able to complete the work required for this thesis. I would like to first thank my scientific mentors including my PhD advisor Todd Yeates for inspiring me to grow to be a better scientist and in so many other ways. I am extremely grateful to him for taking me on as an undergraduate researcher in his lab, and more importantly for allowing me to stick around for so many years. I also need to note my appreciation for Todd's willingness to allow me to pursue my interests out of the lab; allowing me to take a two-year internship with the UCLA Technology Development Group to explore my interests in business and IP law was incredibly helpful to developing a career trajectory, and I'm so thankful Todd supported me on this path in lab by including me in the patenting process on much of my thesis work. Never was there any question that Todd had my interest first, and I wouldn't be where I am today without that support.

I would like to thank other scientific mentors including Dr. Jeffrey H. Miller for sparking my interest in research as an undergraduate student. Working in his research group and under his careful advisement was immensely important to shaping my work ethic and sense of ownership over my experiments. I am truly appreciative of his generosity and mentorship during this time. I'd like to thank Dr. Pascal Egea for immense support with all things graduate school and research. Any time I was stuck with an experiment, his support was available, even when it was just caffeine and encouragement. I'd also like to thank my other thesis committee members of: Dr. Feng Guo, Dr. Yvonne Chen, Dr. William Gelbart, and Dr. Lin Jiang for inspiring my research creativity in diverse areas of science, and for pressing me to think about my experiments from unique perspectives. I should thank Dr. Mark Arbing for his seemingly limitless knowledge and patience, which helped me so much especially in the final year of my dissertation research. I

would like to denote my appreciation of other influential mentors: Dr. Mike Sawaya and Dr. Duilio Cascio for their incredible knowledge and teaching abilities, without whom I wouldn't be the scientist I am today.

I would also like to expressly thank my many peers and mentors that I have overlapped with in the Yeates laboratory: Dr. Joshua Laniado, Dr. Kevin Cannon, Dr. Jessica Ochoa, Dr. Yuxi Liu, Dr. Roger Castells-Graells, Kyle Meador, Matthew Agdanowski, Eric Lee, Roman Aguirre, Brian Nguyen, Angela Agnew, Charlize Espinoza, and Inna Antselovich for support and guidance during my time working in the laboratory.

I need to also thank those friends that mean the world to me and tempered and supported me during my academic journey. Particularly, some who have offered incredible support along the way: Jacqueline, Victoria, Liz, Will, Kat, Yuki, Sean, Joe, Yerim, Sarah, Mike and many others made my time outside of lab so much fun.

Finally, I'd like to thank my family including my mother and father who continually inspire me to great things, and without whom none of this would be possible, my siblings: Jordan and Josh for being the best role models. I'd like to thank my grandparents who inspire me to "work hard but have fun", and that life is best enjoyed with those you love.

…

Chapter 1 is an introduction to application-driven design of protein nanocages.

Chapter 2 is a version of Laniado J*, Cannon KA*, Miller JE*, Sawaya MR, McNamara DE, Yeates TO. Geometric Lessons and Design Strategies for Nanoscale Protein Cages. ACS Nano. 2021 Mar 23;15(3):4277-4286. doi: 10.1021/acsnano.0c07167. Laniado, Miller, Yeates contributed to experimental planning. Cannon contributed to electron microscopy data processing. Laniado, Cannon, Miller, Yeates contributed to writing the manuscript.

Chapter 3 is a version of Miller JE, Srinivasan Y, Dharmaraj NP, Liu A, Nguyen PL, Taylor SD, Yeates TO. Designing Protease-Triggered Protein Cages. J Am Chem Soc. 2022 Jul 20;144(28):12681-12689. doi: 10.1021/jacs.2c02165. Miller and Yeates contributed to experimental planning. Miller et al., contributed to experimentation. Miller and Yeates contributed to writing the manuscript.

Chapter 4 is a summary of my work on a project led by Eric J. Lee and an abridged version of a manuscript in preparation for submission: Lee EJ, Miller JE, Gladkov N, Yeates TO. Design and synthesis of LOCs: a novel, modular nanocage platform for targeted moiety release.

Chapter 5 is a version of the manuscript: Miller JE, Castells-Graells R, Arbing MA, Munoz A, Nguyen B, Espinoza C, Moroz P, Yeates TO. Design of Beta2 Microglobulin Adsorbent Protein Cages.

Appendix 1 is a summary of results from a project titled Structural Investigation of a Protein of Unknown Structure and Function reveals Remaining Challenges of MicroED for Novel Proteins.

Appendix 2 is a version of Lantz C, Wei B, Zhao B, Jung W, Goring AK, Le J, Miller J, Loo RRO, Loo JA. Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order Structure Information for Protein Complexes. J Am Chem Soc. 2022 Dec 7;144(48):21826-21830. doi: 10.1021/jacs.2c06726 with a preface describing the work performed in greater detail.

This work was supported in part by NIH T32GM067555 Biotechnology Training in Biomedical Sciences and Engineering fellowship, and by the UCLA Disseratation Year Fellowship to JEM.

Vita

[Education]

University of California, Los Angeles,
Ph.D. Candidate in Molecular Biology 2017-

University of California, Los Angeles,
B.Sc. in Microbiology, Immunology, and Molecular Genetics, with Honors, *2013-2017*

[Professional and Academic Experience]

Graduate Student Researcher
University of California, Los Angeles, Molecular Biology Institute. 2017-2023
Advisor: Prof. Todd O. Yeates

- Design and structural characterization of synthetic protein cages
- Development of protein cages with controlled disassembly
- Design of beta$_2$ microglobulin binding protein cages
- Identification and characterization of novel bacterial supramolecular assemblies
- Cryo-electron tomography of native propanediol utilization microcompartments

Senior Business Development Fellow
University of California, Los Angeles Technology Development Group (TDG), Aug. 2021-Sept. 2022

- Mentored team of 5 technology fellows enrolled in life science Ph.D. programs on production of marketing materials for out-licensing UCLA inventions
- Collaborated on marketing strategy for immune-oncology invention portfolio
- Acted as scientific advisor on outreach to outside investors and corporate licensees
- Read and interpreted patents for in-house business officers

Business Development Fellow
University of California, Los Angeles Technology Development Group (TDG), Jun. 2020-Aug. 2021

- Produced marketing materials for UCLA life science intellectual property:
    - 50+ Non-confidential invention descriptions
    - 50+ Targeted marketing lists
    - 15+ Marketing pitch decks

Undergraduate Research Assistant
University of California, Los Angeles, Microbiology, Immunology, and Molecular Genetics. 2015-2016
Advisor: Prof. Jeffrey H. Miller

- Investigated mutagenic synergies of chemotherapeutics and antibiotics in bacteria models

[Honors and Awards]

UCLA Graduate Division Dissertation Year Fellowship, 2022
UCLA Biotechnology Training Fellowship 2018-2020

[Selected Publications]

Miller J.E., Castells-Graells R, Arbing M.A., Munoz A., Jiang Y.X., Espinoza C.T., Nguyen B., Moroz P., Yeates T.O. Design of Beta-2 Microglobulin Adsorbent Protein Nanoparticles. *(Submitted)*

Lee E.J., Gladkov N., Miller J.E., Yeates T.O. Design of Ligand-Operable Protein-Cages that Open Upon Specific Protein Binding. *(Submitted)*

Lantz C, Wei B, Zhao B, Jung W, Goring AK, Le J, **Miller J,** Loo RRO, Loo JA. Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order

Structure Information for Protein Complexes. J Am Chem Soc. 2022 Dec 7;144(48):21826-21830. doi: 10.1021/jacs.2c06726.

Miller JE, Srinivasan Y, Dharmaraj NP, Liu A, Nguyen PL, Taylor SD, Yeates TO. Designing Protease-Triggered Protein Cages. J Am Chem Soc. 2022 Jul 20;144(28):12681-12689. doi: 10.1021/jacs.2c02165.

Laniado J*, Cannon KA*, **Miller JE***, Sawaya MR, McNamara DE, Yeates TO. Geometric Lessons and Design Strategies for Nanoscale Protein Cages. ACS Nano. 2021 Mar 23;15(3):4277-4286. doi: 10.1021/acsnano.0c07167.

D'Souza S, **Miller JE**, Ahn J, Subandi R, Lozano D, Ramirez J, Goff M, Davidian C, Miller JH. The Antibiotic Trimethoprim Displays Strong Mutagenic Synergy with 2-Aminopurine. Antimicrob Agents Chemother. 2019 Jan 29;63(2):e01577-18. doi: 10.1128/AAC.01577-18.

[Patents]

“Self Assembling Protein Nanoparticles as Carrier Molecules” (2021). U.S. Patent Application No. 17/002,571 (pending)

“Ligand Operable Protein Cages” (2022). U.S. Patent Application No. PCT/US2023/062186 (pending)

CHAPTER ONE

INTRODUCTION: PROTEIN CAGES: DEFINING, CHARACTERIZING, AND APPLYING

1.1 Main Text

Nanoscale protein compartments called protein cages are widespread in nature. Fulfilling requisition functions for life through their ability to compartmentalize molecules inside their lumen volume. They are found ubiquitously in nature and evolved diverse functions including transport, storage, and sequestration that make use of these structures' capacity for compartmentalization. Understandably, much effort in the field of protein engineering has gone towards engineering novel applications for these assemblies. Examples of these efforts have included developing protein cages as drug delivery vehicles, vaccine platforms, imaging scaffolds, MRI-visible imaging agents, enzymatic display, and other purposes with incredible success. In this chapter we describe the structure and composition of protein cages, then review efforts by researchers to replicate these structures artificially using protein design, and finally describe efforts to apply these structures to applications in biotechnology.

Protein cages are supramolecular assemblies composed of repeated copies of one or more distinct protein chain arranged symmetrically. Despite incredible variation in the size, shape, and number of components of these structures, some commonalities underscore them. Many of the protein cages found in nature can be classified as assembling into the form of one of the Platonic solids: tetrahedron, octahedron, or icosahedron. Structures that assemble into forms that resemble these shapes exhibit remarkable structural complexity and yet these arrangements allow for only a few unique protein chains to compose the entire structure due to the symmetric relationships between components. This fact not only limits the genetic burden on host organisms, but also makes them attractive and efficient design targets.

In nature, many examples of functional protein cages have been identified and characterized. Viruses, as the most abundant biological entities on earth, are the most well-

known example of a protein cage. These infectious particles are composed of protein capsids that assemble around and protect a nucleic acid genome and then disassemble during infection. They benefit extensively from the high symmetry of their capsid structures. With limited space for a genome, encoding many structural capsid components is impossible. Other examples of protein cages found in nature include bacterial nanocompartments which sequester enzymatic reactions to improve flux and reduce toxicity of intermediates[1,2], ferritins which store and subsequently release mineralized iron when needed[3,4], and Arc proteins thought to be involved in cell to cell signaling in the nervous system[5]. All of these examples demonstrate function because of their ability to encapsulate molecules when needed and uncoat in response to other triggers. These few examples serve to demonstrate the remarkable potential of protein cages in developing functional molecular assemblies in the laboratory.

Protein Cage Design

Protein design methodology has rapidly advanced from nascent studies designing single alpha helices, to the more recent application of machine learning in structure prediction and sequence design. Still, as design targets increases in complexity, success rates in generating validated designs decrease. Even so, much progress has been made in the field of protein engineering, enabling the generation of two-dimensional fractals and layers, three-dimensional crystals, 1D filaments, and finite assemblies[6]. Inspired by the rich functionality seen in nature, researchers have sought to engineer bespoke protein cages with characteristics tailored to a desired application. The most common way researchers have undertaken this feat is by starting from naturally symmetric proteins (e.g. cyclically symmetric homo-oligomeric proteins) as building blocks, and bringing these components together with the correct geometries for cage assembly[6]. The principles for designing novel protein cages, along with the first experimental

demonstration, were described by Padilla, Colovos and Yeates in 2001 (ref). The earliest examples of protein cage design targeted a tetrahedral arrangement of proteins using two distinct design methodologies. One, from Padilla and Lai and colleagues solved the problem of bringing two symmetric proteins together in space by linking the two components directly via an alpha-helical linker[7-8]. The other example made use of computation interface design to introduce mutations to the native protein sequences such that the new proteins form a stable interface between the oligomers[9]. Despite the early success of the genetic-fusion approach to cage design, replicating the same feat became challenging, likely due to inherent flexibility in the alpha helices used to link the two proteins together[10,11]. As a result, interface design has become the prevalent technique to design protein cages over the past decade, and has benefited the most from computational advances in protein design software[12,13]. Our work in chapter two of this thesis seeks to expand on existing methods for interface design with a new strategy that makes use of a heterodimeric coiled-coil. In this strategy, each of the helices in the coiled-coil is attached to the proteins in one of the two oligomers making up the cage. The result is similar in technique to the direct genetic fusion demonstrated by Padilla, Lai, Yeates et al., but we hoped would offer improvements in rigidity.

As demonstrated by these design works and others, a modest number of protein engineering studies have succeeded in replicating protein cage structures *de novo* in hopes of applying them to the fields of medicine and biotechnology, and researchers have had success in developing notable applications for these assemblies. Some exciting examples of this include the attachment of enzymes on protein cages for improved flux through metabolic pathways[14], display of small proteins via adapters to enable imaging using cryo-EM[15], antigen display for vaccines[16], amongst many others. We add to this list with an additional application demonstrated

in this thesis: the design of protein cages as nanoparticles for the adsorption of proteins from blood during dialysis. As far as we know this is the first use of protein cages as a dialysis-related device.

While many of these developments have largely made use of intact static assemblies displaying functional moieties for functionalization, they have neglected to reproduce the controlled assembly and reversible encapsulation that makes natural protein cages such effective tools for compartmentalization. While engineering efforts have been able to design cages that display controlled assembly, many of these studies exploit bulk changes in environmental conditions to achieve this structural rearrangement (e.g. pH, ions, redox, or light)[17–23]. To expand on these feats of engineering, in this thesis work we present methods for designing triggerable protein cages that disassemble on-command in response to two different classes of stimuli: proteases cleavage, and protein binding. This enables applications based on principles more reminiscent of biological systems, and potentially useful for applications that require sensitivity to changes in protein expression or activity (Figure 1.1).

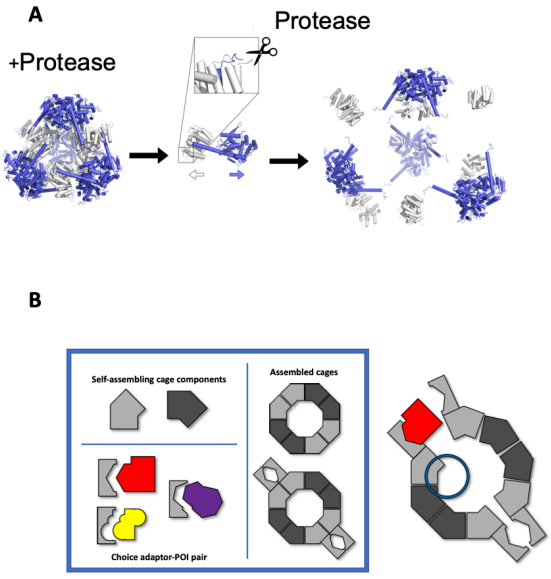


Figure 1. Conceptualization of protein-triggered protein cage design. (A) Protease triggered protein cages are cut by a target protease and then disassemble. (B) Concept of ligand-binding based opening of protein nanocages. Self-assembling protein cage components “A” (light grey) and “B” (dark grey) (top left panel); example adaptor-protein pairs (bottom left panel) that mediate protein binding; self-assembled nanocages, without adaptors (right panel, top) and with fused adaptors (right panel, bottom). At right: ligand-bound complex is incompatible with assembled state of nanocage, driving disruption of the assembly.

All of these engineering feats, including those demonstrated in this thesis make use of existing protein nanocages as starting points. The limited number of protein cages that have been successfully designed and characterized to high resolution, remains a major limitation in application-driven engineering efforts based on these principles, however. New software incorporating machine learning into computational design methods have enabled researchers to build proteins of a predetermined structure *de novo*. This has opened the door to generating structures that mimic natural evolution of protein assembly, including nanocages tailored for specific functions, rather than the top-down approach to application-driven protein design described in in this thesis. While these demonstrations represent advances in the field of protein design, improvements could be made via efforts to design applications from the outset (i.e. combining form and function in the design process), and the results of these efforts may ultimately resemble the exciting and rich functionalities we see in nature more closely.

1.2 References

1. Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat. Microbiol.* **2**, 17029 (2017).
2. Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. & Savage, D. F. Encapsulins: molecular biology of the shell. *Crit. Rev. Biochem. Mol. Biol.* **52**, 583–594 (2017).
3. Lawson, D. M. *et al.* Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**, 541–544 (1991).
4. Edwardson, T. G. W. *et al.* Protein Cages: From Fundamentals to Advanced Applications. *Chem. Rev.* **122**, 9145–9197 (2022).
5. Erlendsson, S. *et al.* Structures of virus-like capsids formed by the Drosophila neuronal Arc proteins. *Nat. Neurosci.* **23**, 172–175 (2020).
6. Yeates, T. O., Liu, Y. & Laniado, J. The design of symmetric protein nanomaterials comes of age in theory and practice. *Curr. Opin. Struct. Biol.* **39**, 134–143 (2016).
7. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm Cage Designed by Using Protein Oligomers. *Science* **336**, 1129–1129 (2012).
8. Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments. Proc. Natl. Acad. Sci. U. S. A. 2001, 98 (5), 2217− 2221.
9. King, N. P. *et al.* Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **336**, 1171–1174 (2012).
10. Lai, Y.-T., Jiang, L., Chen, W. & Yeates, T. O. On the predictability of the orientation of protein domains joined by a spanning alpha-helical linker. *Protein Eng. Des. Sel.* **28**, 491–500 (2015).

11. Lai, Y.-T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6**, 1065–1071 (2014).
12. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
13. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
14. McConnell, S. A. *et al.* Designed Protein Cages as Scaffolds for Building Multienzyme Materials. *ACS Synth. Biol.* **9**, 381–391 (2020).
15. Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.* **10**, 1864 (2019).
16. Marcandalli, J. *et al.* Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **176**, 1420-1431.e17 (2019).
17. Dalmau, M., Lim, S. & Wang, S.-W. pH-Triggered Disassembly in a Caged Protein Complex. *Biomacromolecules* **10**, 3199–3206 (2009).
18. Hoersch, D., Roh, S.-H., Chiu, W. & Kortemme, T. Reprogramming an ATP-driven protein machine into a light-gated nanocage. *Nat. Nanotechnol.* **8**, 928–932 (2013).
19. Jones, J. A., Cristie-David, A. S., Andreas, M. P. & Giessen, T. W. Triggered Reversible Disassembly of an Engineered Protein Nanocage. *Angew. Chem. Int. Ed.* **60**, 25034–25041 (2021).
20. Stupka, I. *et al.* Chemically induced protein cage assembly with programmable opening and cargo release. *Sci. Adv.* **8**, eabj9424 (2022).

21. Zang, J. *et al.* Disulfide-mediated conversion of 8-mer bowl-like protein architecture into three different nanocages. *Nat. Commun.* **10**, 778 (2019).

22. Subramanian, R. H. *et al.* Design of metal-mediated protein assemblies via hydroxamic acid functionalities. *Nat. Protoc.* **16**, 3264–3297 (2021).

23. Yang, M. & Song, W. J. Diverse protein assembly driven by metal and chelating amino acids with selectivity and tunability. *Nat. Commun.* **10**, 5545 (2019).

CHAPTER TWO

DESIGNING PROTEIN CAGES USING GRAFTED COILED-COILS AS INTERFACES

2.1 Abstract

Protein molecules bring a rich functionality to the field of designed nanoscale architectures. High-symmetry protein cages are rapidly finding diverse applications in biomedicine, nanotechnology, and imaging, but methods for their reliable and predictable construction remain challenging. In this study we introduce an approach for designing protein assemblies that combines ideas and favorable elements adapted from recent work. Cubically symmetric cages can be created by combining two simpler symmetries, following recently established principles. Here, two different oligomeric protein components are brought together in a geometrically specific arrangement by their separate genetic fusion to individual components of a heterodimeric coiled-coil polypeptide motif of known structure. Fusions between components are made by continuous α-helices to limit flexibility. After a computational design, we tested 10 different protein cage constructions experimentally, two of which formed larger assemblies. One produced the intended octahedral cage, ~26 nm in diameter, while the other appeared to produce the intended tetrahedral cage as a minor component, crystallizing instead in an alternate form representing a collapsed structure of lower stoichiometry and symmetry. Geometric distinctions between the two characterized designs help explain the different degrees of success, leading to clearer principles and improved prospects for the routine creation of nanoscale protein architectures using diverse methods.

2.2 Introduction

In the past decade, developments in the field of protein design have produced a growing suite of bespoke protein assemblies unseen in nature[1–14]. Highly symmetric, cubic, and icosahedral protein cages have drawn particular attention. Early studies laid out the essential design requirements[15]. When two (or more) symmetric objects (e.g., simple protein oligomers) are held

together in a geometrically specific way, a larger and more complex architecture can be produced by self- assembly. The resulting symmetry is dictated by the combined symmetries of the two underlying components. For the design of high-symmetry protein cages, the symmetry axes belonging to the component oligomers must intersect at an angle according to the Platonic solids (i.e., tetrahedron, cube, icosahedron, Figure S2.8). An example would be a homotrimer and a homodimer that are positioned by design such that their individual symmetry axes intersect at an angle of 54.7° at the center of a tetrahedron. While the geometric requirements for assembling diverse protein materials using the principle of symmetry combinations have been studied for nearly 20 years,[2,7,8,15–19] the practical matter of how to engineer a strictly defined geometric relationship between two complex protein molecules remains challenging, relying on advanced molecular engineering and computational design methods.

A successful method for holding two protein oligomers in a rigid orientation with respect to one another was demonstrated in 2001, when Padilla and co-workers used an α-helical linker to bring together a dimer and a trimer in order to form a protein tetrahedron.[15,20] This design strategy utilizes an α- helical linker fused between terminal helices of the component oligomers, creating a continuous intervening helix rigid enough to hold the oligomers in the correct orientation[21]. A reliable application of this strategy tends to be challenged by helix flexibility,[22] but the approach has successfully produced a protein cube[23] and icosahedron[24] in addition to the original tetrahedron.

Other strategies for orienting oligomers to self-assemble into cages have been demonstrated as well. More than a dozen protein cages have been characterized at near-atomic resolution by X-ray crystallography,[17,20,23,25,26] with several others validated by other techniques such as electron microscopy (EM) [3,10–13,27,28]. Most of the designs that have led to crystal

structures were the result of a computational interface design via the Rosetta protein design software[17,25,26]. In this method, natural protein oligomers are first computationally docked such that their symmetry axes have the desired angle of intersection for a specific point-group symmetry (typically cubic T (tetrahedral), O (octahedral), or I (icosahedral)). Then mutations that could lead to a new interface between oligomers in the specified orientation are suggested computationally by searching for amino acid substitutions that give favorable calculated binding energies. Numerous cages have been successfully designed by the interface design approach developed by King et al.,[17] yet such methods are highly demanding computationally, and experimental success rates tend to be limited by an imperfect knowledge of how to produce tight-binding interfaces (which are often relatively hydrophobic) without causing nonspecific assembly into aberrant structures or aggregation into inclusion bodies when overexpressed in bacteria. Recent efforts to lower the hydrophobicity and improve hydrogen bonding in designed interfaces are improving the performance of interface design approaches[29–31]. Other studies have retained the idea of using genetic fusion to connect oligomeric components but have sought to relax the requirement for rigidity imposed by a continuous helical connection. Flexible connections between oligomers have been used to form ordered two-dimensional (2D) protein layers[19] and several types of protein clusters or cages,[32] some with approximate T, O, and I symmetry[10–12]. Methods based on flexible connections have produced assemblies of the intended forms, but they have largely evaded a detailed structural characterization by crystallography or high-resolution cryo-EM. Protein cages designed by various methods are beginning to find applications in areas as diverse as enzyme scaffolding,[33,34] vaccine design,[35] nucleic acid encapsulation,[36,37] and imaging.[38,39] Thus, a further exploration of design methodologies could

facilitate a more routine production of protein cages tailored for applications in biomedicine and nanotechnology.

In the present study, we built on the original concept of joining symmetric proteins based on continuous α-helices but with a variation intended to reduce the innate flexibility of a single α-helix. This approach effectively replaces the single α- helical linker of the original fusion design method with a (presumptively) more rigid heterodimeric α-helical coiled coil; the two components of the coiled coil are fused separately to the two types of oligomeric subunits. Like previous fusion methods, the computational design demands are modest. Similar to methods based on interface design, the resulting architectures are held together by a noncovalent association between proteins, but here that association (i.e., between coiled coil components) is known in advance. We present the structural characterization of two different cages designed by this method, with some unexpected results that demonstrate crucial lessons about protein cage design principles.

2.3 Results and Discussion

Protein Design Methodology

In this work we set out to design three-dimensional (3D) protein cages obeying tetrahedral or octahedral point-group symmetries, which would self- assemble from two different oligomeric protein components via fusion to heterodimeric coiled coils (Figure 2.1). The general design workflow closely follows that of previous helical fusion cage designs, with some variations (see Methods). The ultimate shape of the self-assembled cage construct is determined by the rotational symmetries of the component oligomers and the angle of intersection between their symmetry axes. On the one hand, combinations of trimeric and dimeric components with their symmetry axes intersecting at an angle of 54.7°, for example, should self-assemble with 12

copies of each subunit (a12b12) to form a tetrahedron (Figure 2.1). On the other hand, combinations of tetrameric and trimeric components intersecting at 54.7° should self-assemble with 24 copies of each subunit (a24b24) to build an octahedron (Figure 2.1). Designs bearing T or O symmetry were generated computationally. [Note that a C2+C3 tetrahedron and a C3+C4 octahedron both require the same angle of intersection for their symmetry axes (Figure S2.8), but this does not extend generally to other symmetry combinations.] All possible pairs of candidate dimers and trimers or tetramers and trimers in the PDB protein structure database were considered. For every candidate pair, a subunit from each oligomer was computationally fused through its α-helical terminus to one component of the c-Fos/c-Jun heterodimeric coiled-coil structure. Following that step, candidate constructions where the symmetry axes carried by the two component oligomers intersected at or nearly at the required angle of 54.7° were identified. From this set of ~100 candidates with favorable symmetry axis configurations, a further curation was performed. We screened all candidates for steric clashes that might occur upon symmetry expansion; this sometimes occurs despite correct angular values, for example, if the dimer component is situated too close to the fused trimeric axis, or vice versa. A final visual inspection was performed to check for valid construction, and potentially problematic experimental cases (e.g., membrane proteins) were removed. Other heuristic criteria for selection included favoring cases where the two oligomers were approximately the same distance from the center of assembly. Such cases produce structures resembling cages, while those with components at extremely different radial positions tend to assemble as spiked balls. Ten designs were chosen for experimental characterization.

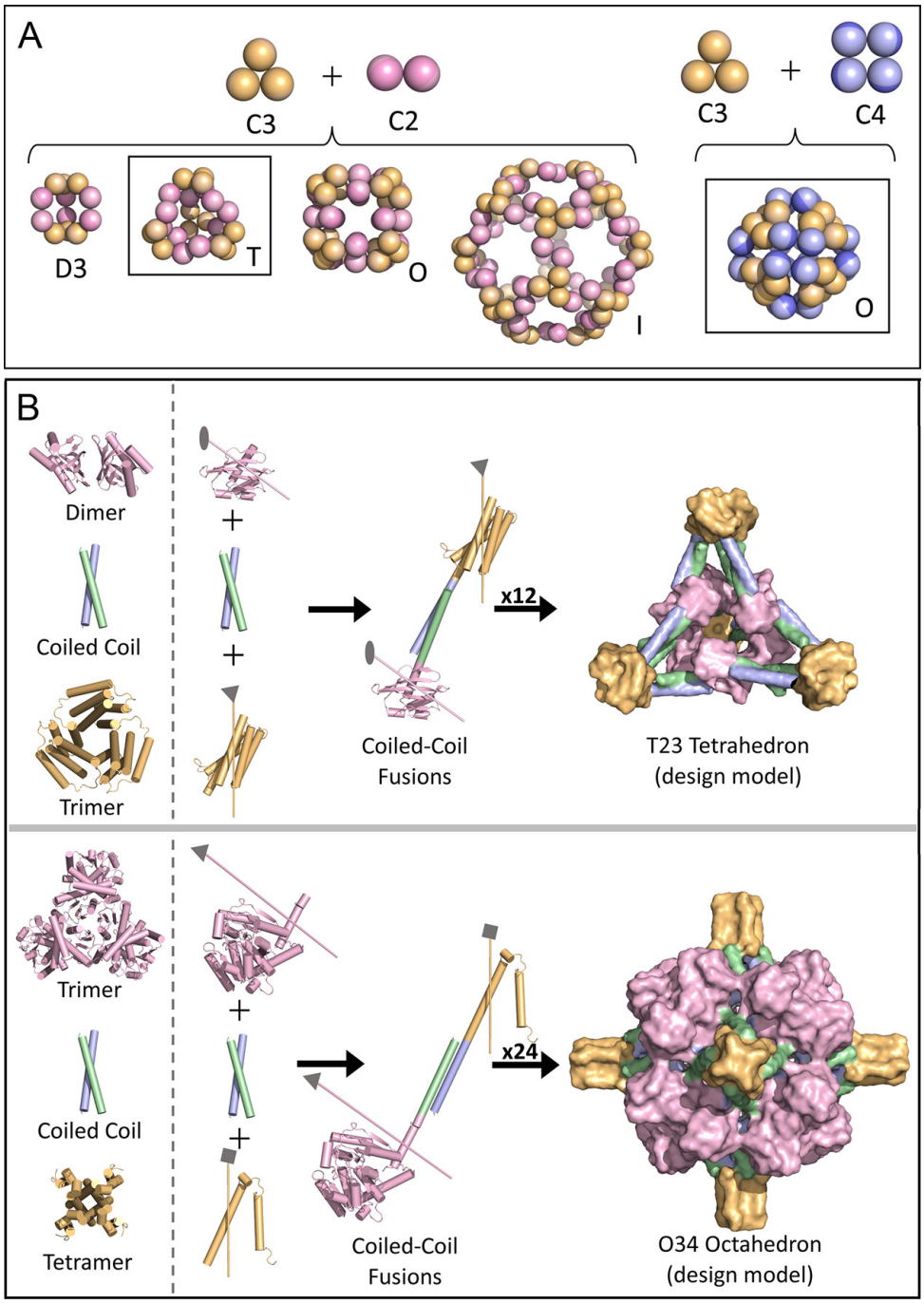


Figure 2.1. Design of two-component protein cages using a coiled-coil helical fusion-based approach. (A) Diagrams showing the possible point symmetry outcomes from combining a dimer and trimer or a trimer and tetramer in specific orientations. The two design types explored in this study—the symmetry T design for the dimer plus trimer and the symmetry O design for the trimer plus tetramer—are boxed. (B) Protein design strategy wherein two distinct oligomeric subunits are fused to separate helices of a heterodimeric coiled-coil linker (blue and green). The

T23 tetrahedral design (top panel) was constructed by combining a C2 dimer (pink) and a C3 trimer (orange). The O34 octahedral design (bottom panel) was constructed by combining a C3 trimer (pink) and a C4 tetramer (orange). A low-resolution surface representation of each design model is shown (top right and bottom right). An oval, a triangle, and a square denote C2, C3, and C4 symmetry axes, respectively. The required angles between symmetry axes are diagrammed in Figure S2.8.

The design approach used here takes advantage of noncovalent interfacial interactions (between the helices of the heterodimeric coiled coil) to hold the two symmetric oligomeric protein components together, as opposed to the genetic fusion or computational interface design strategies used for other cages that have been reported. Coiled-coil motifs have been widely used for protein design, with considerable success and well-characterized principles4,5,10−12,40−45. For our coiled-coil motif we used an engineered version of the c-Fos/c- Jun coiled coil (see the Supporting Information)[41,46,47]. Our coiled-coil interaction serves a similar role as the computationally designed interfaces in King's 2014 work,[25] namely, to bring two different oligomeric components together in a precisely defined fashion. The distinction is that here, instead of requiring a novel interface design for each computational candidate, the noncovalent interaction provided by c-Jun/c- Fos is understood in advance. As with the work of King, the two-component approach allows for a hierarchical assembly (e.g., of complete oligomers) *en route* to the formation of a complete cage, with potential benefits for a robust assembly, compared to approaches where a single polypeptide chain embodies two different oligomerizing domains.

Protein Expression and Characterization of Co-assembly

Synthetic genes encoding the designed proteins were inserted into expression vectors to allow an inducible expression in Escherichia coli (see Methods). Each design included a histidine tag on either the N- or the C-terminus of one of the two protein subunits. For six of the designs,

either one or both of the engineered protein components were not observed in the soluble fraction of the clarified cell lysate on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). For these cases, a low solubility likely arises from a nonspecific aggregation or misfolding caused by the introduction of leucine-rich coiled-coil sequences.

The four remaining designed protein pairs (ccT23−1, ccO34−1, ccO34−3, and ccO23−1) were soluble and coeluted during nickel affinity chromatography and size exclusion chromatography (SEC) (Figure 2.3A, Figure S2.6, and Figure S2.7). Consistent with the target architectures, SEC elution volumes were within the range expected for nanoscale assemblies. Furthermore, a negative stain electron microscopy image of the SEC-purified samples revealed cage-like species in all four cases (Figures 3B, S6, and S7).

Despite these initial results, a further structural characterization of ccO23−1 and ccO34−3 was intractable. The trimeric component of the ccO23−1 design exhibited low levels of soluble expression, and therefore it proved to be infeasible to produce the two-component assembly for subsequent structural studies. Similarly, only a minute fraction of the ccO34−3 sample eluted at the volume expected for the 48-subunit assembly, suggesting that the target design may have only formed as a minor species in solution. This, in addition to issues with sample stability and aggregation, hindered the downstream biophysical analysis of ccO34−3. In contrast, with higher yields and dominant SEC peaks in the expected regions, ccT23−1 and ccO34−1 showed more promise. We therefore directed our attention to further characterize the structures of these two designs.

Crystal Structure of a Collapsed Protein Cage

The ccT23−1 design is a ~20 nm cage with tetrahedral symmetry constructed by combining a C2 dimer and a C3 trimer (Figure 2.1). Its dimeric component was derived from a

putative isomerase of the SnoaL-like family (PDB ID 3DXO, 17.6 kDa monomeric molecular weight), while its trimeric component was derived from the Pyrococcus horikoshii OT3 PH0671 protein (PDB ID 1WY1, 19.5 kDa monomeric molecular weight). When combined with the coiled-coil components, 12 copies of the dimeric subunit and 12 copies of the trimeric subunit would give a total molecular weight of 444 kDa.

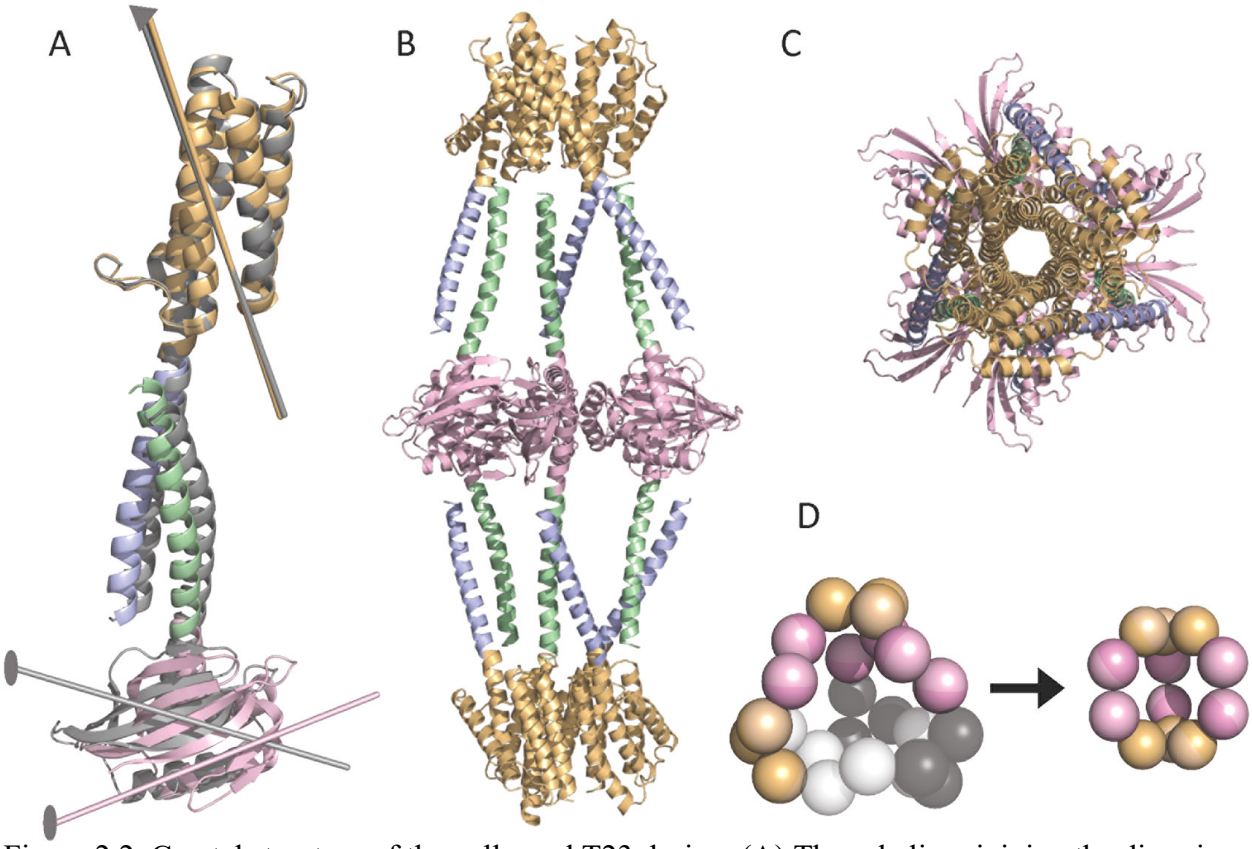


Figure 2.2. Crystal structure of the collapsed T23 design. (A) The α-helices joining the dimeric (pink) and trimeric (orange) subunits in the intended design (gray) are bent in the observed crystal structure (colored), with the symmetry axes of the dimer and trimer shown. The observed crystal structure is an assembly with D3 symmetry, viewed from the side (B) and down its threefold symmetry axis (C). (D) Diagram showing how flexibility can allow a partial assembly intended to be tetrahedral to collapse into a D3 structure. Subunits in the tetrahedral assembly missing from the dihedral assembly are colored in two shades of gray. An oval and a triangle denotes C2 and C3 symmetry axes, respectively. The distribution of deformations between the design and the observed structure is diagrammed in Figure S2.3.

Crystals were obtained by hanging drop vapor diffusion after one to two weeks, and an X-ray diffraction data set was collected with a resolution limit of 4.32 Å (see Methods). Despite the limited diffraction resolution, because the protein components comprising the assembly were all separately known in advance, the structure could be determined by a molecular replacement. On the basis of the cubic space group of the crystals, we initially expected to find the designed tetrahedral cage sitting at a point of symmetry T in the crystal. However, we determined that the space group was in fact P4332, which lacks any site with T symmetry, and only one copy of the a-b heteromeric design could fit in the asymmetric unit. Ultimately, we established that the structure crystallized as an assembly obeying D3 symmetry, sitting at a site in the crystal conforming to that symmetry. Again, cognizant of the limited diffraction resolution, we used omit maps to validate the resulting structure. The entire coiled-coil segment fused to the trimeric component was omitted from phasing, and it appeared clearly in an omit map (Figure S2.1). Ultimately the modest resolution is perhaps unsurprising, given the highly elongated nature of the resulting structure.

Rather than the tetrahedron composed of 12 copies of each subunit that we had designed, the crystal structure obtained from these data revealed a collapsed version of the designed structure, in which only six copies of each subunit (a6b6) instead formed a 222 kDa assembly obeying D3 symmetry (Figure 2.2). Instead of the dimeric and trimeric symmetry axes intersecting at an angle of 54.7°, the coiled-coil linker was bent significantly such that the axes intersected at 90° (Figure 2.2A). The collapsed assembly (renamed ccD3 to accurately reflect its symmetry) resembles a nearly rod-shaped structure ~20.6 nm long and ~10.4 nm wide (Figure 2.2B,C). The coiled-coil linker between oligomers was chosen to reduce flexibility compared to a single helical connection. However, the structural transitions from each oligomer into the

coil−coil do not benefit from the doubled width of the coiled-coil. Indeed, the transition between the dimeric component and the coiled-coil segment appears to be a point where bending was concentrated in the observed structure (Figure S2.3).

Although the crystal structure that was obtained for this design revealed an unintended D3 assembly, a very broad peak in the size exclusion profile that was obtained for the protein in solution suggested a mixture of multiple different assembly species of varying molecular weights (Figure S2.6). Indeed, under a negative stain EM, we observed particles of different sizes in the sample. A 2D classification of manually picked particles revealed a major species with an ∼21 nm diameter, which corresponds closely to the designed tetrahedral cage (Figure S2.5). Although heterogeneity and low particle density prevented us from producing 2D classes that show distinct symmetry elements, many individual particles appear to contain an approximately threefold symmetry (Figure S2.6). Importantly, particle species of higher and lower diameters are also present in the sample; these likely correspond to alternate assembly forms based on different symmetries, such as D3 (as seen in the crystal structure) or even O. Thus, while this cage design case produced a combination of assembled species, apparently including the intended species visible by EM, crystallization selectively favored a smaller, abundant, and possibly better ordered form.

Characterization of a Protein Octahedron

The ccO34−1 design is an ∼26 nm wide octahedron with a large interior cavity ∼10 nm in diameter (Figure 2.1B, bottom). Its protein components were derived from a putrescine carbamoyltransferase trimer (PDB ID 4AM8, 41.6 kDa monomer) and a TraM tetramer (PDB ID 2G7O, 12 kDa monomer). When combined with the coiled-coil components, 24 copies of the trimeric subunit plus 24 copies of the tetrameric subunit would give a total molecular weight of

1.29 MDa. The size of the openings or “windows” joining the interior and exterior space has been characterized for designed and natural protein cages;3 for ccO34−1 the openings are ~30 Å in the narrowest direction.

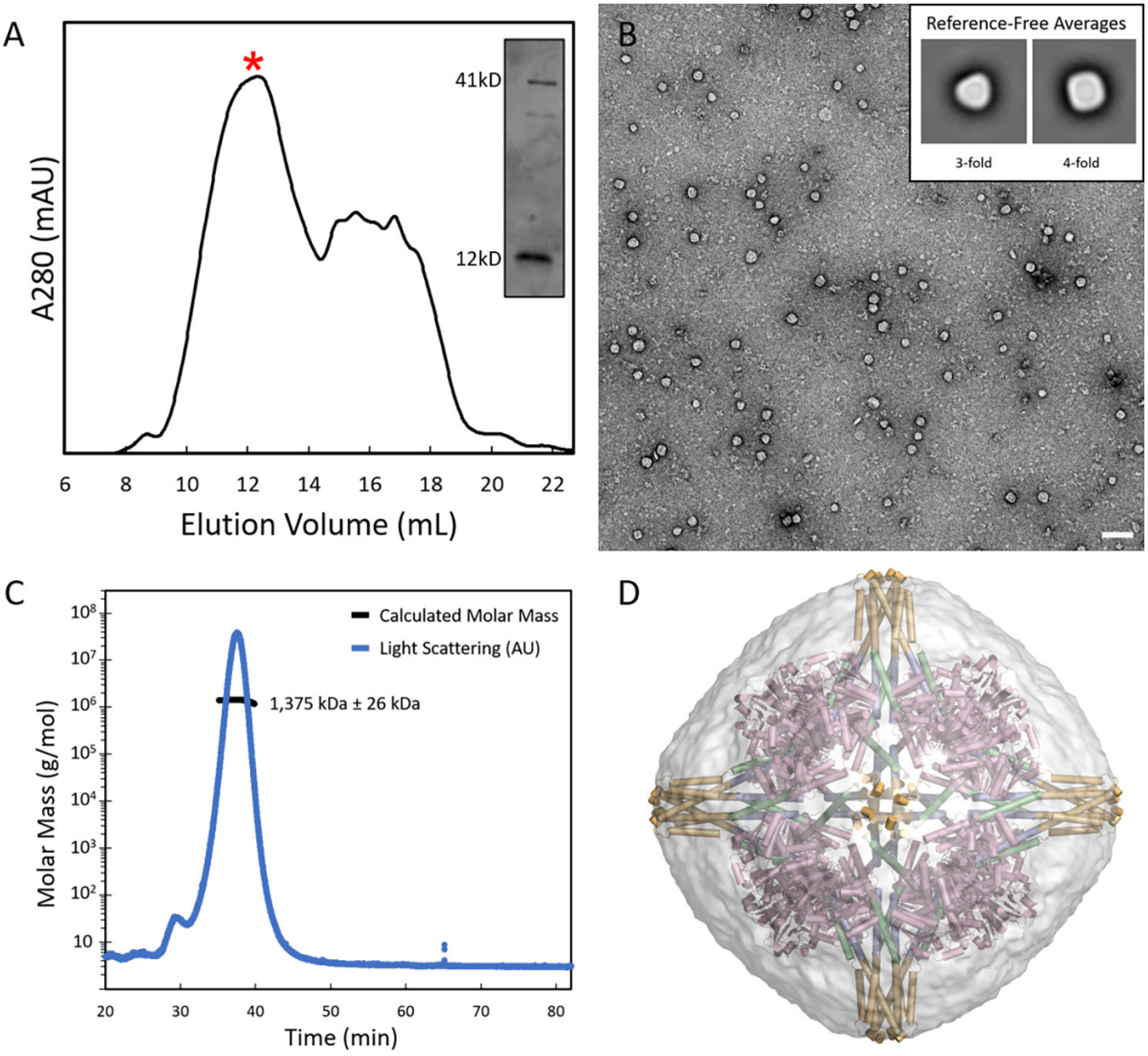


Figure 2.3. Characterization of the ccO34–1 octahedral cage. (A) Purification by SEC produced a major peak (red asterisk) corresponding to a high-molecular weight oligomeric species, comprising both protein cage components, as confirmed by SDS-PAGE (inset). (B) Representative negative stain EM micrograph of the ccO34–1 construct. Scale bar = 100 nm. Selected reference-free class averages of ccO34–1 particles, showing distinct views down the threefold and fourfold axes of the octahedral point-group symmetry, are shown (inset). (C) SEC-MALS of purified cage sample confirmed that the observed species from the major SEC peak has a very close estimated molecular weight to the designed cage. (D) The design model of the

ccO34–1 cage fit into a 3D reconstruction produced by a homogeneous refinement from the negatively stained particles.

Particles ~25 nm in size were readily observed under a negative stain EM, and reference-free 2D class averages revealed several templates corresponding to views down the threefold and fourfold axes of an octahedral assembly (Figure 2.3B). Despite these promising preliminary results, the ccO34−1 cage proved resistant to freezing attempts for a cryo-EM grid preparation, and crystallization trials did not yield protein crystals suitable for X-ray diffraction experiments. Still, a deeper analysis of the cage particles on negatively stained grids provided strong corroborating evidence for the formation of the intended octahedral cage, albeit with a degree of flexibility likely responsible for preventing crystal formation.

From a negative stain EM data set containing ~3500 particles, a 3D reconstruction of a low-resolution electron density map of the structure was pursued. An ab initio 3D reconstruction from these particles performed in cryoSPARC48 suggested an approximately cubic structure, but with some deviations from perfect octahedral symmetry (Figure S2.4). Starting from a version of the design model containing only the trimeric component (i.e., with the tetramer and coiled-coil domains removed) low-pass Fourier-filtered to 20 Å resolution, homogeneous refinement was performed in lower symmetries, D2, D4, and T. The omission of a major component from the reference model and the imposition of a lower-than-expected symmetry served to validate the robustness of the reconstruction (Figure S2.4). In all cases, the reconstruction that was obtained clearly resembled an octahedron (i.e., the higher- symmetry O emerged without being imposed), fitting the dimensions of the design model, and with density appearing for the omitted tetrameric component. The reconstruction obtained with T symmetry imposed was then further refined by imposing O symmetry (Figure 2.3D).

To further confirm the assembly of the ccO34−1 cage, size exclusion chromatography with multiangle light scattering (SEC-MALS) was performed on the cage sample. From the prominent SEC peak that contained the cage species, a molecular weight estimate of 1.38 MDa was calculated. This is within 7% of the molecular weight that would be expected for the designed octahedral cage, providing confirmation of the intended assembly state.

Although structural information for the ccO34−1 cage lacked atomic resolution, the data still indicate a successful octahedral cage formation. In contrast to the ccT23−1 design, which was shown to favor the unintended ccD3 assembly when crystallized, there was no significant evidence of lower- symmetry assembly species in ccO34−1 samples, despite the appearance of some irregular species and larger aggregates in negative stain EM images (Figure 2.3B). Additionally, no major species appeared in 2D class averages that deviated significantly from the size that would be expected for the designed octahedral cage assembly (Figure S2.5). Differences between the design characteristics of the two cages that may have contributed to these contrasting results are discussed below.

Design Implications

Experimental characterizations of the two protein cage designs presented above, along with data from other recent design studies, suggest features of design types that tend to form unanticipated structures[11,23]. Certain combinations of protein oligomers can come together in alternative orientations to produce distinct assemblies having different point-group symmetries. This promiscuity naturally permits alternate experimental outcomes. For example, the combination of a C2 dimer and a C3 trimer is particularly promiscuous. Even when restricted to finite structures, this combination can lead to assemblies of I, O, T, or even D3 symmetries, depending on the relative orientation of the components (Figure S2.8). Indeed, we observed a

heterogeneous assembly for our C2 plus C3 construction here, despite attempts to limit the flexibility. This result parallels that of a previously reported designed protein cube, which was also comprised by a dimer plus trimer combination, held together by a single-helix linker.23 Although a crystal structure ultimately validated the cubic design in that case, supporting experiments indicated that smaller (i.e., T- and D3-symmetric) assemblies were also obtained in abundance. In contrast, the C3 plus C4 combination is one that does not allow other point symmetry combinations besides O, and indeed our design of this type led to a successful characterization at the level of negative stain EM and did not produce smaller species in abundance. The favorable designability for cases where alternate symmetric outcomes are absent has been emphasized in other studies[10−12].

A second geometric feature that appears to affect designability relates to the network properties of different kinds of architectures. The potential geometric difficulty of a design could be affected by the "ring size" in a network diagram describing the assembled structure. Both of the design results in the present study weigh in on this point. For the ccT23−1 design, the C2- and C3-symmetric oligomers form a triangular ring (or graph cycle) containing three oligomers of each type—hence, the ring size is 3 (Figure 2.1). Although it was hypothesized that the hetero coiled-coil fusion design method explored here would lead to rigidly oriented component oligomers, experimental studies showed heterogeneity, and ultimately a crystal structure showed preferential formation of an alternate form. In the collapsed D3 version of the assembly obtained by crystallization, the ring formed by the component oligomers is made up of only two oligomers of each type (ring size of 2) (Figure 2.2D). In essence, the intended design had ring size 3, but the crystal structure revealed a well-ordered alternate assembly with a smaller ring size of 2. For our C3 plus C4 octahedral construction, the ring size for this intended architecture is 2 (the

lowest possible value for a cage), and as noted above this case assembled with good robustness and did not collapse into smaller architectures. Another special case published recently further supports these ideas concerning design robustness.24 In that case, a genetic fusion of a pentamer and a dimer did not give a robust assembly of an icosahedron; the ring size for that network would have been 3. However, favorable results were obtained by fusing three oligomeric components (C2, C3, and C5). Only icosahedral symmetry is jointly compatible with all the component symmetries in that design, and a trimer plus pentamer combination gives an icosahedral network with ring size 2. Because self-assembling structures always proceed through intermediate forms, the possibility for incomplete structures to collapse forming smaller network rings than intended may be an important general consideration for design. Thus, to maximize the chances of design success with somewhat flexible linkages (intentionally or otherwise) between component oligomers, the targeted architecture should be one where the component symmetry types cannot give rise to alternate outcomes with smaller ring sizes. It is also possible that more robust designs could be produced by engineering additional sites of contacts that would favor the desired angular relationships between oligomers, or by negative design, for example, introducing repulsive steric interactions that might disfavor alternative undesirable geometries. We also note that protein solubility presented problems for some candidates. Additional efforts to design interfaces that remain soluble prior to formation of subunit−subunit associations could be fruitful. To the degree that appending single coiled-coil segments might promote unintended associations, other heterodimeric components could be investigated as alternatives to the coiled coil.

2.4 Conclusions

In this study we explored a strategy for designing symmetric protein cages that exploits key advantages evident in recent methods. A single α-helical linker used in earlier work was replaced by a coiled coil with the goal of reducing flexibility. A heteromeric coil−coil allowed the connection of two oligomeric components by a robust and predictable non- covalent interface. Ten computational designs were tested experimentally, with differing degrees of success leading to informative results. A tetrahedral design, based on a dimeric plus trimeric symmetry combination, was found experimentally to be geometrically promiscuous, forming heterogeneous assemblies. Size exclusion experiments and negative stain EM analysis suggest that the mixture likely included structures similar to what was designed, but ultimately a prominent component crystallized and was found to represent a smaller, collapsed structure with D3 symmetry (a6b6 stoichiometry instead of a12b12). A second design based on a non-promiscuous combination of trimer and cyclic tetramer to give symmetry O assembled largely as intended. Flexibility was still evident, limiting high resolution by electron microscopy, but lower- resolution negative stain imaging confirmed the essential design. Experimental trials such as these provide important guidance for optimizing protein design strategies. Given the overall experimental success rates for creating designer protein assemblies, which are currently rather modest, the heteromeric linker approach described here could be a productive avenue, especially when applied to favorable symmetry combinations.

2.5 Methods

Protein Construct Design. The coiled-coil fusion designs were performed using a similar approach to the previously described single- helix oligomer fusion method ([15,20,23,24]) with the primary distinction that the original α-helical linker was substituted by a heterodimeric coiled-coil linker[2] based on c-Fos (coiled-coil segment A) and c-Jun (coiled-coil segment B). The

model of the heterodimeric coiled-coil helical linker used in this study was generated from the crystal structure of the AP-1 c-Fos-c-Jun transcription factor bound to DNA [46]. The slight bending of the helices as a result of their interaction with DNA was accounted for by replacing the affected N-terminal and C-Terminal helical segments of c-Fos and c-Jun respectively with an idealized model of a 10-residue alanine α-helix. The sequences of c-Fos and c-Jun became AAAAAAAAAANRRRELTDTLQAETDQLEDEKSALQTEIAN and KRKLERIARLEEKVKTLKAQNSELASTANMAAAAAAAAAA respectively. The 10-residue alanine extensions in the resulting model were used to align the coiled-coil linker to the helical termini of symmetric oligomers during the design procedure. Furthermore, we identified sequences of a previously designed coiled-coil based on c-Fos-c-Jun with a significantly higher specificity for heterodimerization over homodimerization[41,47]. Therefore, in order to favor hetero over self- association of identical subunits, before ordering the genes encoding our final designs, we substituted the original c-Fos and c-Jun sequences for their engineered counterparts AQLEKELQALEKENAQLEWELQALEKELAQ (ACID-p1) and AQLKKKLQALKKKNAQLKWKLQALKKKLAQ (BASE-p1) respectively.

Homo-oligomeric biological assemblies with C2, C3, and C4 symmetry were downloaded from the Protein Data Bank, and Stride49 was used to analyze their secondary structure. Then, all possible pairs of dimers and trimers or trimers and tetramers with α-helical termini underwent a multistep alignment procedure. First, the N-terminus of coiled-coil segment A was used to align the entire heterodimeric Fos/Jun linker to the C- terminal helix of a subunit from oligomer 1. Then, a subunit from oligomer 2 was aligned at its N-terminal α-helix to the C-terminus of coiled-coil segment B. Additionally, helical extensions ranging from 1 to 4 residues were applied to the N- and C-terminus of coiled-coil segments A and B, respectively, and all

possible coiled-coil linker variations were tested for each pair. For each such geometric construction, the angle and distance between the cyclic symmetry axes of the two oligomers were then computed. Alignments that resulted in a distance of 3 Å or less between symmetry axes and an angle within 5° of 54.7° were considered for further analysis. From this pool of ~100, candidates were then manually inspected, and any that contained major steric clashes were removed. Those exhibiting one or two minor side-chain clashes caused by a helical region in the native sequence of an oligomeric building block were retained. Such steric clashes were mitigated by alanine point mutations in the following designs: ccO34−1 (trimer), ccO34−2 (tetramer), and ccO34−6 (trimer). A total of 10 designs were selected for experimental testing. Amino acid sequences are provided in Table S2.2.

Synthetic Gene Construction. E. coli codon-optimized gene fragments encoding the different protein design constructs were purchased from Integrated DNA Technologies. Sequences encoding the dimer and trimer subunits of the ccT23−1 design were inserted into the pCDFDuet-1 and pET-22b expression vectors (Novagen), respectively, via a Gibson Assembly. Sequences encoding the trimer and tetramer subunits of the ccO34−1 design were both inserted into the same pET-22b plasmid separated by the intergenic region of pETDuet-1 (Novagen).

Protein Expression and Purification. Both the ccT23−1 dimer and trimer subunits were transformed separately into E. coli BL21 (DE3) cells (New England Biolabs). Starter cultures were grown in Luria broth (LB) with overnight shaking at 37 °C and used to inoculate 1 L of LB media supplemented with spectinomycin (pCDFDuet-1) or ampicillin (pET-22b) at 75 μg/mL. Cultures were grown with shaking at 37 °C until reaching an OD600 of 0.6, after which protein expression was induced with 0.4 mM isopropyl β-D-1- thiogalactopyranoside (IPTG). Cells were incubated for a further 4 h with shaking at 37 °C before being harvested by centrifugation at

5000g for 10 min. Harvested cells were suspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 2.8 mM β-mercaptoethanol, 5 mM MgCl2, and 10% glycerol) supplemented with 10 mM imidazole pH 7.5 and ethylenediaminetetraacetic acid (EDTA) free protease inhibitor (Pierce; Thermo Fisher Scientific) and colysed with an Emulsiflex C3 (Avestin). Lysate was clarified by centrifugation at 9500g for 40 min at 4 °C. For a nickel-affinity purification, the soluble lysate fraction was incubated at 4 °C for 90 min with 4 mL of Ni-NTA resin on a rocker before being loaded onto a gravity column and washed with 100 mL of lysis buffer containing 75 mM imidazole. The protein was eluted from the column with 20 mL of lysis buffer containing 250 mM imidazole. The eluted sample was concentrated to 4 mg/mL, then further purified by size exclusion chromatography on a Superose 6 30/100 column (GE Healthcare) equilibrated with SEC buffer (0.5 M NaCl, 5% glycerol, 2.8 mM β-mercaptoethanol, 50 mM Tris pH 7.5, and 5 mM MgCl2). Elution fractions corresponding to the first peak after the void volume were evaluated by SDS-PAGE. Fractions containing both ccT23−1 subunits were used for further characterization (see Protein Crystallization and Figure S2.6).

E. coli BL21 (DE3) cells (New England Biolabs) transformed with the ccO34−1 expression plasmid were grown in P-0.5G noninducing minimal medium overnight.50 The cell culture (10 mL) was then used to inoculate 1 L of ZYP-5052 autoinducing medium supplemented with ampicillin at 75 μg/mL for 65 h at 20 °C (Studier 2005). Cells were harvested by centrifugation at 4000g for 20 min, resuspended in lysis buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0, 5% glycerol, 250 mM NaCl, and 2.8 mM β-mercaptoethanol) supplemented with EDTA free protease inhibitor (Pierce; Thermo Fisher Scientific) and lysed with an Emulsiflex C3 (Avestin). Lysate was centrifuged at 12 000g for 35 min at 4 °C, and the soluble fraction was applied to a 4 mL Ni-NTA gravity column. A series of

25 mL wash buffers (lysis buffer with 25, 40, or 75 mM imidazole) were applied to the column sequentially with increasing amounts of imidazole before eluting the target protein with a lysis buffer containing 250 mM imidazole. The eluate was dialyzed overnight into 25 mM HEPES pH 7.0, 250 mM NaCl, 2% glycerol, and 1.4 mM β-mercaptoethanol at 4 °C. The dialyzed protein was then further purified by SEC on a Superose 6 30/100 column (GE Healthcare) equilibrated with a dialysis buffer. Elution fractions corresponding to the first peak after the void volume were evaluated by SDS-PAGE. Fractions that eluted between 11 and 13 mL contained both ccO34−1 subunits and were used for downstream analysis.

Protein Crystallization. SEC elution fractions (12−15 mL) of the ccT23−1 sample were pooled, concentrated to 2.8 mg/ mL, passed through a 0.22 μm filter, and centrifuged at 9500g for 10 min at 4 °C. The target protein was crystallized using the hanging- drop vapor diffusion method. Crystallization trials were performed at the UCLA Crystallization Facility in 96-well trays set with 210 nL drops using a Mosquito liquid handling device (TPP LabTech). The crystal used for the structure determination was obtained using the Silver Bullets additive screen by Hampton Research. The reservoir contained 90 μL of a crystallization reagent (0.10 M sodium acetate pH 5.4, 66% 2-methyl-2,4-pentanediol) and 10 μL of Silver Bullets reagent B9 (0.25% w/v hexamminecobalt(III) chloride, 0.25% w/v salicylamide, 0.25% w/v sulfanilamide, 0.25% w/v vanillic acid, 0.02 M HEPES sodium pH 6.8). The hanging drop contained a purified protein and reservoir solution in a 2:1 (v/v) ratio. Crystals took ∼7− 14 days to grow at 4°C.

X-ray Diffraction Data Collection and Structure Determination. X-ray diffraction data from a single ccD3 crystal were collected at beamline 24-ID-C of the Advanced Photon Source, Argonne National Laboratory, at a wavelength of 0.9790 Å and temperature of 100 K. Data were collected using 0.5° oscillations and 697 mm detector distance with a DECTRIS PILATUS-6MF

pixel detector. Indexing and integration of the reflections were performed using the XDS software in space group P4332 and scaled with XSCALE to a resolution of 4.32 Å.51 Initial molecular replacement calculations conducted in cubic space groups appeared unreliable, so the structure was solved by molecular replacement in space group P1 using the program Phaser[52] with search models 3DXO and 1WY1 corresponding to the dimer and trimer components of our ccT23 design, respectively. A difference density map revealed positions of the coiled coils, which were then modeled using Coot[53]. We manually rotated and translated the coiled coils from 4G1E to fit the density and made adjustments to the dimer and trimer regions. This model was refined using Phenix[54]. A single heterodimer from the improved model was retrospectively used as a search model for a molecular replacement in cubic space groups P4132 and P4332; the correct solution was evident in the latter. We proceeded with a rigid body refinement and then with simulated annealing and gradient-driven minimization. Because of the limited resolution of the data set, we applied multiple geometric restraints including rotamer restraints, hydrogen bond restraints for the coiled-coil region, and restraints to reference models 3DXO and 1WY1. Furthermore, atomic displacement parameters were grouped throughout the refinement process, and anisotropic displacement was modeled using a single translation- libration-screw group. We alternated between Phenix refinement and manual building for several iterations. Data collection and refinement statistics are reported in Table S2.1.

Purification of ccT23. After a nickel-affinity purification the ccT23 sample was further purified by size exclusion chromatography on a Superose 6 30/100 column (GE Healthcare) equilibrated with 500 mM NaCl, 20 mM HEPES pH 7.5, 5 mM MgCl2, 10% glycerol, and 1.4 mM β-mercaptoethanol. Elution fractions (11−15 mL) corresponding to the first peak after the void volume were pooled and concentrated to 1.1 mg/mL.

Negative Stain Electron Microscopy. Initially, partially assembled complexes of ccT23 appeared to be predominant in negative-stain EM preparations. The purified sample of ccT23was therefore cross-linked using 0.02% glutaraldehyde (Ted Pella) to mitigate the effects of staining and grid preparation, a step that was found to produce better particles (Figure S2.6). The reaction was quenched after 4 min with 100 mM Tris pH 8. The cross-linked sample was diluted to either 0.01 or 0.35 mg/mL in 300 mM NaCl, 15 mM Tris pH 7.5, 5 mM MgCl2, 2% glycerol, and 1.4 mM β- mercaptoethanol for analysis by negative-stain electron microscopy.

Protein samples were applied onto a glow-discharged, 300-mesh carbon-coated Formvar-supported copper grid (Ted Pella), washed with Milli-Q water, and stained with 2% uranyl acetate. Initial sample screening was performed on a Tecnai T12 transmission electron microscope. Micrographs used for further structural characterizations of ccO34−1 and ccT23−1 were collected on an FEI TF20 transmission electron microscope at 29 000× and 50 000× magnification, respectively.

Electron microscopy data for the ccO34−1 construct were processed using cryoSPARC 2.15.0.48 Reference-free 2D class averages were obtained from 7257 autopicked particles. Homoge- neous refinement was performed on the particles using a reference map consisting of a 20 Å filtered version of the design model with the tetrameric component and coiled-coil linker removed. For the final 3D reconstruction, a first round of homogeneous refinement was performed with T symmetry imposed, followed by subsequent rounds which imposed O symmetry.

Data for the ccT23−1 construct were processed using cryoSPARC and Relion 3.1.55 2727 particles were picked manually and extracted in cryoSPARC. Reference-free 2D class averages were produced in Relion.

Size Exclusion Chromatography with Multi-Angle Light Scattering. For SEC-MALS experiments, SEC fractions that contained ccO34−1 were pooled and concentrated to 4 mg/mL. Each concentrated sample was then run on an AKTA pure system equipped with a Superose 6 Increase column (Cytiva Life Sciences) with in-line miniDAWN multiangle light-scattering instruments (Wyatt Technology). Data analysis was performed using the ASTRA software suite (Wyatt Technology).

2.6 Acknowledgements

The authors thank D. Cascio and M. Collazo for help with crystallization trials as well as Prof. H. Zhou and I. Atanasov for guidance on electron microscopy. This work was supported by NSF Grant No. CHE-1629214. This work used NE-CAT beamline 24-ID-C (GM124165) and a Pilatus detector (RR029205) at the APS (DE-AC02-06CH11357). We thank the NE-CAT staff for their assistance with data collection.

2.7 Supplemental Data

Supplemental Tables:

Space group	$P4_332$
Cell dimensions a, b, c (Å)	146.34, 146.34, 146.34
Resolution (Å)	84.5 – 4.32
% Data completeness	99.5 (96.9)
Data redundancy	17.7 (16.8)
R_{merge}	0.10 (2.37)
I/σ(I)	14.40 (1.07)
CC ½ %	100.0 (41.1)
R_{work}	0.278 (0.404)
R_{free}	0.293 (0.425)
Ramachandran (Favored, Allowed, Outliers)	97.8%, 2.2%, 0 %
Average B value Chain A (1354 atoms)	258 $Å^2$
Average B value Chain B (1153 atoms)	279 $Å^2$

Parenthetic values refer to the outer resolution shell.

Table S2.1 Crystallographic table

ccT23-1 Dimer (based on 3DXO)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQATQHLTIAQTYLAAWNEEDNERRRHLV GQAWAENTRYVDPLMQGEGQQGIAAMIEAARQKFPGYRFVLAGTPDGHGNFTRFSW RLISPDGDDVAGGTDVVSLNTEGRIDNVVGFLDGAVSHHHHHH
ccT23-1 Trimer (based on 1WY1)	MSPIIEANGTLDELTSFIGEAKHYVDEEMKGILEEIQNDIYKIMGEIGSKGKIEGISEERIKW LEGLISRYEEMVNLKSFVLPGGTLESAKLDVCRTIARRAERKVATVLREFGIGKEALVYLNR LSDLLFLLARVIEIAAAAQLEKELQALEKENAQLEWELQALEKELAQ
ccO34-1 Trimer (based on 4AM8)	MKRDYVTTETYTKEEMHYLVDLSLKIKEAIKNGYYPQLLKNKSLGMIFQQSSTGTRVSFET AMEQLGGHGEYLAPGQIQLGGHETIEDTSRVLSRLVDILMARVERHHSIVDLANCATIPVI NGMSDYNHPTQELGDLCTMVEHLPEGKKLEDCKVVFVGDATQVCFSLGLITTKMGMN FVHFGPEGFQLNEEHQAKLAKNCEVSGGSFLVTDDASSVEGADFLYTDVWYGLYEAELS EEERMKVFYPKYQVNQEMMDRAGANCKFMHCLPATRGEEVTDEVIDGKNSICFDEAE NRLTSIRGLLVYLMNDYEAKNPYDLIKQAAAKKALEVFLDTQAAAAQLEKELQALEKENA QLEWELQALEKELAQ
ccO34-1 Tetramer (based on 2G7O)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAAQTEFNKLLLECVVKTQSSVAKILGIES LSPHVSGNSKFEYANMVEDIREKVSSEMERFFPKNDDEHHHHHH
ccO34-2 Trimer (based on 2V82)	MHHHHHHQWQTKLPLIAILRGITPDEALAHVGAVIDAGFDAVEIPLNSPQWEQSIPAIV DAYGDKALIGAGTVLKPEQVDALARMGCQLIVTPNIHSEVIRRAVGYGMTVCPGCATAT EAFTALEAGAQALKIFPSSAFGPQYIKALKAVLPSDIAVFAVGGVTPENLAQWIDAGCAG AGLGSDLYRAGQSVERTAQQAAAFVKAYREAVQAQLEKELQALEKENAQLEWELQALE KELAQ
ccO34-2 Tetramer (based on 1E4C)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAAARNKLARQIIDTCLEMTRLGLNQGT AGNVSVRYADGMLITPTGIPYEKLTESHIVFIDGNGKHEEGKLPQSEWRFHMAAYQSRP DANAVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNSIPCAPYATFGTRELSEHVALALK NRKATLLQHHGLIACEVNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIAVVLEKFK TFGLRIEE
ccO34-3 Trimer (based on 4G9Q)	MHHHHHHSSGVDLGTENLYFQSMMTTSNAGAQQPNVEGRRFSPDQVRSVAPALEQY TQQRLYGDVWQRPGLNRRDRSLVTIAALIARGEAPALTYYADQALENGVKPSEISETITHL AYYSGWGKAMATVGPVSEAFAKRGIGQDQLAAVESTPLPLDEEAEAQRATTVGNQFGS VAPGLVQYTTDYLFRDLWLRPDLAPRDRSLVTIAALISVGQVEQITFHLNKALDNGLSEEQ AAEVITHLAFYAGWPNAMSALPVAKAVFEKRRAAQLEKELQALEKENAQLEWELQALEK ELAQ
ccO34-3 Tetramer (based on 1KBJ)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAAAKEDIARKEQLKSLLPPLDNIINLYDF EYLASQTLTKQAWAYYSSGANDEVTHRENHNAYHRIFFKPKILVDVRKVDISTDMLGSH VDVPFYVSATALCKLGNPLEGEKDVARGCGQGVTKVPQMISTLASCSPEEIIEAAPSDKQI QWYQLYVNSDRKITDDLVKNVEKLGVKALFVTVDAPSLGQREKDMKLKFSNTKAGPKA MKKTNVEESQGASRALSKFIDPSLTWKDIEELKKKTKLPIVIKGVQRTEDVIKAAEIGVSGV VLSNHGGRQLDFSRAPIEVLAETMPILEQRNLKDKLEVFVDGGVRRGTDVLKALCLGAKG VGLGRPFLYANSCYGRNGVEKAIEILRDEIEMSMRLLGVTSIAELKPDLLDLSTLKARTVGV PNDVLYNEVYEGPTLTEFEDA
ccO34-4 Trimer (based on 2V82)	MHHHHHHQWQTKLPLIAILRGITPDEALAHVGAVIDAGFDAVEIPLNSPQWEQSIPAIV DAYGDKALIGAGTVLKPEQVDALARMGCQLIVTPNIHSEVIRRAVGYGMTVCPGCATAT EAFTALEAGAQALKIFPSSAFGPQYIKALKAVLPSDIAVFAVGGVTPENLAQWIDAGCAG AGLGSDLYRAGQSVERTAQQAAAFVKAYREAVQAAAQLEKELQALEKENAQLEWELQA LEKELAQ

ccO34-4 Tetramer (based on 2FLF)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAQRAERARLYAAFRQVGEDLFAQGLIS ATAGNFSVRTKGGFLITKSGVQKARLTPEDLLEVPLEGPIPEGASVESVVHREVYRRTGAR ALVHAHPRVAVALSFHLSRLRPLDLEGQHYLKEVPVLAPKTVSATEEAALSVAEALREHRA CLLRGHGAFAVGLKEAPEEALLEAYGLMTTLEESAQILLYHRLWQGAGPALGGGE
ccO34-5 Trimer (based on 4IV5)	MLELPPVASLGGKSITSAEQFSRADIYALIHLASAMQRKIDAGEVLNLLQGRIMTPLFFED SSRTFSSFCAAMIRLGGSVVNFKVEASSINKGETLADTIRTLDSYSDVLVMRHPRQDAIEE ALSVAQHPILNAGNGAGEHPTQALLDTLTIHSELGSVDGITIALIGDLKMGRTVHSLLKLLV RNFSIKCVFLVAPDALQMPQDVLEPLQHEIATKGVIIHRTHALTDEVMQKSDVLYTTRLQ KERFMASTSDDAAALQSFAAKADITIDAARMRLAKEKMIVMHPLPRNDELSTTVDADPR AAYFRQMRYGMFMRMAILWSVLAAAQLEKELQALEKENAQLEWELQALEKELAQ
ccO34-5 Tetramer (based on 2FLF)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAQRAERARLYAAFRQVGEDLFAQGLIS ATAGNFSVRTKGGFLITKSGVQKARLTPEDLLEVPLEGPIPEGASVESVVHREVYRRTGAR ALVHAHPRVAVALSFHLSRLRPLDLEGQHYLKEVPVLAPKTVSATEEAALSVAEALREHRA CLLRGHGAFAVGLKEAPEEALLEAYGLMTTLEESAQILLYHRLWQGAGPALGGGEHHHH HH
ccO34-6 Trimer (based on 3Q1X)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAADNYIYSIAHQLYEMYLQDEDAFHSK RDYPHKKVFTELQKLRKIFFPDFFMKHQKITESHIASELTKLVDYIKDSVTAYNDELFAAQC VMAILEKLPSIKRTLKTDLIAAYAGDPAAPGLSLIIRCYPGFQAVIVYRIAHVLYECGERYYCR EMMESVHSYTSIDIHPGASIKGHFFIDHGVGVVIGETAIIGEWCRIYQSVTLGAMHFQEE GGVIKRGTKRHPTVGDYVTIGTGAKVLGNIIVGSHVRIGANCWIDRDVDSNQTVYISEHP THFVKPCTTKGMKNDTEIIAIIPSSPLANSPSILEHHHHHH
ccO34-6 Tetramer (based on 3RPZ)	MSNAMNVPFWTEEHVGATLPERDAESHKGTYGTALLLAGSDDMPGAALLAGLGAMR SGLGKLVIGTSENVIPLIVPVLPEATYWRDGWKKAADAQLEETYRAIAIGPGLPQTESVQQ AVDHVLTADCPVILDAGALAKRTYPKREGPVILTPHPGEFFRMTGVPVNELQKKRAEYAK EWAAQLQTVIVLKGNQTVIAFPDGDCWLNPTGNGALAKGGTGDTLTGMILGMLCCHE DPKHAVLNAVYLHGACAELWTDEHSAHTLLAHELSDILPRVWKRFEAAAAQLEKELQAL EKENAQLEWELQALEKELAQ
ccO34-7 Trimer (based on 3OER)	MSVESSTDGQVVPQEVLNLPLEKAHEEADDYLDHLLDSLEELSEAHPDCIPDVELSHGVM TLEIPAFGTYVINKQPPNKQIWLASPLSGPNRFDLLNGEWVSLRNGTKLTDILTEEVEKAIS EAAAQLEKELQALEKENAQLEWELQALEKELAQ
ccO34-7 Tetramer (based on 2Z7B)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAQARRKVFEELVTATKILLNEGIMDTFG HISARDPEDPASFFLAQKLAPSLITVDDIQRFNLDGETSDNRPSYLERYIHSEIYKTRPDVQC VLHTHSPAVLPYCFVDTPLRPVTHMGAFIGESVPVYEIRDKHGDETDLFGGSPDVCADIA ESLGSQTVVLMARHGVVNVGKSVREVVFRAFYLEQEAAALTAGLKIGNVKYLSPGEIKTA GKLVGAQIDRGWNHWSQRLRQAGLAHHHHHH
ccO34-8 Trimer (based on 2P4S)	MYTYDTLQEIATYLLERTELRPKVGIICGSGLGTLAEQLTDVDSFDYETIPHFPVSTVAGHV GRLVFGYLAGVPVMCMQGRFHHYEGYPLAKCAMPVRVMHLIGCTHLIATNAAGGANP KYRVGDIMLIKDHINLMGFAGNNPLQGPNDERFGPRFFGMANTYDPKLNQQAKVIAR QIGIENELREGVYTCLGGPNFETVAEVKMLSMLGVDAIGMSTVHEIITARHCGMTCFAFS LITNMCTMSYEEEEEHCHDSIVGVGKNREKTLGEFVSRIVKHIHYEAAQLEKELQALEKEN AQLEWELQALEKELAQ
ccO34-8 Tetramer (based on 2R9G)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAQHYDVISAFQKSIRGSDVDAALHYLAR LVEAGDLASICRRLMVIGYEDIGLGNPAAAARTVNAVLAAEKLGLPEARIPLADVVVDLCL SPKSNSAYMALDAALADIREGKAGDVPDHLRDSHYKGAKSLNRGVGYQYPHHFDQAW VNQQYLPDKLKNAQYYQPKDTGKYEQALGQQYYRIKEWKEHHHHHH

ccO23-1 Dimer (based on 3OCU)	MAQLKKKLQALKKKNAQLKWKLQALKKKLAQAEHANMQLQQQAVLGLNWMQDSGE YKALAYQAYNAAKVAFDHAKVAKGKKKAVVADLNETMLDNSPYAGWQVQNNKPFDG KDWTRWVDARQSRAVPGAVEFNNYVNSHNGKVFYVTNRKDSTEKSGTIDDMKRLGFN GVEESAFYLKKDKSAKAARFAEIEKQGYEIVLYVGDNLDDFGNTVYGKLNADRRAFVDQ NQGKFGKTFIMLPNANYGGWEGGLAEGYFKKDTQGQIKARLDAVQAWDGKHHHHH H
ccO23-1 Trimer (based on 1VL0)	MKILITGANGQLGREIQKQLKGKNVEVIPTDVQDLDITNVLAVNKFFNEKKPNVVINCAA HTAVDKCEEQYDLAYKINAIGPKNLAAAAYSVGAEIVQISTDYVFDGEAKEPITEFDEVNP QSAYGKTKLEGENFVKALNPKYYIVRTAWLYGDGNNFVKTMINLGKTHDELKVVHDQV GTPTSTVDLARVVLKVIDEKNYGTFHCTCKGICSWYDFAVEIFRLTGIDVKVTPCTTEEFPR PAKRPKYSVLRNYMLELTTGDITREWKESLKEYIDLLQMAAAQLEKELQALEKENAQLEW ELQALEKELAQ

Table S2.2 Amino Acid Sequences

Supplemental Figures:

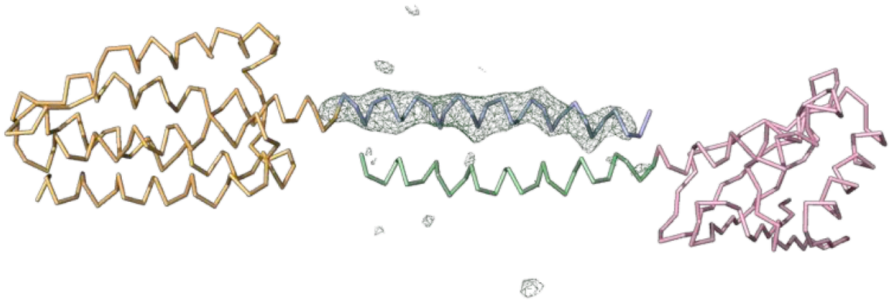


Figure S2.1. Electron density map of the ccD3 structure. An omit map (green mesh) was used to establish the correctness of the molecular replacement solution. The density map, computed after omitting the coiled-coil segment (blue) fused to the trimer subunit (orange), clearly reveals positive density for the omitted region. The dimer subunit is shown in pink with its coiled-coil segment in green.

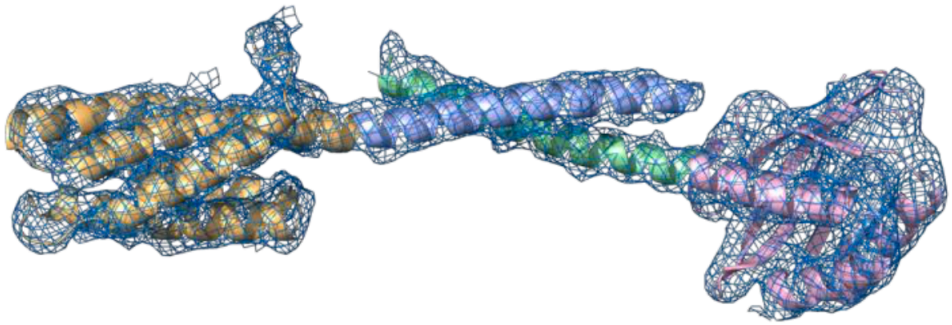


Figure S2.2. Electron density map for the ccD3 crystal structure. A 2Fo-Fc electron density map (blue mesh) contouring the asymmetric unit of the ccD3 crystal structure at 1.1 σ.

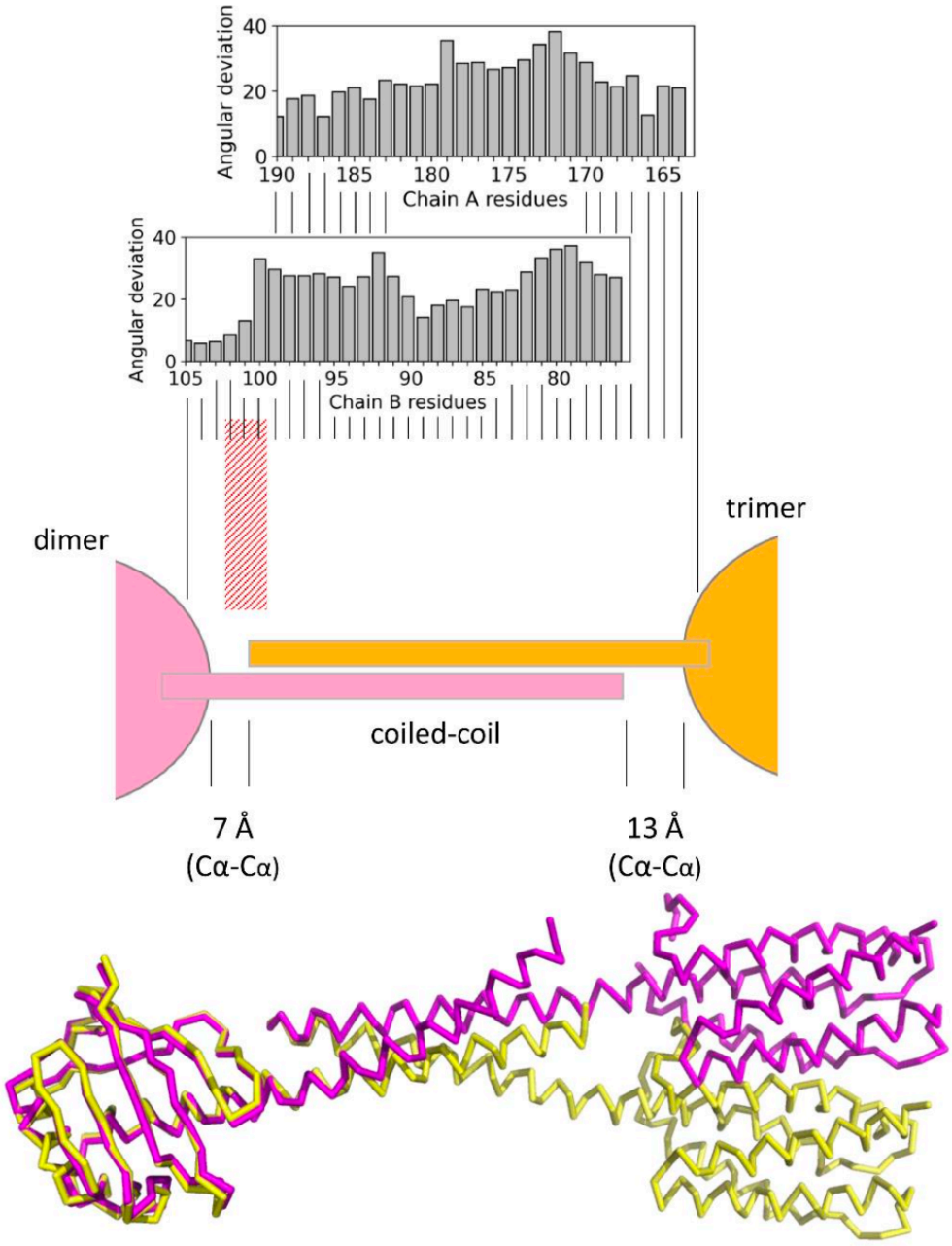


Figure S2.3. Analysis of deviations of the observed ccD3 structure from the designed model. (bottom) The designed (purple) and observed (yellow) structures are shown, overlapped on the dimer components, showing bending that begins near the left side of the coiled-coil region. (top) The bar graphs at the top show the calculated rotational deviation at each C-alpha position between the designed and observed models. The values reflect cumulative error, which begins near zero at the dimer (because the models begin in alignment at the dimer) and progresses to

angular values approaching the difference between the intended tetrahedral symmetry angle of 54.7° (between 2-fold and 3-fold axes) and the 90° value realized in the D3 structure, a difference of about 35°. The calculations over chain B (the dimer component fused to one coiled-coil segment) show that angular error arises steeply near the transition into the coiled-coil (red shading).

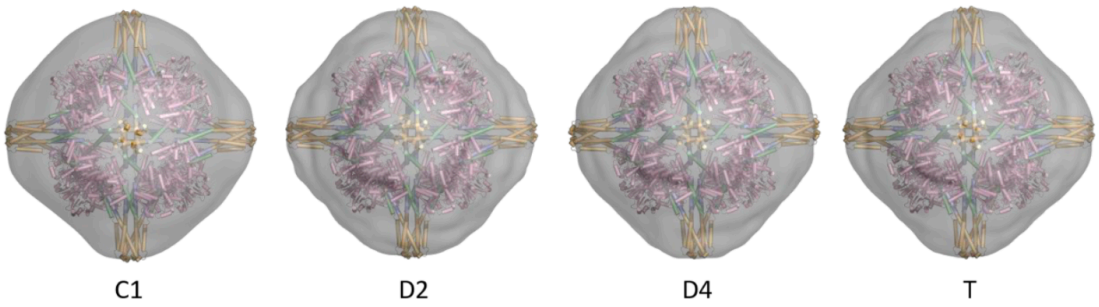


Figure S2.4. 3D single particle electron microscopy reconstructions produced by homogeneous refinement with various types of symmetry imposed for the ccO34-1 octahedral cage. Regardless of the lower symmetry that was imposed, the reconstructions conform to near-octahedral symmetry. The reconstructions were generated using as a reference a low resolution (20 Å) map of a partial model from which major components – the tetramer (colored in yellow) and coiled-coil components – were removed as a test of robustness.

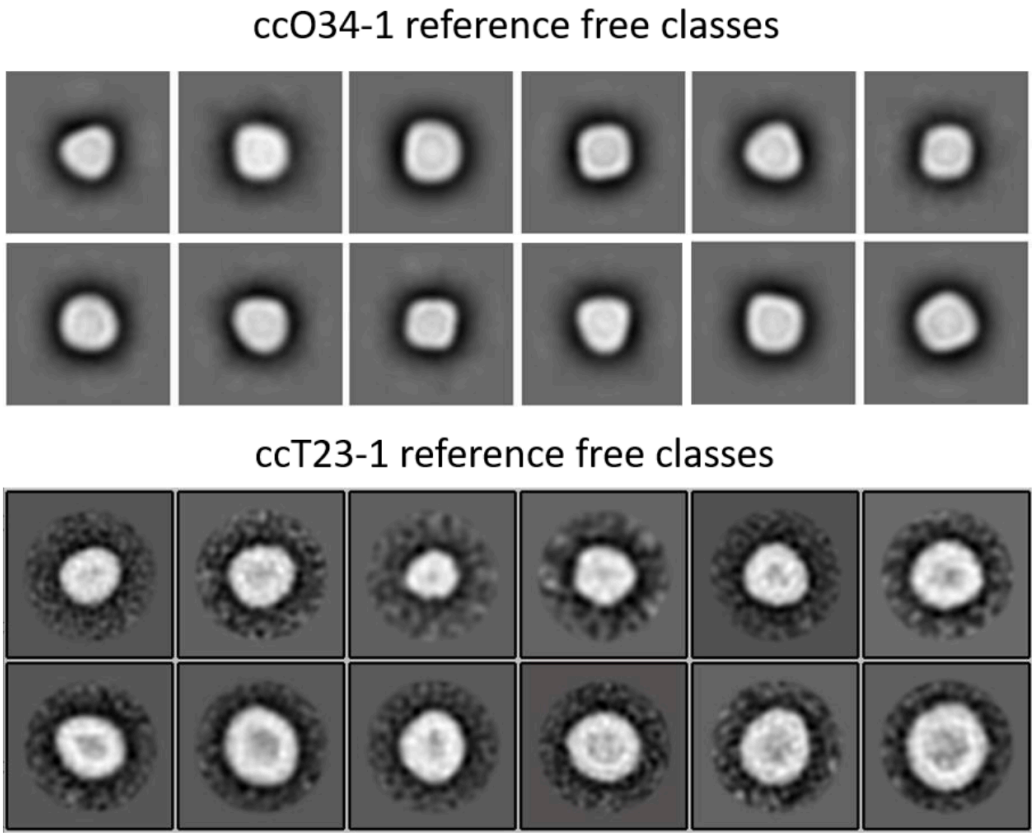


Figure S2.5. Reference-free 2D class averages of particles for the ccO34-1 (top) and ccT23-1 (bottom) cages from negative stain EM images. The ccO34-1 construct produced cages of relatively uniform size (roughly 25 nm), whereas the ccT23-1 construct showed particles with a range of sizes. Medium size particles of ccT23-1 are in the range of 20-22 nm, which is close to the correct size for the designed tetrahedron. The smallest class observed is 14 nm in size, which could correspond to smaller species like the D3 assembly which was observed by X-ray crystallography. The largest classes reach about 29 nm in size, which could correspond to assemblies with more subunits, *e.g.* octahedra with 24 subunits. Classes are ordered from highest to lowest relative abundance (top left to bottom right).

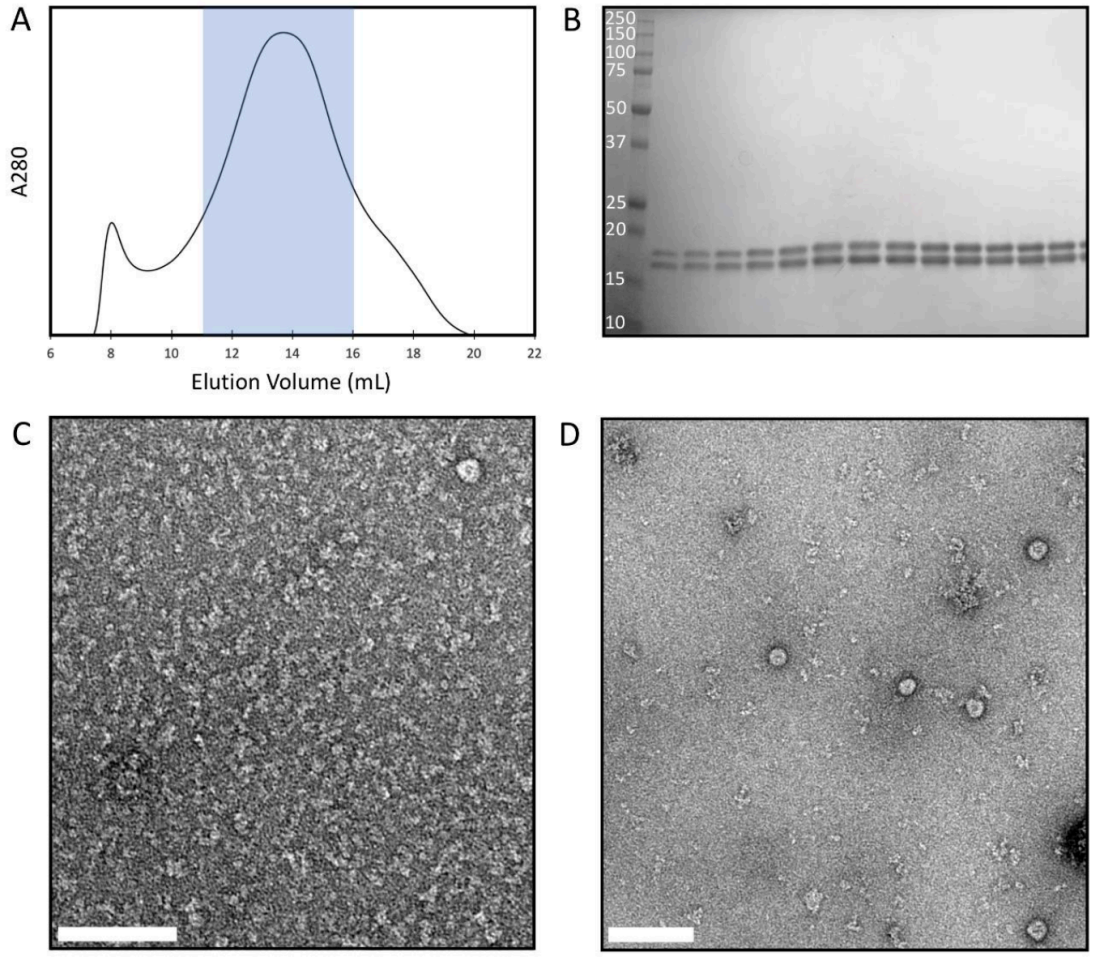


Figure S2.6. Purification and glutaraldehyde crosslinking of ccT23-01. A) A broad major absorbance peak (11 mL – 17 mL) occurs immediately after the void volume on SEC. B) Bands for both dimer and trimer subunits appear on SDS-PAGE in all of the 0.5 mL SEC elution fractions corresponding to the peak (shaded blue in A). The left-most lane of the gel contains the Precision Plus Protein Standard marker (Bio-Rad). Molecular weights are in kDa. C) At low concentrations employed for preparation of EM grids (approx. 0.01 mg/ml), mainly small assembly species were observed for ccT23 while cage-like particles were scarce. The ccT23 sample was crosslinked with 0.02% glutaraldehyde for 4 minutes at a higher protein concentration (approx. 1.1 mg/mL) and was then diluted to approx. 0.01 mg/mL. D) Cage-like particles are significantly more abundant in the crosslinked sample under negative stain EM. Negative stain micrographs were collected on a Tecnai T12 transmission electron microscope. Scale bars (white) are 100 nm. SEC elution fractions used for negative stain EM are highlighted in blue.

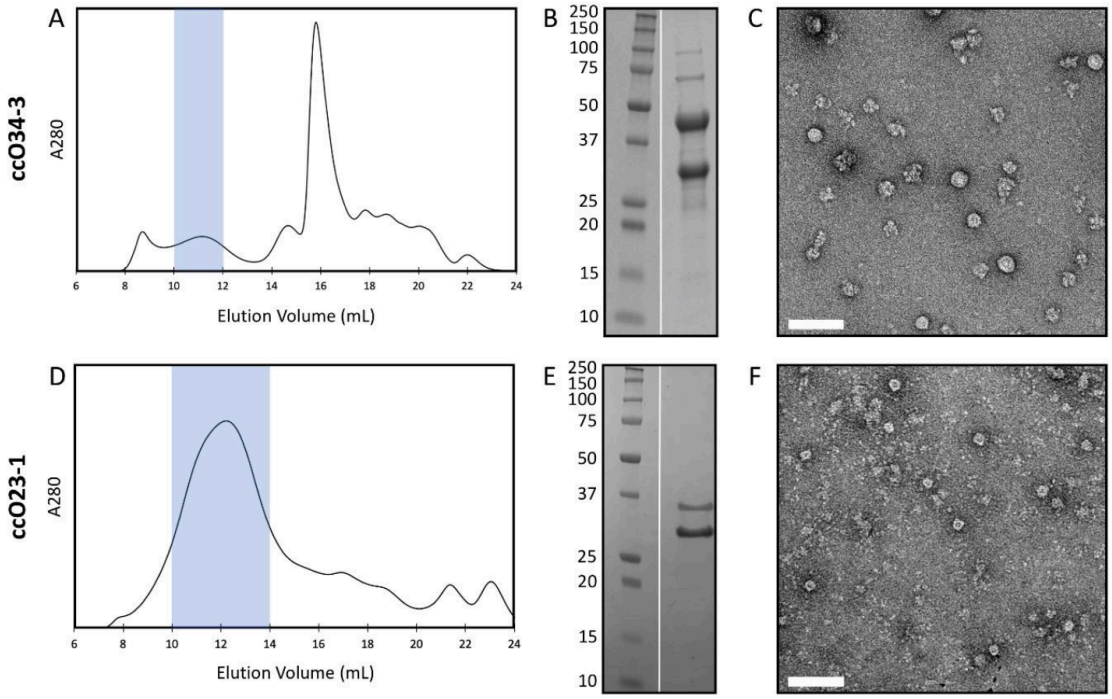


Figure S2.7. Characterization of ccO34-3 and ccO23-1 designs. A) The ccO34-3 construct exhibits a small yet narrow absorbance peak (10mL-12mL) immediately after the void volume on SEC. B) Bands for both trimer and tetramer subunits appear on SDS-PAGE of the pooled elution fractions of the SEC peak (shaded blue). C) Although protein aggregates and partial assemblies are prominent under negative stain EM, cage-like assemblies are readily observed. D) The ccO23-1 construct exhibits a broad absorbance peak (9mL-14mL) immediately after the void volume on SEC. E) Bands for both dimer and trimer subunits appear on SDS- PAGE of the pooled elution fractions of the SEC peak (shaded blue). F) Cage-like assemblies as well as protein aggregates are readily observed under negative stain EM. Left SDS-PAGE lanes contain the Precision Plus Protein Standard marker (Bio-Rad). Molecular weights are in kDa. Negative stain micrographs were collected on a Tecnai T12 transmission electron microscope and scale bars (white) are 100 nm. SEC elution fractions used for negative stain EM and SDS-PAGE are highlighted in blue.

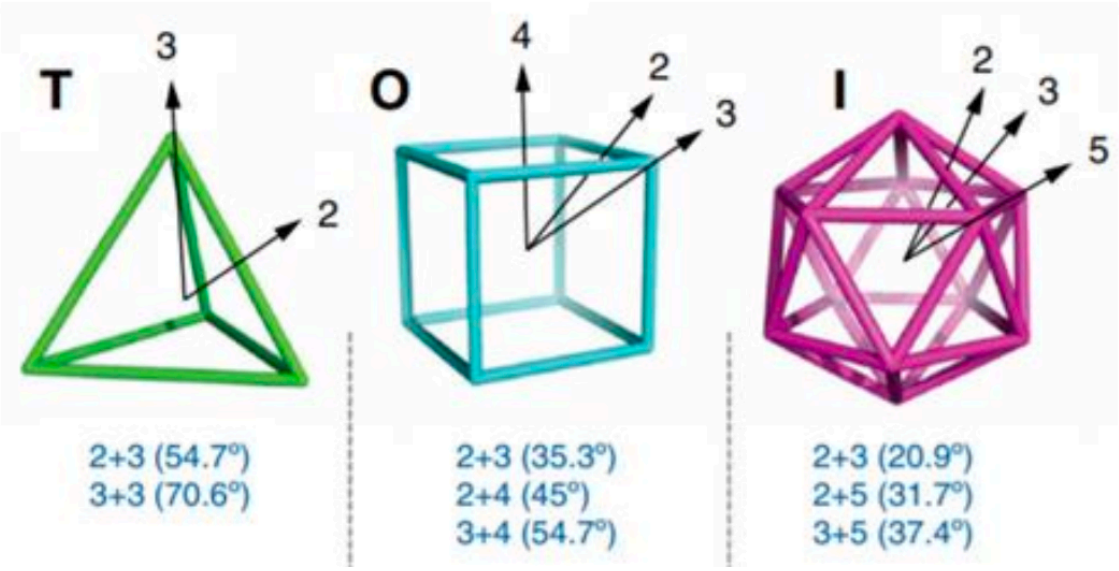


Figure S2.8. Diagram of Platonic solids showing the angles between intersecting symmetry axes of different types for the cubic symmetries: tetrahedral (T), octahedral (O) and icosahedral (I). For symmetry D3 (not shown) 2-fold symmetry axes are perpendicular to a 3-fold axis of rotational symmetry. (Reprinted with permission from Yeates, T.O.; Liu, Y.; Laniado, J. The Design of Symmetric Protein Nanomaterials Comes of Age in Theory and Practice. *Curr. Opin. Struct. Biol.* 2016, *39*, 134-143. Copyright 2016 Elsevier).

2.8 References

(1) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. Nature 2016, 537 (7620), 320−327.

(2) Yeates, T. O.; Liu, Y.; Laniado, J. The Design of Symmetric Protein Nanomaterials Comes of Age in Theory and Practice. Curr. Opin. Struct. Biol. 2016, 39, 134−143.

(3) Cannon, K. A.; Ochoa, J. M.; Yeates, T. O. High-Symmetry Protein Assemblies: Patterns and Emerging Applications. Curr. Opin. Struct. Biol. 2019, 55, 77−84.

(4) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. Self-Assembling Cages from Coiled-Coil Peptide Modules. Science 2013, 340 (6132), 595−599.

(5)Ljubetič,A.;Lapenta,F.;Gradišar,H.;Drobnak,I.;Aupič,J.; Strmšek,Ž.;Lainščěk,D.;Hafner-Bratkovič,I.;Majerle,A.;Krivec, N.;Benčina,M.;Pisanski,T.;Velič̆kovič́,T.Ć.;Round,A.;Carazo,J. M.; Melero, R.; Jerala, R. Design of Coiled-Coil Protein-Origami Cages That Self-Assemble in Vitro and in Vivo. Nat. Biotechnol. 2017, 35 (11), 1094−1101.

(6) Sasaki, E.; Böhringer, D.; Van De Waterbeemd, M.; Leibundgut, M.; Zschoche, R.; Heck, A. J. R.; Ban, N.; Hilvert, D. Structure and Assembly of Scalable Porous Protein Cages. Nat. Commun. 2017, 8, 1−10.

(7) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A. Self-Assembly of Coherently Dynamic, Auxetic, Two-Dimensional Protein Crystals. Nature 2016, 533 (7603), 369− 373.

(8) Gonen, S.; DiMaio, F.; Gonen, T.; Baker, D. Design of Ordered Two-Dimensional Arrays Mediated by Noncovalent Protein-Protein Interfaces. Science 2015, 348 (6241), 1365−1368.

(9) Lanci, C. J.; MacDermaid, C. M.; Kang, S.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. Computational Design of a Protein Crystal. Proc. Natl. Acad. Sci. U. S. A. 2012, 109 (19), 7304−7309.

(10) Sciore, A.; Su, M.; Koldewey, P.; Eschweiler, J. D.; Diffley, K. A.; Linhares, B. M.; Ruotolo, B. T.; Bardwell, J. C. A.; Skiniotis, G.; Marsh, E. N. G. Flexible, Symmetry-Directed Approach to Assembling Protein Cages. Proc. Natl. Acad. Sci. U. S. A. 2016, 113 (31), 8681−8686.

(11) Badieyan, S.; Sciore, A.; Eschweiler, J. D.; Koldewey, P.; Cristie- David, A. S.; Ruotolo, B. T.; Bardwell, J. C. A.; Su, M.; Marsh, E. N. G. Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 2017, 18 (19), 1888−1892.

(12) Cristie-David, A. S.; Chen, J.; Nowak, D. B.; Bondy, A. L.; Sun, K.; Park, S. I.; Banaszak Holl, M. M.; Su, M.; Marsh, E. N. G. Coiled- Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability. J. Am. Chem. Soc. 2019, 141 (23), 9207−9216.

(13) Malay, A. D.; Miyazaki, N.; Biela, A.; Chakraborti, S.; Majsterkiewicz, K.; Stupka, I.; Kaplan, C. S.; Kowalczyk, A.; Piette, B. M. A. G.; Hochberg, G. K. A.; Wu, D.; Wrobel, T. P.; Fineberg, A.; Kushwah, M. S.; Kelemen, M.; Vavpetič, P.; Pelicon, P.; Kukura, P.; Benesch, J. L. P.; Iwasaki, K.; et al. An Ultra-Stable Gold-Coordinated Protein Cage Displaying Reversible Assembly. Nature 2019, 569 (7756), 438−442.

(14) Lach, M.; Künzle, M.; Beck, T. Proteins as Sustainable Building Blocks for the Next Generation of Bioinorganic Nanomaterials. Biochemistry 2019, 58 (3), 140−141.

(15) Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments. Proc. Natl. Acad. Sci. U. S. A. 2001, 98 (5), 2217− 2221.

(16) Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. Annu. Rev. Biophys. 2017, 46 (1), 23−42.

(17) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. Science 2012, 336 (6085), 1171−1174.

(18) Ringler, P.; Schulz, G. E. Self-Assembly of Proteins into Designed Networks. Science 2003, 302 (5642), 106−109.

(19) Sinclair, J. C.; Davies, K. M.; Vénien-Bryan, C.; Noble, M. E. M. Generation of Protein Lattices by Fusing Proteins with Matching Rotational Symmetry. Nat. Nanotechnol. 2011, 6 (9), 558−562.

(20) Lai, Y. T.; Cascio, D.; Yeates, T. O. Structure of a 16-nm Cage Designed by Using Protein Oligomers. Science 2012, 336 (6085), 1129−1129.

(21) Kwon, N.-Y.; Kim, Y.; Lee, J.-O. The Application of Helix Fusion Methods in Structural Biology. Curr. Opin. Struct. Biol. 2020, 60, 110−116.

(22)Lai,Y.-T.;Jiang,L.;Chen,W.;Yeates,T.O.Onthe PredictabilityoftheOrientationofProteinDomainsJoinedbya SpanningAlpha-HelicalLinker.ProteinEng.,Des.Sel.2015,28(11), 491−500.

(23) Lai, Y.-T.; Reading, E.; Hura, G. L.; Tsai, K.-L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. a.; Robinson, C. V.; Yeates, T. O. Structure of a Designed Protein Cage That Self-Assembles into a Highly Porous Cube. Nat. Chem. 2014, 6 (12), 1065−1071.

(24) Cannon, K. A.; Nguyen, V. N.; Morgan, C.; Yeates, T. O. Design and Characterization of an Icosahedral Protein Cage Formed by a Double-Fusion Protein Containing Three Distinct Symmetry Elements. ACS Synth. Biol. 2020, 9 (3), 517−524.

(25) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. Accurate Design of Co- Assembling Multi-Component Protein Nanomaterials. Nature 2014, 510 (7503), 103−108.

(26) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes. Science 2016, 353 (6297), 389−395.

(27) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R.; Yi, S.; Davis, T. N.; Gonen, T.; King, N. P.; Baker, D. Design of a Hyperstable 60-Subunit Protein Icosahedron. Nature 2016, 535 (7610), 136−139.

(28) Ueda, G.; Antanasijevic, A.; Fallas, J. A.; Sheffler, W.; Copps, J.; Ellis, D.; Hutchinson, G. B.; Moyer, A.; Yasmeen, A.; Tsybovsky, Y.; Park, Y.-J.; Bick, M. J.; Sankaran, B.; Gillespie, R. A.; Brouwer, P. J.; Zwart, P. H.; Veesler, D.; Kanekiyo, M.; Graham, B. S.; Sanders, R. W.; et al. Tailored Design of Protein Nanoparticle Scaffolds for Multivalent Presentation of Viral Glycoprotein Antigens. eLife 2020, 9, No. e57659.

(29) Bale, J. B.; Park, R. U.; Liu, Y.; Gonen, S.; Gonen, T.; Cascio, D.; King, N. P.; Yeates, T. O.; Baker, D. Structure of a Designed Tetrahedral Protein Assembly Variant Engineered to Have Improved Soluble Expression. Protein Sci. 2015, 24 (10), 1695−1701.

(30) Boyken, S. E.; Chen, Z.; Groves, B.; Langan, R. A.; Oberdorfer, G.; Ford, A.; Gilmore, J. M.; Xu, C.; DiMaio, F.; Pereira, J. H.; Sankaran, B.; Seelig, G.; Zwart, P. H.; Baker, D. De Novo

Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network− Mediated Specificity. Science 2016, 352 (6286), 680−687.

(31) Cannon, K. A.; Park, R. U.; Boyken, S. E.; Nattermann, U.; Yi, S.; Baker, D.; King, N. P.; Yeates, T. O. Design and Structure of Two New Protein Cages Illustrate Successes and Ongoing Challenges in Protein Engineering. Protein Sci. 2020, 29 (4), 919−929.

(32) Miyamoto, T.; Hayashi, Y.; Yoshida, K.; Watanabe, H.; Uchihashi, T.; Yonezawa, K.; Shimizu, N.; Kamikubo, H.; Hirota, S. Construction of a Quadrangular Tetramer and a Cage-Like Hexamer from Three-Helix Bundle-Linked Fusion Proteins. ACS Synth. Biol. 2019, 8 (5), 1112−1120.

(33) McConnell, S. A.; Cannon, K. A.; Morgan, C.; McAllister, R.; Amer, B. R.; Clubb, R. T.; Yeates, T. O. Designed Protein Cages as Scaffolds for Building Multienzyme Materials. ACS Synth. Biol. 2020, 9 (2), 381−391.

(34) Phippen, S. W.; Stevens, C. A.; Vance, T. D. R.; King, N. P.; Baker, D.; Davies, P. L. Multivalent Display of Antifreeze Proteins by Fusion to Self-Assembling Protein Cages Enhances Ice-Binding Activities. Biochemistry 2016, 55 (49), 6811−6820.

(35) 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, K. K.; Veesler, D.; Correnti, C. E.; Stewart, L. J.; et al. Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. Cell 2019, 176 (6), 1420−1431e17.

(36) Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U.; Ellis, D.; Bale, J. B.; Ke, S.; Lenz, G. H.; Yehdego, A.; Ravichandran, R.; Pun, S. H.; King, N. P.; Baker,

D. Evolution of a Designed Protein Assembly Encapsulating Its Own RNA Genome. Nature 2017, 552 (7685), 415−420.

(37) Edwardson, T. G. W.; Mori, T.; Hilvert, D. Rational Engineering of a Designed Protein Cage for SiRNA Delivery. J. Am. Chem. Soc. 2018, 140 (33), 10439−10442.

(38) Liu, Y.; Gonen, S.; Gonen, T.; Yeates, T. O. Near-Atomic Cryo- EM Imaging of a Small Protein Displayed on a Designed Scaffolding System. Proc. Natl. Acad. Sci. U. S. A. 2018, 115 (13), 3362−3367.

(39) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo- EM Structure of a Small Protein Bound to an Imaging Scaffold. Nat. Commun. 2019, 10 (1), 1864.

(40) Fletcher, J. M.; Boyle, A. L.; Bruning, M.; Bartlett, G. J.; Vincent, T. L.; Zaccai, N. R.; Armstrong, C. T.; Bromley, E. H. C.; Booth, P. J.; Brady, R. L.; Thomson, A. R.; Woolfson, D. N. A Basis Set of de Novo Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. ACS Synth. Biol. 2012, 1 (6), 240−250.

(41) O'Shea, E. K.; Lumb, K. J.; Kim, P. S. Peptide 'Velcro': Design of a Heterodimeric Coiled Coil. Curr. Biol. 1993, 3 (10), 658−667.

(42) Gradišăr, H.; Bozǐc,̌ S.; Doles, T.; Vengust, D.; Hafner-Bratkovic,̃ I.;Mertelj,A.;Webb,B.;Šali,A.;Klavzăr,S.;Jerala,R. Design of a Single-Chain Polypeptide Tetrahedron Assembled from Coiled-Coil Segments. Nat. Chem. Biol. 2013, 9 (6), 362−366.

(43) Reinke, A. W.; Grant, R. A.; Keating, A. E. A Synthetic Coiled- Coil Interactome Provides Heterospecific Modules for Molecular Engineering. J. Am. Chem. Soc. 2010, 132 (17), 6025−6031.

(44) Ogihara, N. L.; Weiss, M. S.; Eisenberg, D.; Degrado, W. F. The Crystal Structure of the Designed Trimeric Coiled Coil Coil-VaLd: Implications for Engineering Crystals and Supramolecular Assemblies. Protein Sci. 1997, 6 (1), 80−88.

(45) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. A Designed Heterotrimeric Coiled Coil. Biochemistry 1995, 34 (37), 11645− 11651.

(46) Glover, J. N. M.; Harrison, S. C. Crystal Structure of the Heterodimeric BZIP Transcription Factor C-Fos−c-Jun Bound to DNA. Nature 1995, 373 (6511), 257−261.

(47) O'Shea, E. K.; Rutkowski, R.; Stafford, W. F.; Kim, P. S. Preferential Heterodimer Formation by Isolated Leucine Zippers from Fos and Jun. Science 1989, 245 (4918), 646−648.

(48) Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; Brubaker, M. A. CryoSPARC: Algorithms for Rapid Unsupervised Cryo-EM Structure Determination. Nat. Methods 2017, 14 (3), 290−296.

(49) Heinig, M.; Frishman, D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. Nucleic Acids Res. 2004, 32, W500−W502.

(50) Studier, F. W. Protein Production by Auto-Induction in High- Density Shaking Cultures. Protein Expression Purif. 2005, 41 (1), 207−234.

(51) Kabsch, W. XDS. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2010, 66 (2), 125−132.

(52) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. J. Appl. Crystallogr. 2007, 40 (4), 658−674.

(53) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2004, 60 (12), 2126−2132.

(54) Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. PHENIX:

Building New Software for Automated Crystallographic Structure Determination. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2002, 58 (11), 1948−1954.

(55) Scheres, S. H. W. RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination. J. Struct. Biol. 2012, 180 (3), 519−530.

CHAPTER THREE

DESIGNING PROTEASE-TRIGGERED PROTEIN CAGES

3.1 Abstract

Proteins that self-assemble into enclosed polyhedral cages, both naturally and by design, are garnering attention for their prospective utility in the fields of medicine and biotechnology. Notably, their potential for encapsulation and surface display are attractive for experiments that require protection and targeted delivery of cargo. The ability to control their opening or disassembly would greatly advance the development of protein nanocages into widespread molecular tools. Toward the development of protein cages that disassemble in a systematic manner and in response to biologically relevant stimuli, here we demonstrate a modular protein cage system that is opened by highly sequence-specific proteases, based on sequence insertions at strategically chosen loop positions in the protein cage subunits. We probed the generality of the approach in the context of protein cages built using the two prevailing methods of construction: genetic fusion between oligomeric components and (non-covalent) computational interface design between oligomeric components. Our results suggest that the former type of cage may be more amenable than the latter for endowing proteolytically controlled disassembly. We show that a successfully designed cage system, based on oligomeric fusion, is modular with regard to its triggering protease. One version of the cage is targeted by an asparagine protease implicated in cancer and Alzheimer's disease, whereas the second version is responsive to the blood-clotting protease, thrombin. The approach demonstrated here should guide future efforts to develop therapeutic vectors to treat disease states where protease induction or mis-regulation occurs.

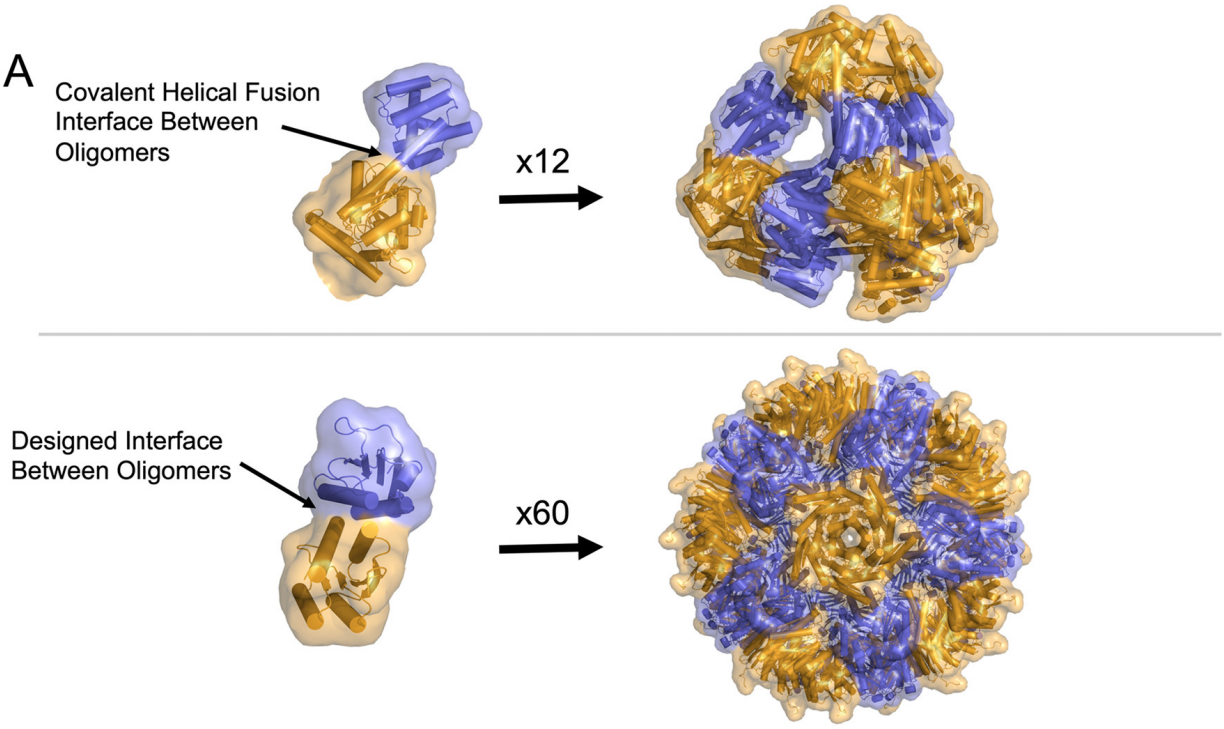
3.2 Introduction

Natural protein assemblies underlie many of the biological functions required to sustain life. The evolution of biochemically complex processes, from protein synthesis to cellular

motility, has been advanced by the capacity of proteins to assemble into complex molecular machines. (1) Nanocages represent a class of protein assemblies typically characterized by highly symmetric shell-like architectures, enclosing an internal lumen space. Diverse examples are seen in nature, including viral capsids, bacterial microcompartments, and ferritins[1–4]. Taking inspiration from these natural assemblies, emerging bioengineering methods have made it possible to design novel nanocage assemblies with new and rich functionality de novo[5–11]. Among diverse chemical types that have been explored for nanotechnology applications, designed materials based on protein molecules offer special benefits in terms of chemical and functional complexity, as well as biocompatibility.

While nature has enjoyed the benefits of time in evolving complex protein cages and containers, engineering comparable structures in the laboratory has been enabled by careful exploitation of principles of symmetry. Bringing simpler oligomeric building blocks together in precise combinations and orientations can produce designed protein cages with atomically defined structures of great size and complexity[12–17]. Engineering multiple protein molecules to come together in a precisely defined fashion is a challenging endeavor, and indeed experimental success rates in designing protein cages remain modest, yet two main methods have proven effective. In the first approach, two simpler (symmetric) oligomeric components are brought together by genetic fusion, with the larger resulting protein subunit inheriting the symmetry elements from both of its parts; connecting components bearing helical termini with a continuous alpha helical linker between them improves the ability to control geometric outcomes[12,18–24,24]. In the second approach, computational methods are used to infer amino acid surface mutations that will drive two oligomeric components together based on surface-matching and specific non-covalent forces[14,16,25]. We refer to those cage types here as fusion-based and interface-based (as

diagrammed in Figure 3.1A). These methods have produced numerous extraordinary novel assemblies, with applications aimed at both internal encapsulation[6,14,26,27] and polyvalent display[28–31], yet with few exceptions those efforts have focused on outcomes that are mainly static in nature.



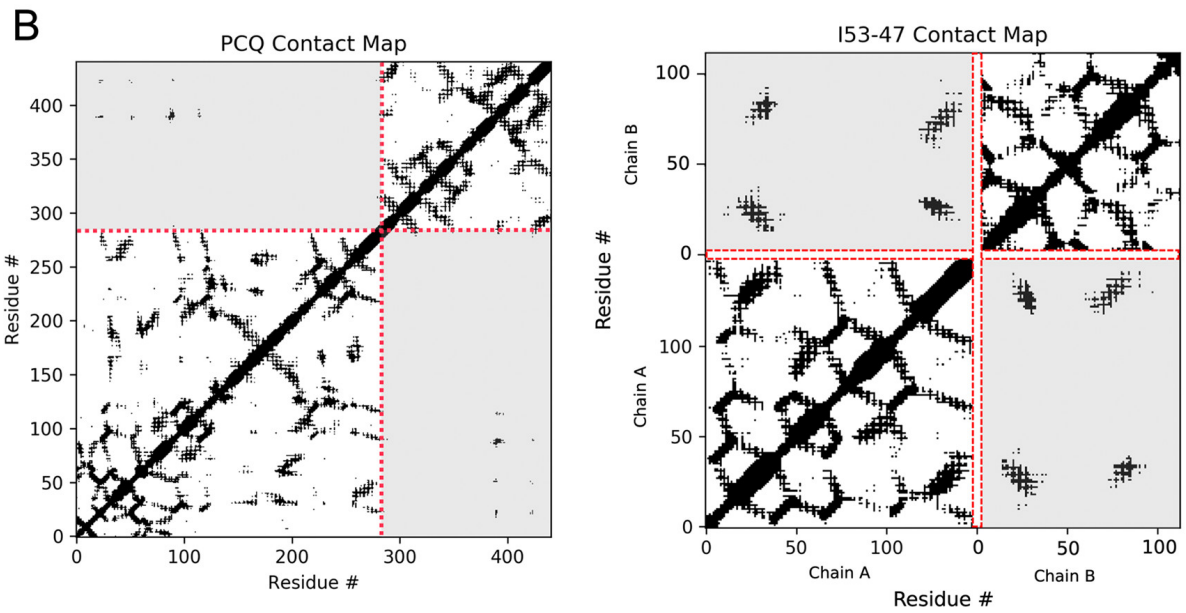


Figure 3.1. Two distinct methods of protein cage design, highlighting differences in how the two oligomeric components are held together. (A, top) In the helix-fusion approach, (12) the two oligomeric protein components (blue and gold) are genetically fused into a single protein chain, preferably by a continuous alpha helical linker. The example shown, known as PCQ and subject to further design in the present study, is a combination of a dimer and trimer to form a tetrahedrally symmetric cage (12 subunits). (A, bottom) In the interface approach, (25) the two oligomeric components are held together by non-covalent forces based on a computationally designed interface. The example shown, known as I53-47 and subject to further design in the present study, is a combination of a trimer and a pentamer to form an icosahedral cage (60 copies of an A and B subunit). (B) Residue–residue contact maps for PC-Q and I53-47. The light gray regions depict contacts between the two oligomeric components. Note that for the fusion-type cage (PCQ, left), cleavage near the indicated position produces fragments with few contacts (in the off-diagonal rectangular region). In contrast, for the interface-type cage (I53-47, right), there are no cleavage positions that produce fragments free of substantial interactions.

Toward the development of dynamically controllable protein cages, more reminiscent of functional protein systems in biology, early studies have exploited sensitivity to bulk environmental changes such as pH, metal chelation, and oxidation or reduction. The well-known pH sensitivity of certain amino acid side chains (e.g., histidine) to ionization can be exploited at a protein–protein interface, with the repulsion of like-charges promoting disassembly. Several studies have demonstrated systems to respond to acidification, mimicking certain viral capsids [32,33]. Other studies have exploited disulfide bonds to chemically link protein subunits, conferring stability in oxidizing environments outside the cell but not in intracellular environments, which are typically reducing[33–36]. Similarly, engineered metal-mediated interactions between proteins can be reversibly broken by chelators of specific metal ions[11,37–40], though connections to the cell state are less direct. In other studies, control of protein assembly by light and by ligand binding have been explored[10,41]. With DNA nanotechnology, a cage opened by ligand binding has been described[42]. These efforts highlight the potential utility of designing cages with dynamic or controllable assembly properties.

Regarding potential biomedical applications for designed protein cages, a major advance would come from systems that respond to highly specific disease biomarkers and factors of clinical relevance. In the field of therapeutic design, proteases represent important players because of the frequency with which they appear as disease biomarkers, including in cancer[43]. To exploit this, efforts have been made to develop drugs that target the activity of these proteases, for example, by activating inactive prodrugs bearing specific sequence motifs by cleavage in cellular environments expressing high levels of a specific protease[44]. Inspired by these observations, in this study, we seek to leverage the idea of protease-specific activation, now with application to designed protein cages. We present a novel system wherein designed protein nanocages respond to the sole input signal of a sequence-specific protease with efficient disassembly. We refer to these as protease-triggered protein cages (PTPCs). Parallels can be drawn between these engineered systems and diverse viral capsids, where proteolysis may be crucial for assembly and maturation, or disassembly in the acidic environment of the endosome[45,46]

We further investigate the applicability of this engineering strategy to protein cages built according to different design principles. Given the clinical relevance of proteases as biomarkers for various disease states, our findings set the stage for future work to develop therapeutic nanocage systems tailored to respond under conditions of specific protease upregulation.

3.3 Results

Design Principles for PTPCs

To render an otherwise unaffected protein subunit susceptible to a sequence-specific protease, a sequence recognized by the protease must be inserted into the protein. For reasons of conformational accessibility, loop regions between secondary structure elements are the natural

sites for such insertions. Such loops must face outward on the cage surface for reasonable cleavage efficiency. In addition, more nuanced considerations arise. For a chosen site to be effective, the two molecular fragments created by cleavage should prefer to separate, for example, by gaining the entropy associated with free movement. However, the inserted amino acid sequence, however long it might be, must not promote such separation in the absence of cleavage; the two parts must remain associated despite the increased entropy allowed by the reduced spatial constraint of the inserted segment. The underlying design goal for the present study is therefore subject to a potentially narrow thermodynamic window where the conditions noted above might be met.

The success of a candidate design must depend on the extent of the intermolecular forces that remain holding the covalently distinct parts of the protein(s) together after a proteolytic cleavage. We reasoned that the two different types of designed protein cages that have been developed─fusion-based and interface-based─might present different scenarios with respect to the issue of stabilizing intermolecular interactions since the protein domains are held together in very different ways in those two cases (Figure 3.1A). Intuition suggests that, for a fusion-based cage, there could be sites near the connecting joint whereupon cleavage would produce two molecular fragments with relatively modest contacts between them. (N.B. The limiting case would be cleavage directly in the connecting segment, with no substantial contacts at all between the resulting parts, but the sequence of the connecting segment is generally not a suitable choice for proteolytic sequence insertion since preserving the sequence and structure of the connecting segment is critical for ensuring correct cage assembly at the outset]. On the contrary, for an interface-based cage, where the components are held together by numerous non-covalent

interactions dispersed throughout the linear sequence of the protein, there may be no single site that would produce fragments held together by only a few interactions.

Residue–residue pairwise contact maps help illustrate the relationship between a candidate cleavage site in a protein chain and the remaining intermolecular forces between resulting fragments. Figure 3.1B shows contact maps for representatives of the two protein cage types. Shaded areas depict the region corresponding to interactions between original oligomeric components that make up each cage type (i.e., dimer and trimer for the fusion-based cage known as PCQ, or trimer and pentamer for the interface-based cage known as I53-47). Predictably, far fewer interactions are present between components making up the fusion-based cage (PCQ) compared to the number between the components making up interface-based cage (I53-47) (Figure 3.1B). This type of analysis provides a framework for the experimental tests in our study. It also helps frame a hypothesis that fusion-type cages might be more amenable subjects than interface-type cages for developing robust PTPCs. We note, however, that such contact maps reflect static structural views and do not capture important molecular interactions that might occur after proteolysis.

Design and Testing of Thrombin-Targeted PTPCs

As subjects for our engineering and experimental studies, we selected protein nanocages covering the two types noted above. For the fusion-type, we employed a tetrahedral (12 subunit) cage we refer to here as PCQ (following the moniker Padilla cage quad mutant in earlier papers [47]). This design fuses a trimeric component to a dimeric component by an alpha helical connection. For the interface-type cage, we selected two closely related cages, referred to as I53-47 and I53-50, which are icosahedral assemblies based on a designed interface that holds together a pentameric and trimeric component; these two cages are designed from a common

pentameric component with unique trimeric components differentiating the two cages [14]. These cages were selected for the present redesign studies based on the expression levels, solubility, and stability of the original cage designs.

We based our first set of PTPC designs on thrombin cleavage, motivated by the clinical relevance and efficient cleavage associated with that enzyme. We used the amino acid sequence (LVPRGS) as the thrombin recognition sequence. A minimal linker was used to flank the protease site for first tests, with additional residues added in some experiments to improve the accessibility where necessary (Supporting Information, Table S3.1). Based on analysis of the three-dimensional structure of the cages and intermolecular contacts as discussed above, five thrombin-based sequence insertion locations were investigated for the PCQ cage (cage designs designated PCQth140, PCQth256, PCQth302*, PCQth336, and PCQth353*), and seven were investigated for the I53-47 and I53-50 cages (designated 47thPENT41*, 50thTRI45*, 50thTRI63, 50thTRI125*, 50thTRI150*, 50thPENT43*, and 50thPENT67) along with a mutant containing two insertions (denoted 50thPENT43_67*) (Supporting Information, Figures S3.1 & S3.5). Of these designs, two of the PCQ cage designs and five of the designs based on I53-47 and I53-50 cages were successfully cloned, expressed, and characterized biochemically (denoted with asterisks above); other cases presented difficulty in cloning or expressing soluble proteins.

Using synthetic DNA sequences encoding the designed amino acid sequences, protein cage mutants were expressed in Escherichia coli using inducible vectors under the control of the lac promoter. All designs encoded a terminal 6xHistidine tag on the protein subunit (or on one of the two unique subunits for the two-component interface design cases) to enable metal affinity purification. SDS-PAGE was used to confirm the protein composition of the purified cages, whereas size exclusion chromatography (SEC) and negatively stained electron microscopy were

used to assess correct assembly of the modified proteins into cages as intended (Figures 3.2 and 3.3, Supporting Information, Figures S3.2 and S3.3). In some cases, minor sequence variations were made and analyzed in order to arrive at constructions producing intact protein cages in good yield (Supporting Information, Table S3.1). Versions of PCQ (PCQth302), I53-47 (47thPENT41), and I53-50 (50thTRI45, 50thTRI125, and 50thPENT43) that formed assemblies of the correct size, as assessed by SEC or dynamic light scattering (DLS), were selected for subsequent protease experiments (Supporting Information, Figures S3.2 and S3.3).

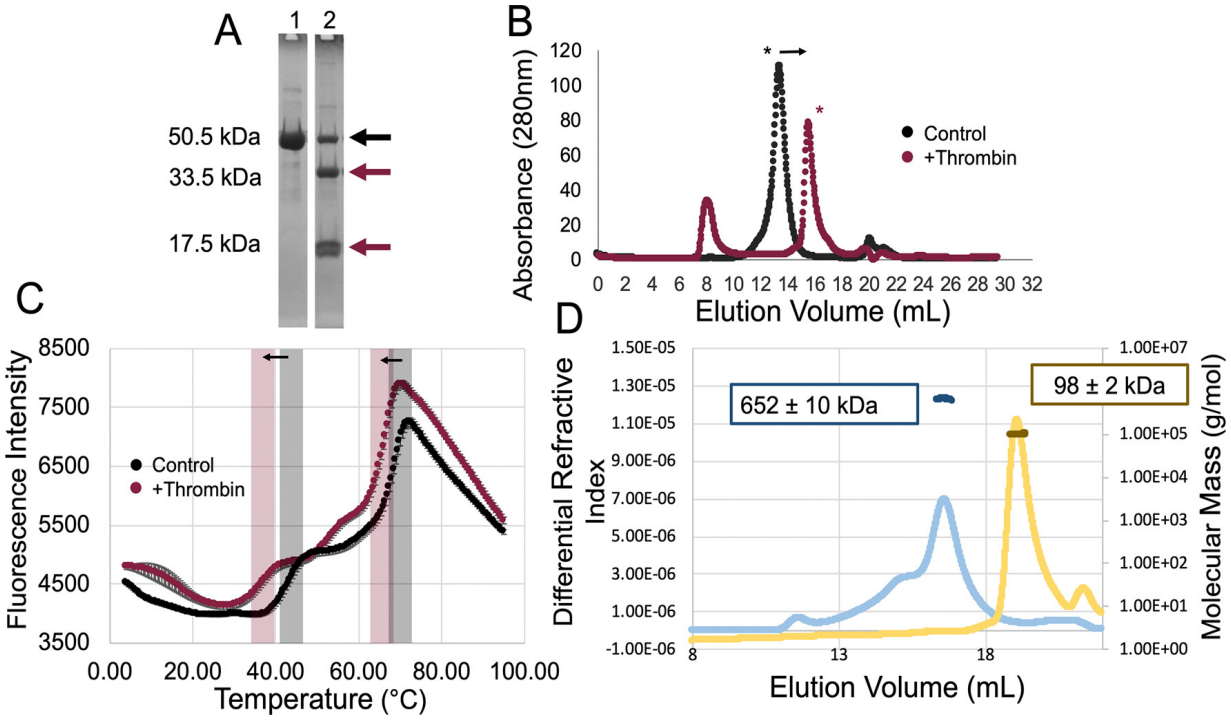


Figure 3.2. Cleavage and disassembly of fusion-type cage PCQth302 by the protease thrombin. (A) SDS-PAGE indicates a high cleavage efficiency of the protein nanocage by thrombin at the designed site by the appearance of degradation products (red arrows) of expected molecular weight (lane 2). The uncleaved control is shown in lane 1 for comparison. (B) SEC shows essentially complete conversion of the intact cage (black asterisk) to a smaller degradation product (red asterisk). (C) Thermal shift assay showing control conditions (black traces) and the effects of adding thrombin (red) demonstrate a destabilizing effect on the assembly, and exposure of hydrophobic protein surfaces, when thrombin is introduced. The conserved peaks are highlighted in red and black for experimental and control conditions, respectively. (D) SEC-MALS of PCQth302 indicates that the molecular weight of the control condition (650 kDa) agrees well with the predicted molecular weight of the assembled cage (610 kDa), while thrombin treatment produces mainly smaller 98 kDa species.

Cleavage and Characterization of Fusion-Based Cages

We first investigated the effect of incubating the fusion-based cage designs with the triggering protease thrombin. Protease reactions were performed on purified cage samples. By incubating each of the mutants with thrombin, we identified designs that produced degradation products of expected molecular weight, as determined by SDS-PAGE analysis. In investigating PCQth302, we observed conversion of the 50.8 kDa intact protein chain into degradation products of 33.5 and 17.5 kDa, which agree with the expected molecular weights of protein fragments cut at the inserted canonical thrombin cleavage site (Figure 3.2A). Note that for the PCQth302 design, the cleavage position was relatively close to the fusion point between oligomeric components (Figure 3.1B). PCQth302 was cleaved with an efficiency of approximately 84% based on densitometric analysis of the SDS-PAGE gel after 48 h at 37 °C. Interestingly, we never observed 100% cleavage efficiency, possibly due to aggregation induced by partial cage disassembly and concomitant occlusion of protease cleavage sites.

We characterized the effects of proteolytic cleavage using various techniques. In samples where thrombin was added to PCQth302, we visualized a marked shift to a later elution volume (15.6 mL) by SEC when compared with control cages incubated under the same conditions but without thrombin, which eluted earlier (13.4 mL) (Figure 3.2B). This observation suggests a conversion of the protein cage assembly into smaller subunit components. We also observed the appearance of an early eluting peak, corresponding to elution in the void volume of the column used. This likely corresponds to aggregated protein that resulted from changes in the solubility of PTPCs after cleavage by thrombin. With the goal of better characterizing the timescale of cage destruction, we performed DLS on cage samples at time points after addition of thrombin. We

observed that the conversion of cage into smaller components happened under an hour after thrombin was added, requiring only 20% of protein to be proteolyzed (Supporting Information, Figure S3.6). We also investigated the effect of the proteolysis on the stability of the cage proteins in solution using thermal shift assay, where increased fluorescence indicates unfolding or exposure of hydrophobic protein surfaces in solution. In proteolyzed samples, we observed a global shift in the trace to lower temperatures (Figure 3.2C). This result indicates that the proteolyzed cage sample was destabilized by the cleavage event and underwent denaturation earlier in the protocol than did unproteolyzed samples. Lastly, to more precisely investigate the molecular weights of protein species occurring after proteolysis, we subjected samples to SEC with inline multi-angle light scattering (SEC-MALS). Analysis of peaks eluted from SEC from the control experiment without protease resulted in molecular mass calculations, in close agreement with the predicted molecular weight of the assembled tetrahedral cage (652 kDa vs 610 kDa) (Figure 3.2D). In control samples, the more-prominent earlier-eluting shoulder might be explained by the observation that the original cage design (PC-quad) exists in an equilibrium of compact and expanded states of the protein assembly[47]. In analyzing the proteolyzed sample, the estimated molecular weight of the dominant species in solution (97.6 kDa) was very close to the calculated molecular weight of the trimeric degradation product calculated based on the predicted cut of thrombin in the design model (99.5 kDa) (Figure 3.2D). Another less intense peak eluted later from the proteolyzed sample (Figure 3.2D), possibly corresponding to the smaller dimeric degradation product that would result from thrombin cleavage, but light scattering intensity was insufficient to calculate molecular weight. Together the results demonstrate that the PCQth302 cage is readily disassembled by thrombin proteolysis, as intended.

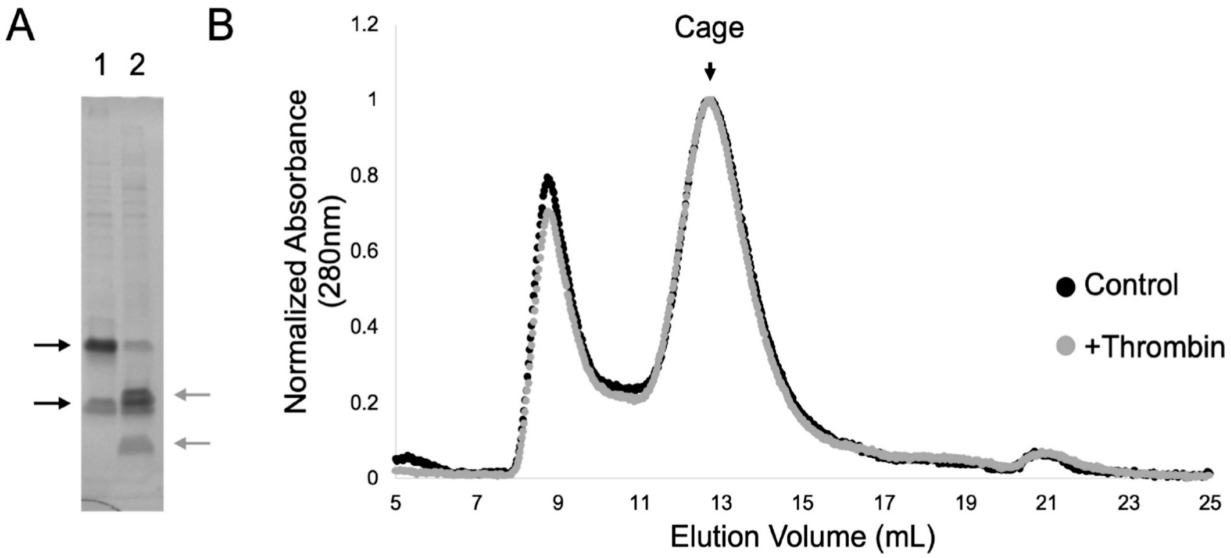


Figure 3.3. Failure of the interface-type cage 47th_pent41 to disassemble after digestion with thrombin. (A) SDS-PAGE (control in lane 1 and thrombin-digested in lane 2) indicates that nearly complete cleavage is achieved. Intact components are indicated by black arrows. Degradation products are indicated by gray arrows. (B) SEC of thrombin-treated 47th_pent41 (grey trace) shows that the assembled form of the cage (black arrow) remains essentially unaffected by cleavage. The non-proteolyzed control curve is shown in black for reference. The elution peak at roughly 8 mL represents the void volume.

Cleavage and Characterization of Interface-Based Cages

For experiments involving the interface-based cages, initial designs with minimal (GS) linkers flanking the protease recognition sequence were not cut efficiently by thrombin in designs based on the I53-50 cage, so we investigated longer linker lengths. The loop insertion positions tested for this series of designs are shown in Supporting Information (Figure S3.3). After observing high yield and abundance of correctly assembled cage in 50thPENT43, we also tested a similar assembly 47th_PENT41 employing the same pentamer and pentamer insertion site (but a unique trimer and designed interface) to expand our design space. These trials resulted in one design (47thPENT41_LL) that could be efficiently cut by thrombin (Figure 3.3). Purified

samples of the linker-lengthened design were incubated with thrombin. Densitometry of SDS-PAGE gels indicated that a cleavage efficiency of roughly 76% was obtained for the protein chain bearing the engineered cleavage site (Figure 3.3A); the interface-type cages here are built from two distinct protein subunit types, only one of which bears the cleavage site. However, despite efficient cleavage evident by denaturing SDS-PAGE, analysis of native species (in non-denaturing solution conditions) by SEC indicated no discernible disassembly of the cage (Figure 3.3B). Attempts to engineer multiple protease cleavage sites into the same protein cage did not result in intact cage assemblies with appreciable yields (Supporting Information, Figure S3.3).

Generating a PTPC Targeted by Other Proteases

Based on the successful disassembly of the fusion-based cage, PCQth302, we next sought to test the modularity of this system by generating variants that could be targeted by other disease-associated proteases. Three new designs were generated with protease recognition sites at the same location as PCQth302. These new mutants were targeted by matrix-metalloproteinase 9 (MMP9), recognizing the amino acid sequence "GPLGIAGQ,"[48] furin, recognizing the sequence "RGRKR,"[49] and asparagine endopeptidase (AEP), recognizing the sequence "GAANG,"[50] named PCQmmp302, PCQfur302, and PCQaep302, respectively. The protease types were selected on the basis of their relatively small cleavage sequences, well-characterized activities, and clinical relevance as disease biomarkers. Of the three designs tested, the one based on an inserted furin cleavage site (PCQfur302) did not assemble into a cage of expected size (Supporting Information, Figure S3.4). The variant based on metalloprotease (PCQmmp302) did assemble correctly, but we could not achieve efficient cleavage by protease treatment (Supporting Information, Figure S3.4b). However, the variant based on insertion of an AEP site (PCQaep302) assembled correctly. AEP is an endosomal protease, which is active and stable

under acidic conditions, so we next assayed the stability of our cage under conditions amenable to AEP activity and found the cage correctly assembled, and was efficiently cut, producing degradation products of the expected molecular weight (Figure 3.4A). Densitometric analysis of the SDS-PAGE gel revealed a cleavage efficiency of 65% (Figure 3.4A), while peak heights in SEC showed that intact cages were only a minority species after proteolysis, with smaller species predominating (Figure 3.4B). This result shows that a given PTPC system can be adapted for response to different proteases, following relatively minor optimization, in some cases.

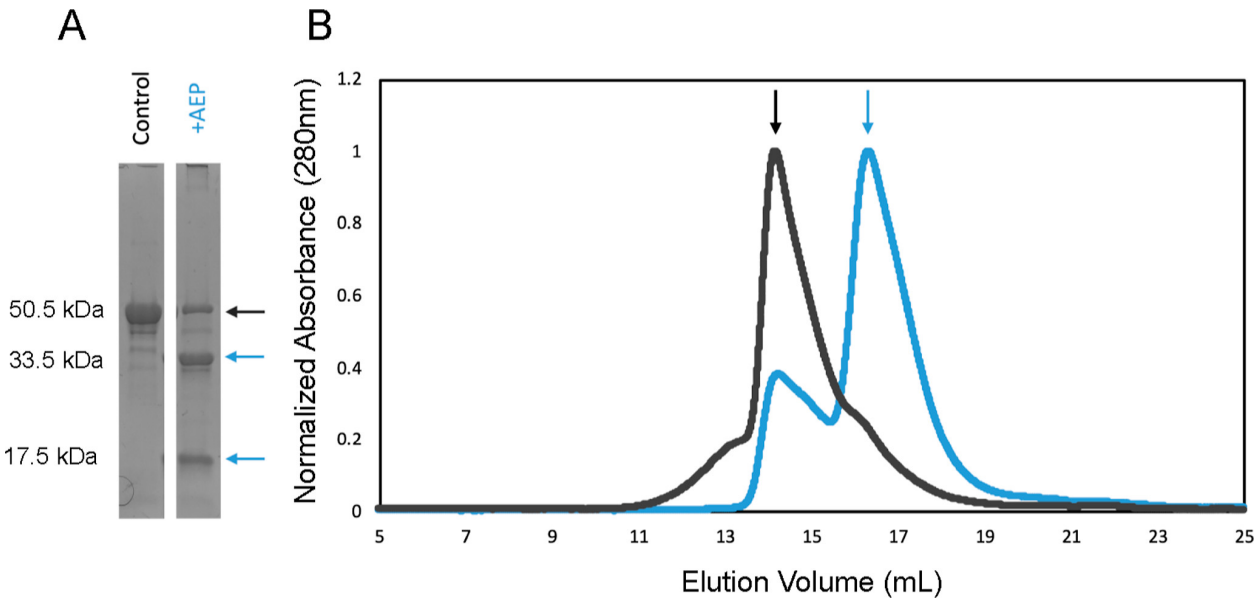


Figure 3.4. Demonstration of triggered disassembly by a second protease, AEP. The PCQ cage was modified to present an AEP cleavage site instead of the original thrombin site. (A) SDS-PAGE of intact PCQaep302 purified from Ni-NTA (left) and PCQaep302 digested with AEP protease (right). The intact protein species (50.5 kDa) is shown with a black arrow. Degradation products of 33.5 and 17.5 kDa are shown by blue arrows. (B) SEC of control (black) and digested (blue) PCQaep302, showing that the assembled form of the cage is largely destroyed by AEP cleavage.

3.4 Discussion and Conclusions

The goal of the present study was to demonstrate an approach for creating protein nanocages capable of being disassembled under highly specific, therapeutically relevant, conditions. To this end, we demonstrated a nanocage system wherein controlled disassembly of

two minor design variants occurs in response to two different, highly specific protease types. The successful result here is notable in that protein cage disassembly is triggered by a highly specified stimulus, connectable to the cell or tissue state in various diseases or clinical settings, rather than a global environmental change. Additionally, the protease trigger used for disassembly here is modular in that a single starting design can be diversified for multiple distinct applications based on relatively straightforward sequence mutations in favorable cases. For the protein cage system we chose as a starting point, only a subset of design choices for a cleavage site proved suitable, showing that systematic experimental studies are generally required to develop robust PTPCs. Similarly, only a subset of protease types we attempted (two of the four) gave successful results in our experiments, showing that more substantial efforts, or alternative protein cage systems, may be needed to achieve success in particular instances. This observation is not altogether surprising when considering that many proteases have some degree of structural dependence to their activity. Furthermore, variations in the active site conformation between distinct proteases may require differing degrees of cleavage site exposure and necessitate some degree of tailoring the design to the protease of interest.

A comparison of outcomes for different kinds of protein cages suggests that the nature and extent of the intermolecular interactions holding the protein subunits together helps dictate whether a protein cage will readily disassemble after targeted proteolysis. A protein cage designed on the basis of genetic fusion of two symmetric oligomers could be decoupled into separable parts via protease cleavage near the location of the linking polypeptide segment. In contrast, we found that cages constructed on the basis of computational interface design, with their more extensive protein–protein interfaces, were more resistant to separation and disassembly following protease treatment, across several experimental choices for the cleavage

position. This was consistent with expectations from analysis of contact patterns using known three-dimensional structures (Supporting Information, Figure S3.1). Although we were not successful in designing a protease-sensitive cage using an interface-based designed cage, our experimental trials do not exhaust the possibilities for creating diverse PTPCs. Analysis of other interface-based cages could identify cases more amenable to controlled proteolytic disassembly. Interestingly, by virtue of their extensive subunit interfaces, the cages based on computational interface design more closely resemble natural viral capsids. An important distinction between our design efforts and evolved protease-responsive sequences found in natural viruses is that our design process requires identifying sites for insertions that would be unlikely to disrupt the cage structure without proteolysis (i.e., site where insertion by itself is not unduly destabilizing). The natural consequence is that cutting at such sites may not always lead to disassembly. The challenge of these opposing concerns in the context of design provides insight into the exquisite nature of evolved proteolytic maturation or disassembly mechanisms in some natural capsids.

In summary, the present findings should guide future efforts to create diverse protein materials that respond to specific cellular stimuli. In particular, we note that the fusion cage tested here[18] may not be especially well-suited for downstream delivery applications, owing to the large size of its pores or windows and the disposition of its chain termini (e.g., for making fusions to bioactive sequences). Even so, applications where activity depends on exposure or release of an internally attached bioactive molecule could be envisioned for such cages. Fusion-type cages were the first created in the laboratory, but successful designs in this category have been limited, chiefly by the challenges of geometric flexibility between joined components. However, the findings of the current study motivate efforts to create additional cages of the fusion-type, specifically focusing on designing systems more tailored for future delivery

applications. Indeed, forms of oligomeric fusion more advanced than the type first developed by Padilla et al.[12] have found utility in recent cage design studies[24,41], opening up broader possibilities for PTPCs.

3.5 Methods

Computational Design

Protein cages were engineered using the crystal structure of the one-component fusion-type tetrahedral protein cage previously designed in our laboratory as a starting model (PDBID 4QES). Protein sequence sites on the surface of protein cages were assessed for engineering potential on the basis of three criteria: secondary structure using assignment by STRIDE[51], solvent-accessible surface area[52], and predicted protease accessibility. During this process, loop regions were favored. Solvent-accessible surface area was quantified and used to rank design candidates to ensure that insertion sites were not buried. Finally, protease accessibility was assessed visually by docking the structure of a globular protease into the insertion site, in hopes of identifying potential clashes that would disallow enzyme activity. Canonical protease cleavage sites were then inserted in sequences on the exterior surface of protein cages based on these criteria.

Gene Synthesis

Codon-optimized gene sequences were ordered from Integrated DNA Technologies or Twist Biosciences with overlapping sequences corresponding to flanking regions around the hindIII and ndeI restriction sites in the pET-22b expression vector. Intergenic sequences for two-component designs were taken from a pETDuet-1 expression plasmid and ordered as a single gene fragment.

Protein Expression and Purification

Designs were cloned into pET-22b expression vectors using Gibson assembly, and mutagenesis and alternate insertions made using established PCR mutagenesis procedures[53]. Genes were verified using Sanger sequencing. Small-scale expression was performed in BL21(DE3) cells grown in 200 mL of cultures using auto-induction media grown for 24 h at 25 C. PCQ-based designs were lysed in 1X phosphate-buffered saline (PBS) pH 7.4 supplemented with 5 mM 2-mecaptoethanol and EDTA-free protease inhibitor tablets (Thermo Fisher Scientific) using an Emulsiflex C3 homogenizer and affinity purified using Ni-NTA agarose resin (Thermo Fisher Scientific) in a gravity flow column. Interface-type cages were lysed in 50 mM Tris–HCl (pH 8.0) with 250 mM NaCl, 5 m M 2-mecaptoethanol, and EDTA-free protease inhibitor tablets and were lysed and purified in the same way. Protein was washed with lysis buffer +100 mM imidazole and eluted in lysis buffer +500 mM imidazole. Eluted protein was dialyzed against imidazole overnight at 4 °C. Protein was then mixed with 1U thrombin/mg protein cage, and samples were run on SDS-PAGE to assess the cleavage efficiency. Designs that were cut by thrombin were screened for dissociation by SEC using a Superose-6 Increase column (Cytiva Life Sciences) attached to a NGC FPLC (Bio-Rad Laboratories).

Thermal Shift Assay

2 µg of protein cage was added to opaque white PCR strips with optically clear caps. SYPRO orange at 5X working concentration was added to each sample, thrombin was added to samples that required it, and samples were held at 4 C until placed in a q-peltier thermocycler (Bio-Rad Laboratories). Samples were warmed at a rate of 0.5 °C/min from 4 to 95 °C. Measurements were taken every 1 min.

SEC-MALS Experiments

For experiments involving SEC-MALS measurements, SEC-purified protein cage was concentrated to >1 mg/mL. The concentrated, filtered samples were injected using an AKTA-pure instrument equipped with a Superose 6 Increase column at 25 °C (Cytiva Life Sciences) with inline miniDAWN multi-angle light-scattering and refractive index instruments (Wyatt Technology). Data analysis was performed using ASTRA software (Wyatt Technology).

Dynamic Light Scattering

SEC-purified cage proteins were assayed in a 96-well plate using a Wyatt DynaPro plate Reader (Wyatt Technology). Data analysis was performed and figures generated using Wyatt Technology software.

Protease Digestion

Thrombin was purchased from Sigma-Aldrich and stored in buffer containing 30% glycerol, 50 mm Tris–HCl pH 8.0, 100 mm NaCl, and 10 mm CaCl2. Thrombin cleavage reactions took place in 1 × PBS buffer. Recombinant human AEP was purchased from R&D systems and activated for 2 h at 37 °C in 50 mM sodium acetate (pH 4.0) and 100 mM sodium chloride prior to use. AEP-sensitive protein cage digestion was assayed in 50 mM 2-(N-morpholino)ethanesulfonic acid (MES) (pH 5.0) and 300 mM sodium chloride for 48 h at 37 °C.

Negative Stain Electron Microscopy

Protein cage samples were taken from SEC purification and applied to glow-discharged carbon-coated copper grids (Ted Pella). 4.5 μL of the sample was applied to the carbon side of the grid and allowed to rest for 1 min. The sample was then wicked away using the filter paper and blotted with a drop of water twice, wicking away excess water each time. Stain was finally applied to the grid in two successive applications, the first time wicking immediately and the

second time allowing the stain to incubate for 30 s before finally drying the sample completely using the filter paper. Grids were visualized using a Tecnai-12 microscope (FEI).

Interface Analysis

Protein residue contact maps were generated using matplotlib using pdb atomic coordinate files corresponding to the cages analyzed in this study with a 10 Å all-atom distance cutoff.

3.6 Acknowledgements

The authors thank Sri Kosuri for early discussions on the project, as well as Pascal Egea and Mark Arbing for helpful discussions, and Amy B. Guo, Omar M. Ebrahim, Alan J. Chen, Amy K. Pai, and Megan Satyadi, and other members of the UCLA iGEM team.

3.7 Supplemental Data

SUPPLEMENTAL TABLE 1. SEQUENCES OF REFERENCED PTPC DESIGNS.

Construct Name:	Protein Sequence:
PCQth140	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGLVPRGSGAAPQEFFDGIVA AVKADRYAFYTGFFNDFYNLDENLGTRISEE AVRNSWNTAASGGFFAAAAAPTTWYTDFRA DIPRIDVPALILHGTGDRTLPIENTARVFHKAL PSAEYVEVEGAPHGLLWTHAEEVNTALLAFL AKAQEAQKQKLLTEVETYVLSIIPSGPLKAEI AQRLEDVFAGKNTDLEVLMEWLKTRPILSPL TKGILGFVFTLTVPSERGLQRRRFVQNALNG NGDPNNMDKAVKLYRKLKREITFHGAKEISL SYSAGALASCMGLIYNRMGAVTTEVAFGLVC ATCEQIADSQHRSHRQ
PCQth256	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGAAPQEFFDGIVAAVKADR YAFYTGFFNDFYNLDENLGTRISEEAVRNSW NTAASGGFFAAAAAPTTWYTDFRADIPRIDV PALILHGTGDRTLPIENTARVFHKALPSAEYV EVEGLVPRGSGAPHGLLWTHAEEVNTALLAF LAKAQEAQKQKLLTEVETYVLSIIPSGPLKAE IAQRLEDVFAGKNTDLEVLMEWLKTRPILSP LTKGILGFVFTLTVPSERGLQRRRFVQNALNG NGDPNNMDKAVKLYRKLKREITFHGAKEISL SYSAGALASCMGLIYNRMGAVTTEVAFGLVC ATCEQIADSQHRSHRQ

Construct Name:	Protein Sequence:
PCQth302	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGAAPQEFFDGIVAAVKADR YAFYTGFFNDFYNLDENLGTRISEEAVRNSW NTAASGGFFAAAAAPTTWYTDFRADIPRIDV PALILHGTGDRTLPIENTARVFHKALPSAEYV EVEGAPHGLLWTHAEEVNTALLAFLAKAQE AQKQKLLTEVETYVLSIIPSGLVPRGSGPLKA EIAQRLEDVFAGKNTDLEVLMEWLKTRPILS PLTKGILGFVFTLTVPSERGLQRRRFVQNALN GNGDPNNMDKAVKLYRKLKREITFHGAKEIS LSYSAGALASCMGLIYNRMGAVTTEVAFGLV CATCEQIADSQHRSHRQHHHHHH
PCQth336	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGAAPQEFFDGIVAAVKADR YAFYTGFFNDFYNLDENLGTRISEEAVRNSW NTAASGGFFAAAAAPTTWYTDFRADIPRIDV PALILHGTGDRTLPIENTARVFHKALPSAEYV EVEGAPHGLLWTHAEEVNTALLAFLAKAQE AQKQKLLTEVETYVLSIIPSGPLKAEIAQRLE DVFAGKNTDLEVLMEWLKTRPGLVPRGSGIL SPLTKGILGFVFTLTVPSERGLQRRRFVQNAL NGNGDPNNMDKAVKLYRKLKREITFHGAKE ISLSYSAGALASCMGLIYNRMGAVTTEVAFG LVCATCEQIADSQHRSHRQ
PCQth353	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGAAPQEFFDGIVAAVKADR YAFYTGFFNDFYNLDENLGTRISEEAVRNSW NTAASGGFFAAAAAPTTWYTDFRADIPRIDV PALILHGTGDRTLPIENTARVFHKALPSAEYV EVEGAPHGLLWTHAEEVNTALLAFLAKAQE AQKQKLLTEVETYVLSIIPSGPLKAEIAQRLE DVFAGKNTDLEVLMEWLKTRPILSPLTKGIL

Construct Name:	Protein Sequence:
	GFVFTLTGLVPRGSGERGLQRRRFVQNALNG NGDPNNMDKAVKLYRKLKREITFHGAKEISL SYSAGALASCMGLIYNRMGAVTTEVAFGLVC ATCEQIADSQHRSHRQ
PCQfur302	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGAAPQEFFDGIVAAVKADR YAFYTGFFNDFYNLDENLGTRISEEAVRNSW NTAASGGFFAAAAAPTTWYTDFRADIPRIDV PALILHGTGDRTLPIENTARVFHKALPSAEYV EVEGAPHGLLWTHAEEVNTALLAFLAKAQE AQKQKLLTEVETYVLSIIPSGGRGRKRSGPLK AEIAQRLEDVFAGKNTDLEVLMEWLKTRPIL SPLTKGILGFVFTLTVPSERGLQRRRFVQNAL NGNGDPNNMDKAVKLYRKLKREITFHGAKE ISLSYSAGALASCMGLIYNRMGAVTTEVAFG LVCATCEQIADSQHRSHRQHHHHHH
PCQmmp302	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGAAPQEFFDGIVAAVKADR YAFYTGFFNDFYNLDENLGTRISEEAVRNSW NTAASGGFFAAAAAPTTWYTDFRADIPRIDV PALILHGTGDRTLPIENTARVFHKALPSAEYV EVEGAPHGLLWTHAEEVNTALLAFLAKAQE AQKQKLLTEVETYVLSIIPSGGPLGIAGQSGP LKAEIAQRLEDVFAGKNTDLEVLMEWLKTR PILSPLTKGILGFVFTLTVPSERGLQRRRFVQN ALNGNGDPNNMDKAVKLYRKLKREITFHGA KEISLSYSAGALASCMGLIYNRMGAVTTEVA FGLVCATCEQIADSQHRSHRQHHHHHH

Construct Name:	Protein Sequence:
PCQaep302	MPFITVGQENSTSIDLYYEDHGTGTPVVLIHG FPLSGHSWERQSAALLDAGARVITYDRRGFG QSSQPTTGYDYDTFAADLNTVLETLDLQDAV LVGFSMGTGEVARYVSSYGTARIAAVAFLASL EPFLLKTDDNPDGAAPQEFFDGIVAAVKADR YAFYTGFFNDFYNLDENLGTRISEEAVRNSW NTAASGGFFAAAAAPTTWYTDFRADIPRIDV PALILHGTGDRTLPIENTARVFHKALPSAEYV EVEGAPHGLLWTHAEEVNTALLAFLAKAQE AQKQKLLTEVETYVLSIIPSGGAANGSGPLKA EIAQRLEDVFAGKNTDLEVLMEWLKTRPILS PLTKGILGFVFTLTVPSERGLQRRRFVQNALN GNGDPNNMDKAVKLYRKLKREITFHGAKEIS LSYSAGALASCMGLIYNRMGAVTTEVAFGLV CATCEQIADSQHRSHRQHHHHHH
47thPENT41	Trimer: MPIFTLNTNIKATDVPSDFLSLTSRLVGLILSK PGSYVAVHINTDQQLSFGGSTNPAAFGTLMSI GGIEPSKNRDHSAVLFDHLNAMLGIPKNRMY IHFVNLNGDDVGWNGTTFIGS Pentamer: MNQHSHKDYETVRIAVVRARWHADIVDACV EAFEIAMAAILVPRGSGGDRFAVDVFDVPGA YEIPLHARTLAETGRYGAVLGTAFVVNGGIY RHEFVASAVIDGMMNVQLSTGVPVLSAVLTP HRYRDSAEHHRFFAAHFAVKGVEAARACIEI LAAREKIAAGSLEHHHHHH

Construct Name:	Protein Sequence:
50thTRI45	Trimer: MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTGGLVPRGSGVPDADTVIKALSVLKE KGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKA MKLGHDILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCKWFKAGVLAVGVGKA LVKGKPDEVREKAKKFVKKIRGCTE Pentamer: MNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRFAVDVFDVPGAYEIPL HARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVINGMMNVQLNTGVPVLSAVLTPHNY DKSKAHTLLFLALFAVKGMEAARACVEILAAREKIAAGSLEHHHHHH
50thTRI63	Trimer: MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKE KGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGLVPRGSGVFYMPGVMTPTELVKA MKLGHDILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCKWFKAGVLAVGVGKA LVKGKPDEVREKAKKFVKKIRGCTE Pentamer: MNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRFAVDVFDVPGAYEIPL HARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVINGMMNVQLNTGVPVLSAVLTPHNY

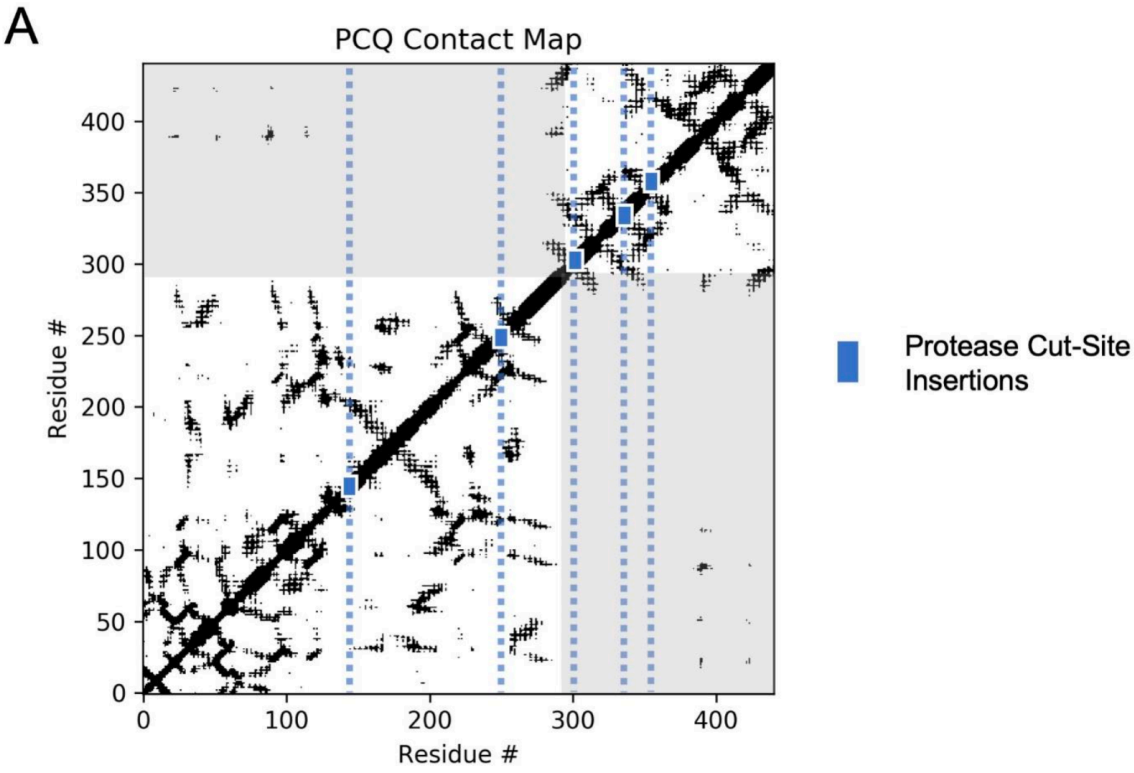
Construct Name:	Protein Sequence:
	DKSKAHTLLFLALFAVKGMEAARACVEILAA REKIAAGSLEHHHHHH
50thTRI125	Trimer: MKMEELFKKHKIVAVLRANSVEEAIEKAVAV FAGGVHLIEITFTVPDADTVIKALSVLKE KGAIIGAGTVTSVEQCRKAVESGAEFIVSPHL DEEISQFCKEKGVFYMPGVMTPTELVKA MKLGLVPRGSGHDILKLFPGEVVGPQFVKA MKGPFPNVKFVPTGGVNLDNVCKWFKAGV LAVGVGKA LVKGKPDEVREKAKKFVKKIRGCTE Pentamer: MNQHSHKDHETVRIAVVRARWHAEIVDACV SAFEAAMRDIGGDRFAVDVFDVPGAYEIPL HARTLAETGRYGAVLGTAFVVNGGIYRHEFV ASAVINGMMNVQLNTGVPVLSAVLTPHNY DKSKAHTLLFLALFAVKGMEAARACVEILAA REKIAAGSLEHHHHHH

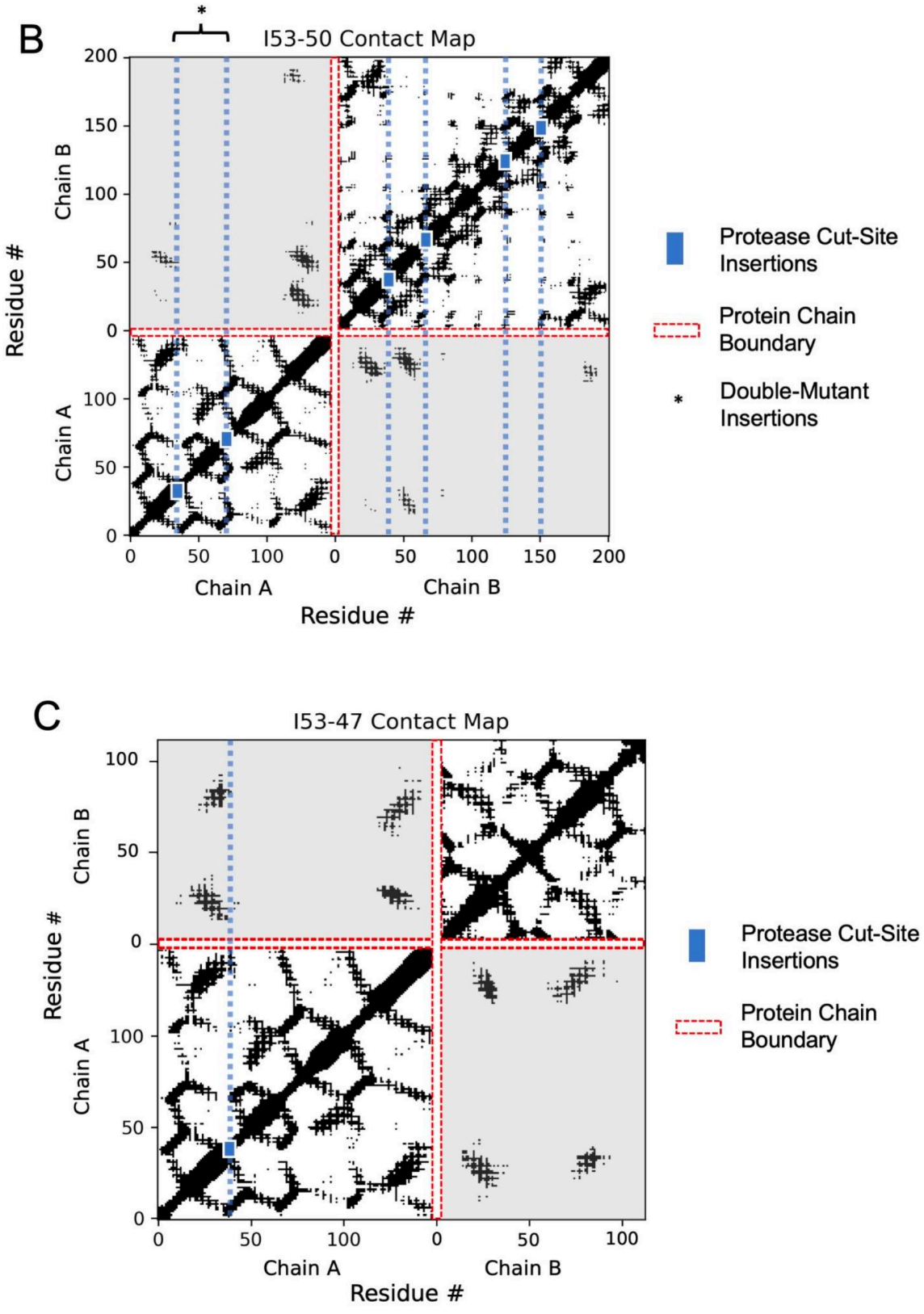
Construct Name:	Protein Sequence:
50thTRI150	Trimer: MKMEELFKKHKIVAVLRANSVEEAIEKAVAV FAGGVHLIEITFTVPDADTVIKALSVLKE KGAIIGAGTVTSVEQCRKAVESGAEFIVSPHL DEEISQFCKEKGVFYMPGVMTPTELVKA MKLGHDILKLFPGEVVGPQFVKAMKGPFPLV PRGSGNVKFVPTGGVNLDNVCKWFKAGVL AVGVGKA LVKGKPDEVREKAKKFVKKIRGCTE Pentamer: MNQHSHKDHETVRIAVVRARWHAEIVDACV SAFEAAMRDIGGDRFAVDVFDVPGAYEIPL HARTLAETGRYGAVLGTAFVVNGGIYRHEFV ASAVINGMMNVQLNTGVPVLSAVLTPHNY DKSKAHTLLFLALFAVKGMEAARACVEILAA REKIAAGSLEHHHHHH
50thPENT43	Trimer: MKMEELFKKHKIVAVLRANSVEEAIEKAVAV FAGGVHLIEITFTVPDADTVIKALSVLKE KGAIIGAGTVTSVEQCRKAVESGAEFIVSPHL DEEISQFCKEKGVFYMPGVMTPTELVKA MKLGHDILKLFPGEVVGPQFVKAMKGPFPN VKFVPTGGVNLDNVCKWFKAGVLAVGVGK A LVKGKPDEVREKAKKFVKKIRGCTE Pentamer: MNQHSHKDHETVRIAVVRARWHAEIVDACV SAFEAAMRDIGGLVPRGSDRFAVDVFDVPGA YEIPLHARTLAETGRYGAVLGTAFVVNGGIY RHEFVASAVINGMMNVQLNTGVPVLSAVLTP HNY DKSKAHTLLFLALFAVKGMEAARACVEILAA REKIAAGSLEHHHHHH

Construct Name:	Protein Sequence:
50thPENT67	Trimer: MKMEELFKKHKIVAVLRANSVEEAIEKAVAV FAGGVHLIEITFTVPDADTVIKALSVLKEKGA IIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEI SQFCKEKGVFYMPGVMTPTELVKAMKLGHD ILKLFPGEVVGPQFVKAMKGPFPNVKFVPTG GVNLDNVCKWFKAGVLAVGVGKALVKGKP DEVREKAKKFVKKIRGCTE Pentamer: MNQHSHKDHETVRIAVVRARWHAEIVDACV SAFEAAMRDIGGDRFAVDVFDVPGAYEIPLH ARTLAETLVPRGSRYGAVLGTAFVVNGGIYR HEFVASAVINGMMNVQLNTGVPVLSAVLTPH NYDKSKAHTLLFLALFAVKGMEAARACVEIL AAREKIAAGSLEHHHHHH
50th43_67	Trimer: MKMEELFKKHKIVAVLRANSVEEAIEKAVAV FAGGVHLIEITFTVPDADTVIKALSVLKEKGA IIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEI SQFCKEKGVFYMPGVMTPTELVKAMKLGHD ILKLFPGEVVGPQFVKAMKGPFPNVKFVPTG GVNLDNVCKWFKAGVLAVGVGKALVKGKP DEVREKAKKFVKKIRGCTE Pentamer: MNQHSHKDHETVRIAVVRARWHAEIVDACV SAFEAAMRDIGGGGGGLVPRGSGGGGGGDR FAVDVFDVPGAYEIPLHARTLAETGGGGGGL VPRGSGGGGGGRYGAVLGTAFVVNGGIYRH EFVASAVINGMMNVQLNTGVPVLSAVLTPHN YDKSKAHTLLFLALFAVKGMEAARACVEILA AREKIAAGSLEHHHHHH

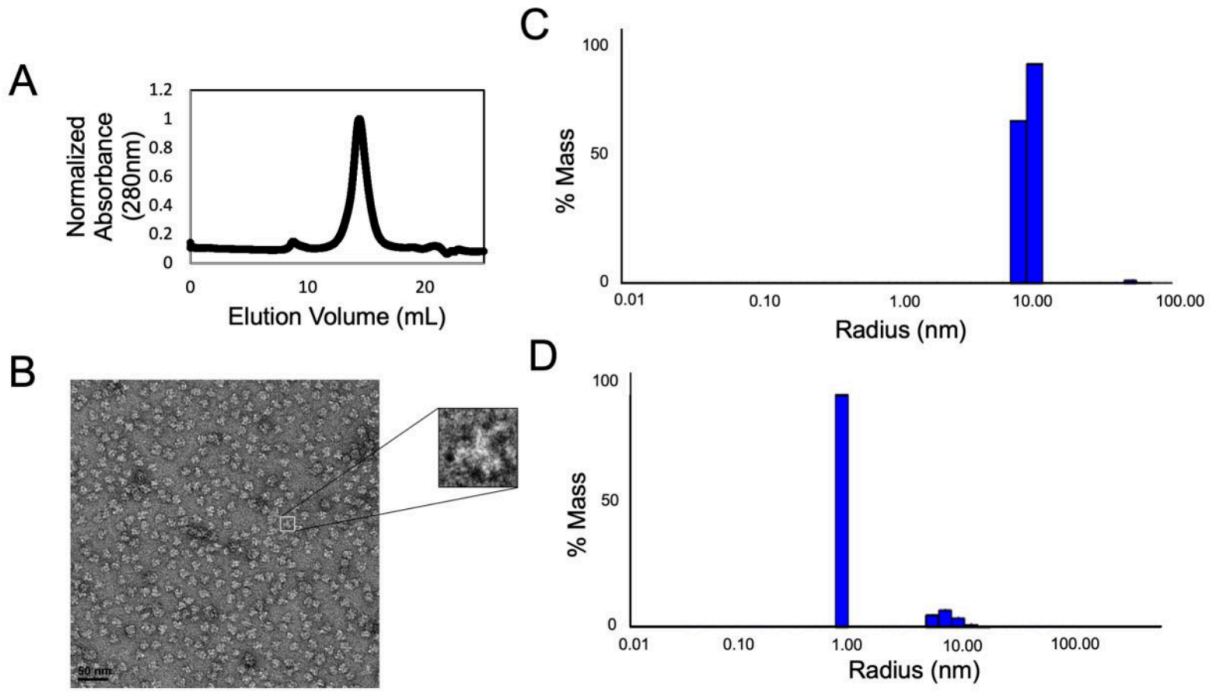
Construct Name:	Protein Sequence:
47thPENT41_LL	Trimer: MPIFTLNTNIKATDVPSDFLSLTSRLVGLILSK PGSYVAVHINTDQQLSFGGSTNPAAFGTLMSI GGIEPSKNRDHSAVLFDHLNAMLGIPKNRMY IHFVNLNGDDVGWNGTT Pentamer: MNQHSHKDYETVRIAVVRARWHADIVDACV EAFEIAMAAIGGGGGGLVPRGSGGGGGGDR FAVDVFDVPGAYEIPLHARTLAETGRYGAVL GTAFVVNGGIYRHEFVASAVIDGMMNVQLST GVPVLSAVLTPHRYRDSAEHHRFFAAHFAVK GVEAARACIEILAAREKIAAGSLEHHHHHH

Supplemental Data Table S3.1

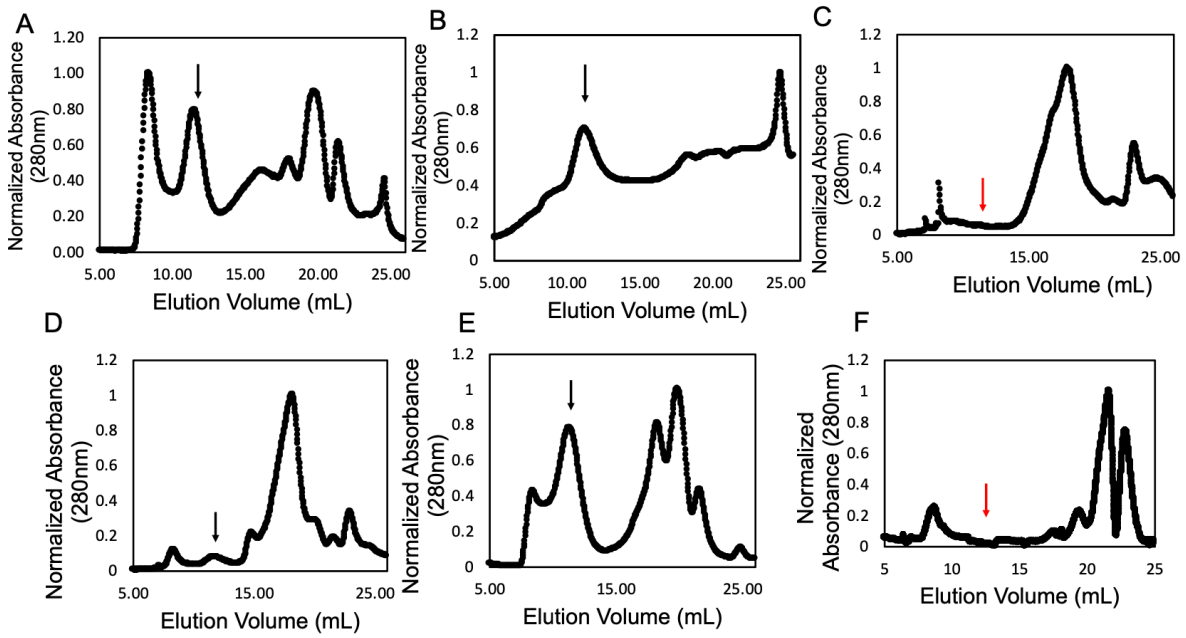



B
I53-50 Contact Map
*
Chain B
Residue #
Chain A
Chain A
Chain B
Residue #
Protease Cut-Site Insertions
Protein Chain Boundary
Double-Mutant Insertions
C
I53-47 Contact Map
Chain B
Residue #
Chain A
Chain A
Chain B
Residue #
Protease Cut-Site Insertions
Protein Chain Boundary

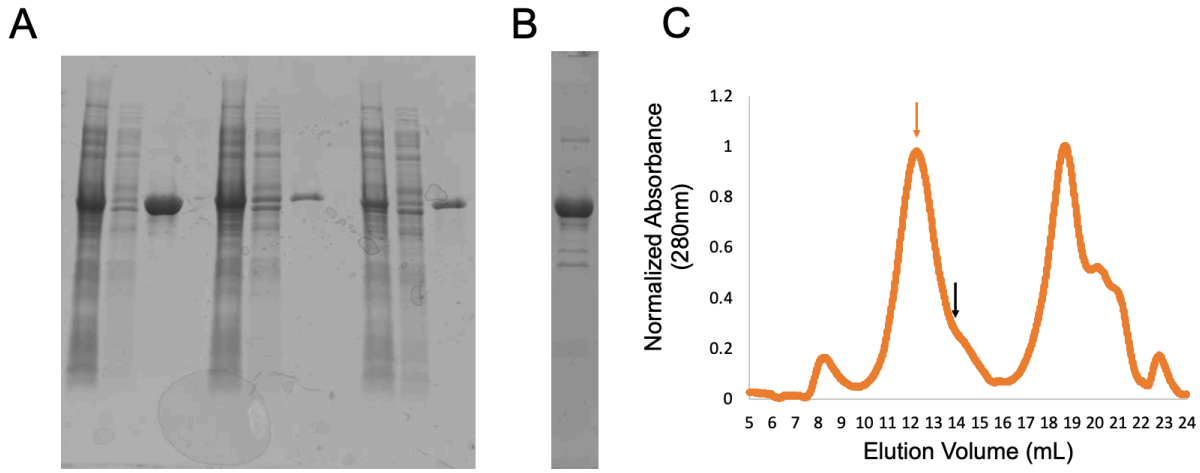
Supplemental Data Figure S3.1. Attempted protease cut site insertions for the protease-triggered protein cage (PTPC) designs examined in the present study. The insertion sites are displayed in the context of residue- residue contact maps, to help illustrate the extent of the contacts that would need to be disrupted in order to separate the resulting fragments. PTPC design variants are indicated for each of the three cages investigated in this study: PCQ, I53-50, and I53-47. Loop insertion locations are plotted as blue rectangles, with dashed lines indicating the separation of degradation products that would be divided by proteolysis at the corresponding site. (A) PCQ PTPC insertion locations. (B) I53-50 PTCP insertion locations. Insertions comprising the 50th43_67 double mutant are denoted with asterisks. Two chains comprising assembled cage design are separated by red dashed lines. Chain A is the pentameric component, and chain B is the trimeric component (C) I53-47 PTCP insertion location. Two chains comprising assembled cage design are separated by red dashed lines. Chain A is the pentameric component and chain B is the trimeric component.



Supplemental Data Figure S3.2. Assessing helical fusion-type cage PTPC mutants for assembly into cages. (A) PCQth302 SEC chromatogram indicates elution of redesigned protein at volume consistent with assembly into 12-component cage. (B) Negatively stained electron micrograph of SEC-purified PCQth302 reveals symmetric assemblies with dimensions that agree with that of protein cage design model. (C) Dynamic light scattering data performed on PCQth302 reveals majority species in solution is appropriately sized for proper protein cage assembly. (D) DLS data collected on PCQth353 indicates design does not assemble into cage species in abundance.

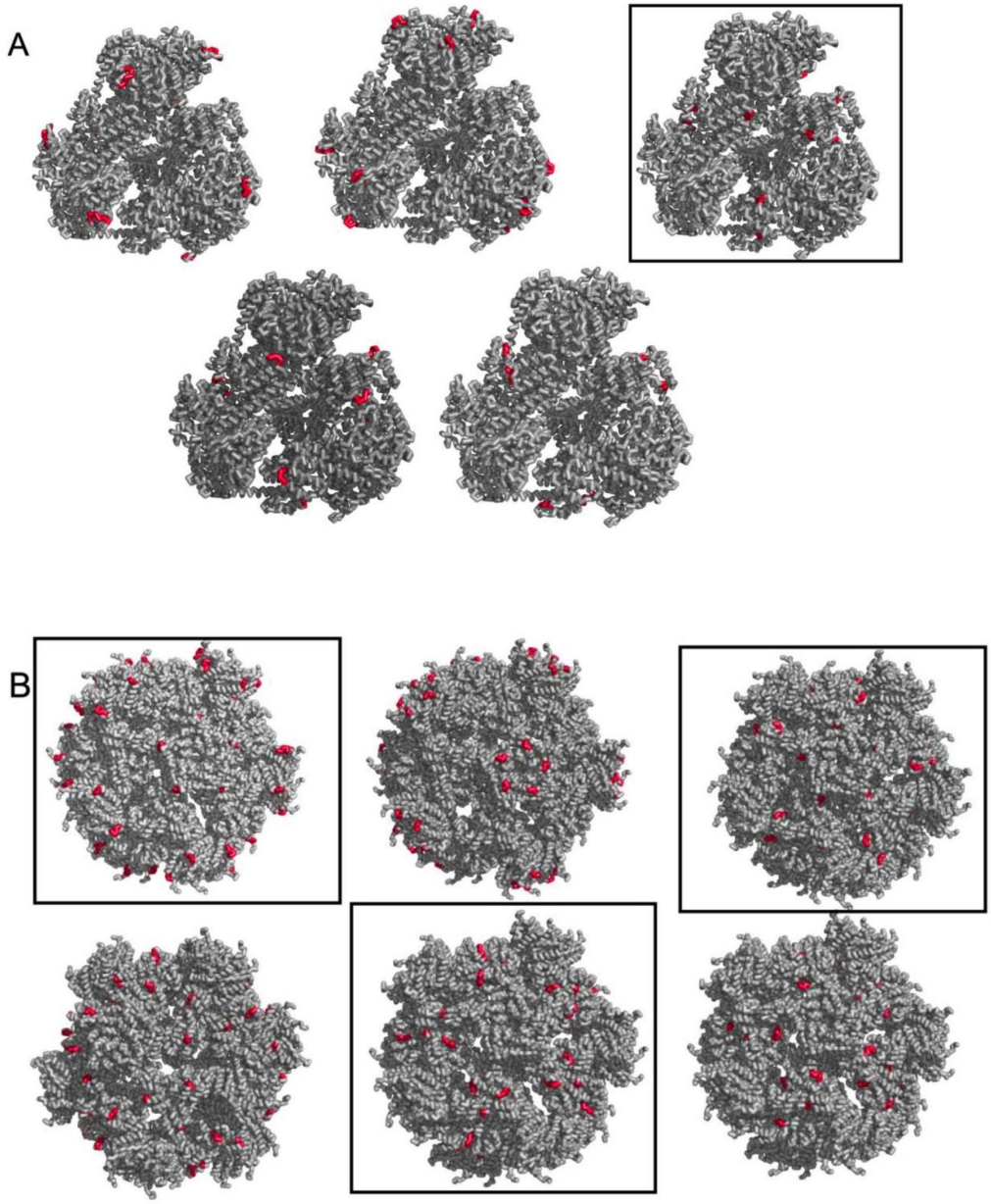


Supplemental Data Figure S3.3. Assessing interface-based cage PTPC mutants for correct assembly using SEC. Cages 47th_PENT41 (A), 50th_TRI45 (B), 50thTRI125 (D) 50thPENT43 (E) elute on SEC with a peak that indicates assembly into the correct icosahedral cage species (denoted with black arrow). 50thTRI150 (C) and 50thPENT43_67 (F) do not elute on SEC with significant signal at elution volumes that agree with correctly assembled cage; expected elution volume indicated with red arrow.

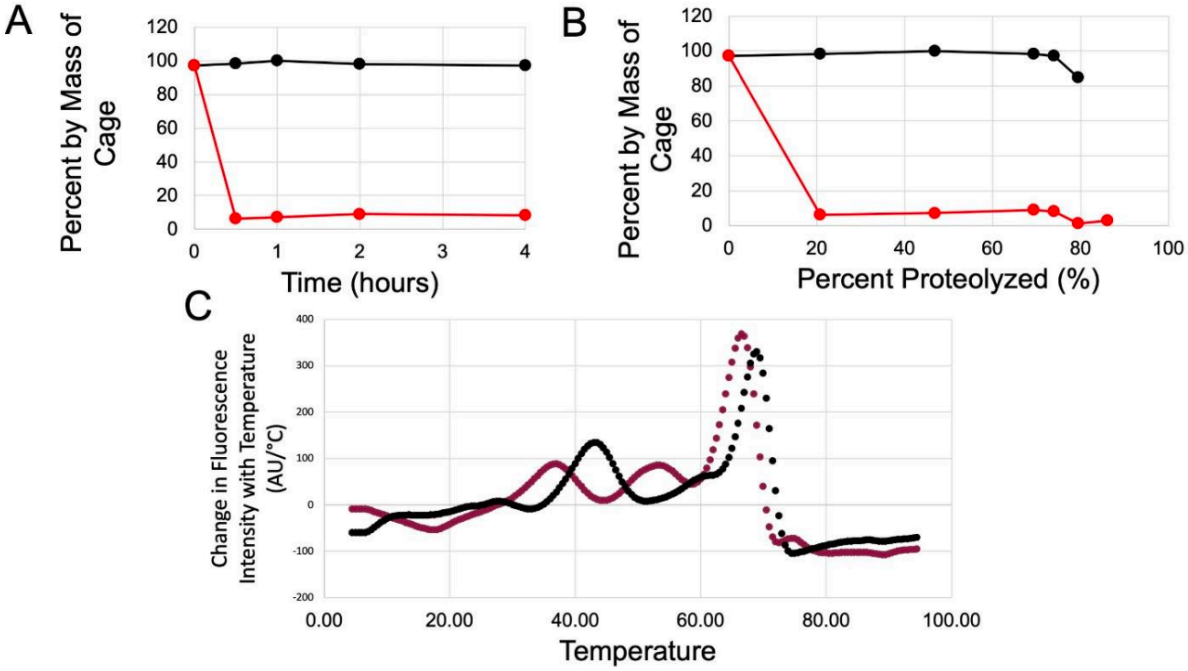


Supplemental Data Figure S3.4. SDS-PAGE of purification for helix-fusion type PTPCs with alternate protease specificity. (A) Total soluble lysate, wash fraction, and elution fraction from purification of PCQmmp302 (lanes 1-3), PCQfur302 (lanes 5-7), and PCQaep302 (lanes 9-11).

(B) SDS-PAGE of PCQmmp302 protease cleavage assay indicates minimal digestion by target protease. (C) SEC analysis reveals PCQfur302 does not assemble into the expected cage structure, eluting earlier than expected elution volume for PCQ mutants (majority elution volume denoted with orange arrow, expected elution volume denoted with black arrow).



Supplemental Data Figure S3.5. Protease cleavage insertion sites shown on 3D structures of cages PC-quad and I53-50. (A) Insertion sites mapped on structure from left to right top to bottom (boxed cages assembled as expected into cages): PCQth140, PCQth256, PCQth302, PCQth336, PCQth353. (B) Insertion sites mapped on the structure from left to right top to bottom (boxed cages assembled as expected into cages): 50thPENT43, 50thPENT67 , 50thTRI45, 50thTRI63, 50thTRI125, 50thTRI150.



Supplemental Data Figure S3.6. Dynamic light scattering and thermal shift assay of helix-fusion type PTPC PCQth302. (A) Interpretation of DLS data in terms of fraction of PCQth302 retaining cage structure as a function of time after protease addition for control (black) and +thrombin condition (red). (B) Interpretation of DLS data in terms of fraction of PCQth302 retaining cage structure as a function of cleavage efficiency as assessed by SDS-PAGE for +thrombin condition (red), or the corresponding timepoint for the control condition (black). (C) Derivative plot of thermal shift assay depicted in Figure 3.2 showing control conditions (black) and the effects of adding thrombin (red). The curves demonstrate a destabilizing effect on the assembly, manifesting as exposure of hydrophobic protein surfaces, when thrombin is introduced. For DLS experiments, mass percentages attributed to the assembled cage were computed by summing over diameters that fell within the range of 12-25nm.

3.8 References

(1) Ahnert, S. E.; Marsh, J. A.; Hernandez, H.; Robinson, C. V.; Teichmann, S. A. Principles of Assembly Reveal a Periodic Table of Protein Complexes. Science 2015, 350 (6266), aaa2245–aaa2245. https://doi.org/10.1126/science.aaa2245.

(2) Nichols, R. J.; Cassidy-Amstutz, C.; Chaijarasphong, T.; Savage, D. F. Encapsulins: Molecular Biology of the Shell. Crit. Rev. Biochem. Mol. Biol. 2017, 52 (5), 583–594. https://doi.org/10.1080/10409238.2017.1337709.

(3) Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G. Solving the Structure of Human H Ferritin by Genetically Engineering Intermolecular Crystal Contacts. Nature 1991, 349 (6309), 541–544. https://doi.org/10.1038/349541a0.

(4) Yeates, T. O.; Kerfeld, C. A.; Heinhorst, S.; Cannon, G. C.; Shively, J. M. Protein-Based Organelles in Bacteria: Carboxysomes and Related Microcompartments. Nat. Rev. Microbiol. 2008, 6 (9), 681–691. https://doi.org/10.1038/nrmicro1913.

(5) Edwardson, T. G. W.; Hilvert, D. Virus-Inspired Function in Engineered Protein Cages. J. Am. Chem. Soc. 2019, 141 (24), 9432–9443. https://doi.org/10.1021/jacs.9b03705.

(6) Edwardson, T. G. W.; Mori, T.; Hilvert, D. Rational Engineering of a Designed Protein Cage for SiRNA Delivery. J. Am. Chem. Soc. 2018, 140 (33), 10439–10442. https://doi.org/10.1021/jacs.8b06442.

(7) Cannon, K. A.; Ochoa, J. M.; Yeates, T. O. High-Symmetry Protein Assemblies: Patterns and Emerging Applications. Curr. Opin. Struct. Biol. 2019, 55, 77–84. https://doi.org/10.1016/j.sbi.2019.03.008.

(8) Waghwani, H. K.; Uchida, M.; Fu, C.-Y.; LaFrance, B.; Sharma, J.; McCoy, K.; Douglas, T. Virus-Like Particles (VLPs) as a Platform for Hierarchical Compartmentalization. Biomacromolecules 2020, 21 (6), 2060–2072. https://doi.org/10.1021/acs.biomac.0c00030.

(9) Wang, Y.; Douglas, T. Protein Nanocage Architectures for the Delivery of Therapeutic Proteins. Current Opinion in Colloid & Interface Science 2021, 51, 101395. https://doi.org/10.1016/j.cocis.2020.101395.

(10) Hoersch, D.; Roh, S.-H.; Chiu, W.; Kortemme, T. Reprogramming an ATP-Driven Protein Machine into a Light-Gated Nanocage. Nat Nanotechnol 2013, 8 (12), 928–932. https://doi.org/10.1038/nnano.2013.242.

(11) Malay, A. D.; Miyazaki, N.; Biela, A.; Chakraborti, S.; Majsterkiewicz, K.; Stupka, I.; Kaplan, C. S.; Kowalczyk, A.; Piette, B. M. A. G.; Hochberg, G. K. A.; Wu, D.; Wrobel, T. P.; Fineberg, A.; Kushwah, M. S.; Kelemen, M.; Vavpetič, P.; Pelicon, P.; Kukura, P.; Benesch, J. L. P.; Iwasaki, K.; Heddle, J. G. An Ultra-Stable Gold-Coordinated Protein Cage Displaying Reversible Assembly. Nature 2019, 569 (7756), 438–442. https://doi.org/10.1038/s41586-019-1185-4.

(12) Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments. Proc. Natl. Acad. Sci. U.S.A. 2001, 98 (5), 2217–2221. https://doi.org/10.1073/pnas.041614998.

(13) Laniado, J.; Yeates, T. O. A Complete Rule Set for Designing Symmetry Combination Materials from Protein Molecules. Proc Natl Acad Sci U S A 2020, 117 (50), 31817–31823. https://doi.org/10.1073/pnas.2015183117.

(14) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. Accurate Design of Megadalton-Scale Two-Component

Icosahedral Protein Complexes. Science 2016, 353 (6297), 389–394. https://doi.org/10.1126/science.aaf8818.

(15) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R.; Yi, S.; Davis, T. N.; Gonen, T.; King, N. P.; Baker, D. Design of a Hyperstable 60-Subunit Protein Dodecahedron. [Corrected]. Nature 2016, 535 (7610), 136–139. https://doi.org/10.1038/nature18010.

(16) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. Accurate Design of Co-Assembling Multi-Component Protein Nanomaterials. Nature 2014, 510 (7503), 103–108. https://doi.org/10.1038/nature13404.

(17) Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. Annu. Rev. Biophys. 2017, 46 (1), 23–42. https://doi.org/10.1146/annurev-biophys-070816-033928.

(18) Lai, Y.-T.; Cascio, D.; Yeates, T. O. Structure of a 16-Nm Cage Designed by Using Protein Oligomers. Science 2012, 336 (6085), 1129–1129. https://doi.org/10.1126/science.1219351.

(19) Laniado, J.; Cannon, K. A.; Miller, J. E.; Sawaya, M. R.; McNamara, D. E.; Yeates, T. O. Geometric Lessons and Design Strategies for Nanoscale Protein Cages. ACS Nano 2021, 15 (3), 4277–4286. https://doi.org/10.1021/acsnano.0c07167.

(20) Cannon, K. A.; Nguyen, V. N.; Morgan, C.; Yeates, T. O. Design and Characterization of an Icosahedral Protein Cage Formed by a Double-Fusion Protein Containing Three Distinct Symmetry Elements. ACS Synth Biol 2020, 9 (3), 517–524. https://doi.org/10.1021/acssynbio.9b00392.

(21) Lai, Y.-T.; Jiang, L.; Chen, W.; Yeates, T. O. On the Predictability of the Orientation of Protein Domains Joined by a Spanning Alpha-Helical Linker. Protein Engineering, Design and Selection 2015, 28 (11), 491–500. https://doi.org/10.1093/protein/gzv035.

(22) Badieyan, S.; Sciore, A.; Eschweiler, J. D.; Koldewey, P.; Cristie-David, A. S.; Ruotolo, B. T.; Bardwell, J. C. A.; Su, M.; Marsh, E. N. G. Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 2017, 18 (19), 1888–1892. https://doi.org/10.1002/cbic.201700406.

(23) Cristie-David, A. S.; Chen, J.; Nowak, D. B.; Bondy, A. L.; Sun, K.; Park, S. I.; Banaszak Holl, M. M.; Su, M.; Marsh, E. N. G. Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability. J. Am. Chem. Soc. 2019, 141 (23), 9207–9216. https://doi.org/10.1021/jacs.8b13604.

(24) Vulovic, I.; Yao, Q.; Park, Y.-J.; Courbet, A.; Norris, A.; Busch, F.; Sahasrabuddhe, A.; Merten, H.; Sahtoe, D. D.; Ueda, G.; Fallas, J. A.; Weaver, S. J.; Hsia, Y.; Langan, R. A.; Plückthun, A.; Wysocki, V. H.; Veesler, D.; Jensen, G. J.; Baker, D. Generation of Ordered Protein Assemblies Using Rigid Three-Body Fusion. Proc Natl Acad Sci USA 2021, 118 (23), e2015037118. https://doi.org/10.1073/pnas.2015037118.

(25) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. Science 2012, 336 (6085), 1171–1174. https://doi.org/10.1126/science.1219364.

(26) Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U.; Ellis, D.; Bale, J. B.; Ke, S.; Lenz, G. H.; Yehdego, A.; Ravichandran, R.; Pun, S. H.; King, N. P.;

Baker, D. Evolution of a Designed Protein Assembly Encapsulating Its Own RNA Genome. Nature 2017, 552 (7685), 415–420. https://doi.org/10.1038/nature25157.

(27) Edwardson, T. G. W.; Tetter, S.; Hilvert, D. Two-Tier Supramolecular Encapsulation of Small Molecules in a Protein Cage. Nat Commun 2020, 11 (1), 5410. https://doi.org/10.1038/s41467-020-19112-1.

(28) McConnell, S. A.; Cannon, K. A.; Morgan, C.; McAllister, R.; Amer, B. R.; Clubb, R. T.; Yeates, T. O. Designed Protein Cages as Scaffolds for Building Multienzyme Materials. ACS Synth Biol 2020, 9 (2), 381–391. https://doi.org/10.1021/acssynbio.9b00407.

(29) 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, K. K.; Veesler, D.; Correnti, C. E.; Stewart, L. J.; Baker, D.; Loré, K.; Perez, L.; King, N. P. Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. Cell 2019, 176 (6), 1420-1431.e17. https://doi.org/10.1016/j.cell.2019.01.046.

(30) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. Nat Commun 2019, 10 (1), 1864. https://doi.org/10.1038/s41467-019-09836-0.

(31) Terasaka, N.; Azuma, Y.; Hilvert, D. Laboratory Evolution of Virus-like Nucleocapsids from Nonviral Protein Cages. Proc Natl Acad Sci USA 2018, 115 (21), 5432–5437. https://doi.org/10.1073/pnas.1800527115.

(32) Dalmau, M.; Lim, S.; Wang, S.-W. PH-Triggered Disassembly in a Caged Protein Complex. Biomacromolecules 2009, 10 (12), 3199–3206. https://doi.org/10.1021/bm900674v.

(33) Jones, J. A.; Cristie-David, A. S.; Andreas, M. P.; Giessen, T. W. Triggered Reversible Disassembly of an Engineered Protein Nanocage. Angew. Chem. Int. Ed. 2021, 60 (47), 25034–25041. https://doi.org/10.1002/anie.202110318.

(34) Stupka, I.; Azuma, Y.; Biela, A. P.; Imamura, M.; Scheuring, S.; Pyza, E.; Woźnicka, O.; Maskell, D. P.; Heddle, J. G. Chemically Induced Protein Cage Assembly with Programmable Opening and Cargo Release. Sci. Adv. 2022, 8 (1), eabj9424. https://doi.org/10.1126/sciadv.abj9424.

(35) Zang, J.; Chen, H.; Zhang, X.; Zhang, C.; Guo, J.; Du, M.; Zhao, G. Disulfide-Mediated Conversion of 8-Mer Bowl-like Protein Architecture into Three Different Nanocages. Nat Commun 2019, 10 (1), 778. https://doi.org/10.1038/s41467-019-08788-9.

(36) Schafer, F. Q.; Buettner, G. R. Redox State and Redox Environment in Biology. In Signal Transduction by Reactive Oxygen and Nitrogen Species: Pathways and Chemical Principles; Forman, H. J., Fukuto, J., Torres, M., Eds.; Kluwer Academic Publishers: Dordrecht, 2004; pp 1–14. https://doi.org/10.1007/0-306-48412-9_1.

(37) Subramanian, R. H.; Zhu, J.; Bailey, J. B.; Chiong, J. A.; Li, Y.; Golub, E.; Tezcan, F. A. Design of Metal-Mediated Protein Assemblies via Hydroxamic Acid Functionalities. Nat Protoc 2021, 16 (7), 3264–3297. https://doi.org/10.1038/s41596-021-00535-z.

(38) Yang, M.; Song, W. J. Diverse Protein Assembly Driven by Metal and Chelating Amino Acids with Selectivity and Tunability. Nat Commun 2019, 10 (1), 5545. https://doi.org/10.1038/s41467-019-13491-w.

(39) Huard, D. J. E.; Kane, K. M.; Tezcan, F. A. Re-Engineering Protein Interfaces Yields Copper-Inducible Ferritin Cage Assembly. Nat Chem Biol 2013, 9 (3), 169–176. https://doi.org/10.1038/nchembio.1163.

(40) Ni, T. W.; Tezcan, F. A. Structural Characterization of a Microperoxidase Inside a Metal-Directed Protein Cage. Angewandte Chemie International Edition 2010, 49 (39), 7014–7018. https://doi.org/10.1002/anie.201001487.

(41) Divine, R.; Dang, H. V.; Ueda, G.; Fallas, J. A.; Vulovic, I.; Sheffler, W.; Saini, S.; Zhao, Y. T.; Raj, I. X.; Morawski, P. A.; Jennewein, M. F.; Homad, L. J.; Wan, Y.-H.; Tooley, M. R.; Seeger, F.; Etemadi, A.; Fahning, M. L.; Lazarovits, J.; Roederer, A.; Walls, A. C.; Stewart, L.; Mazloomi, M.; King, N. P.; Campbell, D. J.; McGuire, A. T.; Stamatatos, L.; Ruohola-Baker, H.; Mathieu, J.; Veesler, D.; Baker, D. Designed Proteins Assemble Antibodies into Modular Nanocages. Science 2021, 372 (6537), eabd9994. https://doi.org/10.1126/science.abd9994.

(42) Douglas, S. M.; Bachelet, I.; Church, G. M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. Science 2012, 335 (6070), 831–834. https://doi.org/10.1126/science.1214081.

(43) Duffy, M. J. Proteases as Prognostic Markers in Cancer. Clin Cancer Res 1996, 2 (4), 613–618.

(44) Poreba, M. Protease-activated Prodrugs: Strategies, Challenges, and Future Directions. FEBS J 2020, 287 (10), 1936–1969. https://doi.org/10.1111/febs.15227.

(45) Saunders, K.; Sainsbury, F.; Lomonossoff, G. P. Efficient Generation of Cowpea Mosaicvirus Empty Virus-like Particles by the Proteolytic Processing of Precursors in Insect Cells and Plants. Virology 2009, 393 (2), 329–337. https://doi.org/10.1016/j.virol.2009.08.023.

(46) Bruinsma, R. F.; Wuite, G. J. L.; Roos, W. H. Physics of Viral Dynamics. Nat Rev Phys 2021, 3 (2), 76–91. https://doi.org/10.1038/s42254-020-00267-1.

(47) Lai, Y.-T.; Hura, G. L.; Dyer, K. N.; Tang, H. Y. H.; Tainer, J. A.; Yeates, T. O. Designing and Defining Dynamic Protein Cage Nanoassemblies in Solution. Sci. Adv. 2016, 2 (12), e1501855. https://doi.org/10.1126/sciadv.1501855.

(48) Kratz, F.; Drevs, J.; Bing, G.; Stockmar, C.; Scheuermann, K.; Lazar, P.; Unger, C. Development and in Vitro Efficacy of Novel MMP2 and MMP9 Specific Doxorubicin Albumin Conjugates. Bioorganic & Medicinal Chemistry Letters 2001, 11 (15), 2001–2006. https://doi.org/10.1016/S0960-894X(01)00354-7.

(49) Tian, S.; Huajun, W.; Wu, J. Computational Prediction of Furin Cleavage Sites by a Hybrid Method and Understanding Mechanism Underlying Diseases. Sci Rep 2012, 2 (1), 261. https://doi.org/10.1038/srep00261.

(50) Stern, L.; Perry, R.; Ofek, P.; Many, A.; Shabat, D.; Satchi-Fainaro, R. A Novel Antitumor Prodrug Platform Designed to Be Cleaved by the Endoprotease Legumain. Bioconjugate Chem. 2009, 20 (3), 500–510. https://doi.org/10.1021/bc800448u.

(51) Heinig, M.; Frishman, D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. Nucleic Acids Research 2004, 32 (Web Server), W500–W502. https://doi.org/10.1093/nar/gkh429.

(52) Mitternacht, S. FreeSASA: An Open Source C Library for Solvent Accessible Surface Area Calculations. F1000Res 2016, 5, 189. https://doi.org/10.12688/f1000research.7931.1.

(53) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. BMC Biotechnol 2008, 8 (1), 91. https://doi.org/10.1186/1472-6750-8-91.

CHAPTER 4

DESIGN OF A LIGAND-OPERABLE PROTEIN CAGE BASED ON A NATURAL SULFUR OXYGENASE ASSEMBLY

4.1 Preamble

This work is part of a larger study, led by fellow graduate student in the Yeates laboratory, Eric J. Lee. In that work, ligand-operable cage (LOC) design is demonstrated using a variety of target ligands, and by using artificial cages as scaffolds. In this chapter of my thesis, I describe my contribution to the work, which involved the design of a LOC based on a naturally occurring protein cage scaffold.

4.2 Abstract

Protein nanocages have diverse applications in medicine and biotechnology, particularly molecular delivery. However, while many studies have demonstrated the ability of protein nanocages to predictably encapsulate moieties, methods for subsequent "opening" of the nanocage for cargo release are limited. Prior studies have relied on non-specific environmental stimuli, such as drastic shifts in pH, addition of denaturants, degradation, etc. to disrupt an assembled nanocage. A modular platform with a specific, protein target-based mechanism of nanocage opening is notably lacking. To address this engineering limitation, we present a new class of designed protein cages, the Ligand-Operable Cage (LOC). LOCs primarily comprise a protein nanocage core and a fused surface binder. The geometry of the LOC is designed so that binding of a target protein ligand (or multiple copies of the ligand) to the surface binder is sterically incompatible with retention of the assembled state of the cage. Therefore, tight binding of a target ligand drives cage disassembly by mass action, subsequently exposing the encapsulated cargo. LOCs are also modular, where simply substituting the surface binder sequence can reprogram the nanocage to open in response to any target protein ligand of interest. These findings have exciting implications for targeted molecular delivery and detection.

4.3 Introduction

Cellular compartmentalization is crucial to the biochemical processes that sustain life. In eukaryotic organisms, lipid membranes separate incompatible regions of the cell into organelles. In prokaryotes, spatial compartmentalization is often achieved with the aid of proteinaceous barriers to diffusion that can form enclosed polyhedral shells. These shell-like protein assemblies are not limited to prokaryotes, however. Sometimes referred to as cages, they are found ubiquitously in nature and have evolved diverse functions – e.g. transport, storage, and sequestration -- that make use of these structures' capacity for compartmentalization. Understandably, there is significant interest in the field of protein engineering to replicate such assemblies in the laboratory with bespoke functionalities and applications. More specifically, unnatural or designed protein cages have been used in the fields of medicine and biotechnology for drug delivery vehicles[1], vaccine platforms[2], molecular and medical imaging scaffolds[3,4], enzymatic display, and other purposes. These developments have largely made use of intact, static assemblies displaying functional moieties for functionalization, neglecting to reproduce the controlled assembly and reversible encapsulation that makes natural protein cages such effective tools for compartmentalization. Towards the development of dynamically controllable protein cages, more reminiscent of functional protein systems in biology, pioneering studies have exploited sensitivity to bulk environmental changes such as pH[5], metal chelation[6], and oxidation or reduction[7]. Regarding potential biomedical applications for designed protein cages, a major advance would come from systems that respond to highly specific disease biomarkers and factors of clinical relevance. Towards this end, we demonstrate the design of protein-binding triggered cage destruction. Our cages which we term ligand-operable cages (LOCs), can sense proteins via fusion of a class of antibody-mimetic molecules called designed ankyrin repeat proteins

(DARPins) displayed on their exterior surfaces. After binding to the target, the cage protein disassembles into components, exposing the cage lumen to the surrounding environment.

4.4 Results

Our first example LOC design utilized the natural nanocage Sulfur Oxygenase Reductase (SOR). Our rationale was that, because natural nanocages often exhibit some degree of reversibility between assembled and disassembled states (in comparison to synthetic cages with designed interfaces), it might be more feasible to induce a disrupted state in their assemblies by design. Regarding design, SOR was an appropriate candidate because it possesses C-terminal helices present on its surface, permitting us to readily model and fuse appropriately oriented DARPins, which possess N-terminal helices. For this trial, we used a DARPin sequence that binds sfGFP, which was previously designed, for feasible monitoring of assembly states via absorbance tracking at 280nm (corresponding to aromatic side-chain absorbance) versus 490nm (sfGFP-specific). We attached this DARPin sequence via rigid alpha-helical fusion from C-terminus of the SOR component to N-terminus of the DARPin. This resulted in rigid attachment of the DARPin to the exterior surface of the SOR cage (Figure 4.1a). Alpha helical geometry dictates that the addition or removal of residues results in rotation around the helical axis. Approximately 3.6 residues make up a complete rotation around that axis. Due to this known quality of alpha helices, we were able to tune the length of this linker such that the binding site on the DARPin was directed towards the cage itself. The result of this design strategy is the cage is stable (i.e. exhibits no steric clashes) in its apo state, but is destabilized by binding to the target protein (sfGFP) owing to steric clashes between the target proteins and the cage core. That is, cage assembly is effectively disallowed when the DARPins are bound to sfGFP. We rationalize that this will destabilize the LOCs such that they favor a disassembled state in the presence of

their target protein, at concentrations sufficient to drive binding. We first assayed whether the assembly was affected by attachment of the fused DARPin binders, and observed the intended assembly via SEC elution profile (Figure 4.1c). We then added the cargo protein sfGFP to the cage, and again assayed the mixture by SEC. After mixing, the peak corresponding to assembly of the LOC was no longer apparent on the 280nm chromatogram. However, we observed significant 280nm signal overlayed with 490nm at a later elution volume where we would expect disassembled LOC proteins to elute.

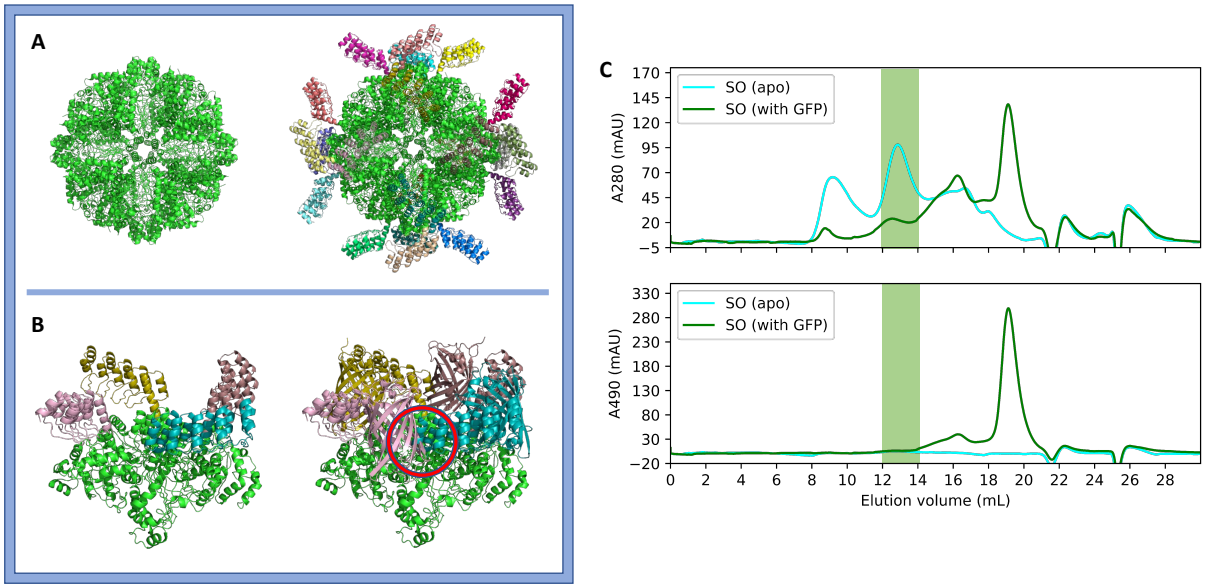


Figure 4.1. Cognate ligand-binding induced disruption of assembled SOR-N3C-α-sfGFP. A) Biological assembly of sulfur oxygenase (PDB: 2YAV) (left); Sulfur Oxygenase modeled with pair-fitted surface α-sfGFP DARPins. B) Magnified view of SO-DARPins at tetrameric "vertex" of the assembly. In the absence of sfGFP, no steric clash modeled (left); in the presence of bound sfGFP (pdb code 6nhv), steric clashes are modeled between bound ligand and neighboring DARPin backbone (right).

4.5 Discussion

LOCs are a novel class of protein nanocages that "open" upon binding a specific protein target to expose its encapsulated moiety. LOCs possess advantages over previously engineered Virus-like particles (VLP) and other nanocages for delivery[8–10]. These prior applications are limited, requiring changes in environmental conditions, such as pH, or spontaneous mechanisms such as pore diffusion or degradation. Other studies have mainly used nanocages as protective carriers, extending the circulation half-life of otherwise serum-unstable cargoes[11,12]. LOCs not only open using a target-based mechanism, but are also modular due to the ease of DARPin sequence mutation to generate binders against new targets.

Even so, this platform comes with a significant shortcoming. The method requires that a DARPin be attachable to a protein cage via rigid alpha helical linker. This requires the existence of a terminal alpha helix to be present on the cage and that it is important enough to the fold of the cage proteins that straining it via ligand-binding will destabilize the assembly. To our knowledge only a select few cages fit these design criteria. Furthermore, ligands targeted by our system need to be of a particular size range; large enough that attachment to our cage induces a steric effect on the cage, but not so large that the DARPin binding site is completely unavailable via diffusion. Nevertheless this work represents a major advance in protein design methods to generate protein cages with reversible assembly; it is the first to exploit specific protein binding as the mechanism of action.

4.6 Methods

LOC Design

Computational alpha helix fusions were modeled similarly to previous works. For α-LOC, we chose nanocage core T33-51 for its subunit terminal helices of at least 6 amino acids in

length. For its DARPin fusion partners, we favored DARPins whose cognate-bound structure is known to clearly predict steric clashes caused by DARPin-ligand binding.

Models were generated by first pairwise fitting an ideal alpha helix to the last six helical residues on the C-terminus of component A. Any residues that lay closer to the actual terminus of the sequence but were not helically structured were ignored for the purposes of pair-fitting, and subsequently deleted to accommodate a fully helical extension. Next, we aligned the N-terminal helical residues of the DARPin fusion partner similarly to previous methods. Again, only structured helical residues were considered for splicing in the final sequence. The DARPin component of the model then slid along the ideal helix until an appropriate orientation with the fewest possible steric clashes in the absence of the cognate ligand, and the highest number of steric clashes in the presence of bound ligand, was visually curated. Final amino acid sequences were redesigned intuitively if necessary to avoid potential local clashes, preferring helix-forming residues. The shortest possible amino acid length for the helical extension was chosen.

Recombinant Protein Expression and Purification

All gene fragments encoding SOR-N3C-α-sfGFP, T33-51-N3C-α-MBP, sfGFP, MBP, and 11S were purchased from IDT and sequencing of plasmids was verified via Sanger Sequencing (Azenta Life Sciences/Genewiz). All genes of interest were inserted into a pET-22b vector (Novagen) via Gibson Assembly.

Expression was performed using BL21 (DE3) cells (New England Biolabs). Transformed cells were grown in LB supplemented with 100μg/mL ampicillin at 37°C to an OD_{600} between 0.8-1.0, before 100μM IPTG was added. Following induction, cultures were grown overnight at 18°C before harvesting via centrifugation at 3500g for 10 minutes.

Cells were resuspended and pressure-homogenized via Emulsiflex (Avestin) in a buffer containing 50mM Tris-HCl pH 8.0, 2% w/v glycerol, protease inhibitor tablet (Pierce), and 20mM imidazole (cage samples contained 500mM NaCl, all other samples contained 250mM NaCl). Lysates were clarified via centrifugation at 20,000 x g for at least 30 minutes.

Lysates were first gravity purified over Ni-NTA resin, washed, and eluted at increasingly higher concentrations of imidazole (50-250mM). Elution fractions containing the cage proteins (confirmed by SDS-PAGE) were then further purified via NGC chromatography system (Bio-Rad) on a Superose 6 Increase 10/300 GL resin column (GE Life Sciences). Sample runs were conducted at 4C with a flow rate of 0.3 ml/min. SEC fractions containing the purest and correctly sized assemblies were then used for subsequent experiments.

SOR LOC Disassembly Assay

0.5 mg of gravity-purified SOR-N3C-α-sfGFP cage protein was mixed with 2x molar excess of sfGFP for 20 minutes at 37C. Sample mixture was then injected and analyzed via SEC.

4.7 References

1. Edwardson, T. G. W., Mori, T. & Hilvert, D. Rational Engineering of a Designed Protein Cage for siRNA Delivery. *J. Am. Chem. Soc.* **140**, 10439–10442 (2018).
2. Marcandalli, J. *et al.* Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **176**, 1420-1431.e17 (2019).
3. Nandwana, V. *et al.* Engineered ferritin nanocages as natural contrast agents in magnetic resonance imaging. *RSC Adv.* **7**, 34892–34900 (2017).
4. Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat Commun* **10**, 1864 (2019).
5. Jones, J. A., Cristie-David, A. S., Andreas, M. P. & Giessen, T. W. Triggered Reversible Disassembly of an Engineered Protein Nanocage. *Angew. Chem. Int. Ed.* **60**, 25034–25041 (2021).
6. Malay, A. D. *et al.* An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **569**, 438–442 (2019).
7. Stupka, I. *et al.* Chemically induced protein cage assembly with programmable opening and cargo release. *Sci. Adv.* **8**, eabj9424 (2022).
8. Han, M., Kickhoefer, V. A., Nemerow, G. R. & Rome, L. H. Targeted Vault Nanoparticles Engineered with an Endosomolytic Peptide Deliver Biomolecules to the Cytoplasm. *ACS Nano* **5**, 6128–6137 (2011).
9. Kickhoefer, V. A. *et al.* Targeting Vault Nanoparticles to Specific Cell Surface Receptors. *ACS Nano* **3**, 27–36 (2009).

10. Ashley, C. E. *et al.* Cell-Specific Delivery of Diverse Cargos by Bacteriophage MS2 Virus-like Particles. *ACS Nano* **5**, 5729–5745 (2011).
11. Schoonen, L., Nolte, R. J. M. & Van Hest, J. C. M. Highly efficient enzyme encapsulation in a protein nanocage: towards enzyme catalysis in a cellular nanocompartment mimic. *Nanoscale* **8**, 14467–14472 (2016).
12. Zhang, T. *et al.* Encapsulation of anthocyanin molecules within a ferritin nanocage increases their stability and cell uptake efficiency. *Food Research International* **62**, 183–192 (2014).

CHAPTER 5

DESIGN OF BETA$_2$ MICROGLOBULIN ADSORBENT PROTEIN CAGES

5.1 Abstract

Beta-2 microglobulin (B2M) is an immune complex protein that is found on the surface of all nucleated human cells. B2M is naturally shed from cell surfaces into the plasma, followed by renal excretion. In patients with impaired renal function, B2M will accumulate in organs and tissues leading to significantly reduced life expectancy and quality of life. While current hemodialysis methods have been successful in managing electrolyte as well as small and large molecule disturbances arising in chronic renal failure, they have shown only modest success in managing plasma levels of B2M and similar sized proteins, while sparing important proteins such as albumin. We describe a systematic protein design effort aimed at adding the ability to selectively remove specific, undesired waste proteins such as B2M from the plasma of chronic renal failure patients. A novel nanoparticle built using a tetrahedral protein cage as a scaffold that presents 12 copies of a B2M-binding nanobody is described. The designed nanoparticle binds specifically to B2M through protein-protein interactions with nanomolar binding affinity (~4.2 nM). Notably, binding to the nanoparticle increases the effective size of B2M by over 50-fold, offering a potential selective avenue for separation based on size. We present data to support the potential utility of such a nanoparticle for removing B2M from plasma by either size-based filtration or by polyvalent binding to a stationary matrix under blood flow conditions. Such applications could address current shortcomings in the management of problematic mid-sized proteins in chronic renal failure patients.

5.2 Introduction

For patients afflicted by chronic renal failure (CRF), hemodialysis is a lifesaving and life enhancing therapy wherein a hemodialyzer aims to fulfil the blood filtering function of healthy kidneys. However, the methods and technologies used to filter some waste products from the

plasma of CRF patients are poorly effective, resulting in high levels of many uremic toxins that can ultimately diminish long-term prognosis[1]. This limitation in current technology is in large part due to the lack of selectivity with which some waste products are removed from patient dialysate; discrimination of waste proteins from required beneficial proteins occurs primarily by size, diffusion across a semipermeable membrane or via membrane adsorption. Established dialysis technologies are most effective at removing small uremic toxins and balancing volume and electrolytes. but can struggle to remove larger molecules, especially middle-sized proteins (2 to 4 nm)—which diffuse more slowly—while retaining necessary serum components such as albumin and some immunoglobulins. As a consequence, patients on long-term dialysis therapy often experience an unintended increase in the concentration of proteins and molecules that fall in between the molecular weight of small molecules and larger proteins, often referred to as 'middle molecules'.

One of these difficult-to-remove middle molecules is the 11.8 kDa protein beta-2 microglobulin (B2M). The presence of many other middle molecules such as– albumin, immunoglobulins, transthyretin, tumor necrosis factor, various interleukins and other cytokines – all of similar size, some of which must be retained while others must be removed, sets a difficult task for size-based filtration systems that do not recognize specific targets.

B2M is a structural protein of the mutual histocompatability complex I (MHC-I), which resides on the surface of cell membranes of all nucleated cells, and so plays a role in normal immunologic function. B2M is a member of a broad class of proteins known to transition from a functional folded state to a misfolded amyloid state[2–5]. Approximately 2-4 mg/kg/day[6] of B2M is produced per day with 99%[7] excreted by the healthy kidney so that the normal plasma concentration of B2M is less than 2 mcg/mL. CRF leads to reduced renal excretion of B2M

resulting in chronically elevated plasma concentrations of 50-100 mcg/mL, in which the amyloid state will commonly arise. In CRF, the high concentration of circulating B2M often results in amyloid deposition in a condition known as dialysis related amyloidosis (DRA). DRA is characterized by the formation of insoluble deposits of B2M throughout the body, resulting in pain and physiological dysfunction, with effective treatments mostly limited to pain management or surgical intervention[8]. Because of the limited treatment options for DRA, the development of methods to control blood B2M concentrations in patients on hemodialysis is of considerable medical importance.

Many of the efforts to improve the removal of middle molecules such as B2M during hemodialysis rely on tuning the characteristics of the membranes used or on using non-specific protein-adsorbent media. These innovations have included changing the pore size or chemistry of hemodialyzer membranes, or introducing a hydrophobic protein-adsorbent media to bind B2M and remove it from plasma[9]. These advances in technology have led to somewhat limited improvements in patient prognosis, with notable challenges in maintaining efficacy over extended durations of dialysis[10]. More recent studies have demonstrated the utility of using B2M binding proteins (including antibody fragments and nanobodies) or peptides attached to a matrix to sequester B2M and facilitate its removal[11–15]. Methods that remove B2M with greater specificity could improve efficacy without the unintended consequence of removing other similarly sized molecules in serum, including albumin and immunoglobulins.

We sought to investigate whether new ideas in protein engineering could be used to design a high-avidity soluble nanoparticle with favorable performance as an immunosorbent. To explore this, we investigated the viability of designed protein nanocages, displaying B2M binding domains in a polyvalent fashion on their surface, for the removal of B2M from solution.

We report on the successful creation of a novel protein nanoparticle for that purpose, along with important findings on stability and B2M binding and removal capabilities.

5.3 Methods

Cloning and Expression: DNA encoding BAC nanoparticles was ordered from Twist Biosciences and cloned using Gibson assembly into the bacterial pET22b expression vector. Expression of nbBACs was performed by cloning sequence-verified BAC-encoding expression vectors into Shuffle T7 Express lysY cells (New England Biolabs) to allow for proper folding of disulfide bond-containing nanobodies. Cells were grown in autoinduction media[16] at 25C for 48 hours. Cells were harvested at 4000xg and stored at -20C until purification.

BAC Purification: Frozen pelleted cells were solubilized in lysis buffer containing 50mM TRIS pH 8.0, 250mM NaCl, and protease inhibitor tablets (Pierce). Cells were lysed using a C3 Emulsiflex with 4 passages through the instrument. Lysate was clarified by centrifugation at 18000xg for 35 minutes after which the supernatant fraction was passed over a gravity-flow column containing Ni-NTA functionalized agarose beads (Thermo Fisher). The column was washed with lysis buffer containing imidazole of sequential concentrations: 50mM, 75mM, and 100mM. BACs were eluted from the column using lysis buffer supplemented with 500mM imidazole. BAC proteins were then buffer-exchanged using dialysis into assay buffer containing: 50mM TRIS, 150mM NaCl, and 0.02% Tween-20. SEC was performed to isolate correctly assembled nanoparticles using a Superose-6 column (Cytiva Life Sciences). The identity of fractions containing purified BACs were verified using SDS-PAGE.

Purification of B2M: Avi and 6xHis- tagged B2M was expressed and purified using established methods with refolding[17,18]. B2M was then enzymatically biotinylated with the addition of BirA enzyme according to existing protocols.

Binding Affinity Measurements: Binding affinities were determined using an Octet RED96 BLI system (Sartorius). Biosensors functionalized with streptavidin were used to bind biotinylated B2M at a concentration of 2.5ug/ml and tested for binding with soluble BACs. Buffer used for experiments was 50mM TRIS pH 8.0, 150 mM NaCl, 0.02% Tween-20.

Immunoblots: 3ul aliquots of flow through samples from spin columns were applied to nitrocellulose paper, allowed to dry then repeated at the same location once. Established methods for western blotting were then used develop and image the blot. Anti-biotin antibodies were used to recognize biotinylated B2M (Invitrogen). For western blots, Mini-Protean Any-KD SDS-PAGE gels (Bio-Rad) were used to separate proteins, and then transferred to nitrocellulose membrane (GE-Healthcare) using a Trans-Blot SD system (Bio-Rad). The membrane was rocked in blocking buffer: TBST with 5% milk powder for 1 hour at room temperature with rocking. Primary anti-B2M antibodies were purchased from AbClonal (A1562) and used at a dilution of 1:1000 in blocking buffer. Membrane was allowed to incubate with primary antibody for 1 hour at room temperature. The membrane was washed thrice in TBST, then incubated with anti-rabbit goat secondary antibody with HRP conjugate for an additional 1 hour at room temperature. The membrane was washed of excess secondary antibody thrice with TBST, then incubated with Clarity Western ECL substrate (Bio-rad) and imaged on an Azure imager (Azure Biosystems).

Size-excluded Flow Through Experiment: 100nm B2M with and without 1uM BAC was added to the supernatant of a centrifuge Amicon Ultra 100kDa cutoff microcentrifuge concentrator, and then spun for 2.5 minutes at 4000xg. Flow through and supernatant were collected for further experimentation.

Serum Challenge Assay: SEC-purified BAC was added to human serum and incubated at 37C for 4 hours. Serum mixture was then flowed over a column containing HisPur Ni-NTA resin

(Thermo Fisher Scientific) and washed and eluted with buffers identical to those used in IMAC purification. Elution fraction was analyzed by SDS-PAGE and concentrated and injected into a Superose 6 gel filtration column (Cytiva Life Sciences) for oligomeric state determination.
B2M Removal Experiment: 1ml His-Trap columns (Cytiva life sciences) were used as a stationary matrix to immobilize BACs. Following incubation with SEC-purified BACs, human serum supplemented with 50ug/ml of B2M was flowed over the column and the flow-through was collected. Column was washed with binding buffer (50mm Tris pH 8.0, 200mm NaCl, 20mm Imidazole), and then eluted with elution buffer containing 50mm Tris pH 8.0, 200mm NaCl, 500mm Imidazole. These fractions were analyzed by western blot according to the previously described protocol.

5.4 Results

Protein Design Approach –

We first sought to generate novel protein cages displaying B2M binders on their surface; we refer to these as B2M adsorbent cages (BACs). Designed protein nanocages have architectures that resemble viral capsids in certain respects (e.g. taking the forms of Platonic solids such as a tetrahedron, cube, or icosahedron), and have increasingly been at the heart of engineering studies that exploit their large size and symmetry[19–26]. Such cages are composed of tessellating protein molecules – typically one or two distinct protein subunit types each present in multiple copies – which self-assemble to form symmetric supramolecular assemblies or nanoparticles. Their structures make polyvalent attachment and display of additional outward-facing proteins (or other molecule types) a relatively routine engineering process. We noted two potential benefits of applying protein cages to the problem of B2M removal technologies: (1) polyvalent display could lead to advantageous binding properties through well-known avidity

effects (reviewed extensively by Varner, Kane and colleagues[27], and (2) the large size of the nanocage could lead to unique opportunities for filtration, without removing beneficial plasma proteins (which are generally much smaller than the nanocage) in the process. There could also be advantages to technologies that bind and remove B2M while in solution rather than bound to highly dense stationary matrices, as the former approach might limit the chances of inadvertently creating and seeding B2M amyloid.

For design and testing of BAC nanoparticles, we selected two previously characterized protein cages to serve as scaffolds for the development of functionalized nanoparticles. We generated initial designs based on two protein cages that have proven to be robust to various biochemical conditions and engineering modifications in our previous studies: the tetrahedral cage known as T33-51[28], and the icosahedral cage known as I53-50[29] (Figure 1a). The divergent sizes and symmetries of these candidates offered distinct engineering opportunities. Designs were generated by genetically fusing B2M binding nanobodies[30] to the outward-facing termini of the protein cage subunits, thereby creating nanobody BACs (nbBACs) (Figure 1b-c). Both of the candidate cages have a polypeptide chain terminus that extends to the outside surface of the cage, which was important to permit the scaffold to assemble correctly after modification, with the binding domains accessible on the exterior surface to interact with B2M. Note that both cages used here are of the 2-component type, being comprised of two distinct subunit types A and B. Accordingly, the stoichiometry of the tetrahedral T33-51 cage is A12B12 while the icosahedral I53-50 cage has stoichiometry A60B60, meaning the former cage will display 12 copies of a binding domain on its surface whereas the latter cage will display 60 copies.

For the first round of nbBACs, we employed a minimal polypeptide linker (a GGS sequence) to connect a C-terminus of a protein cage component to the N-terminus of the B2M nanobody. For

these starting candidates, we refer to the nanoparticle based on the smaller T33-51 cage as nbBAC1_SL (for short linker) and the nanoparticle based on the I53-50 cage as nbBAC2 (Supplemental Data Table 1).

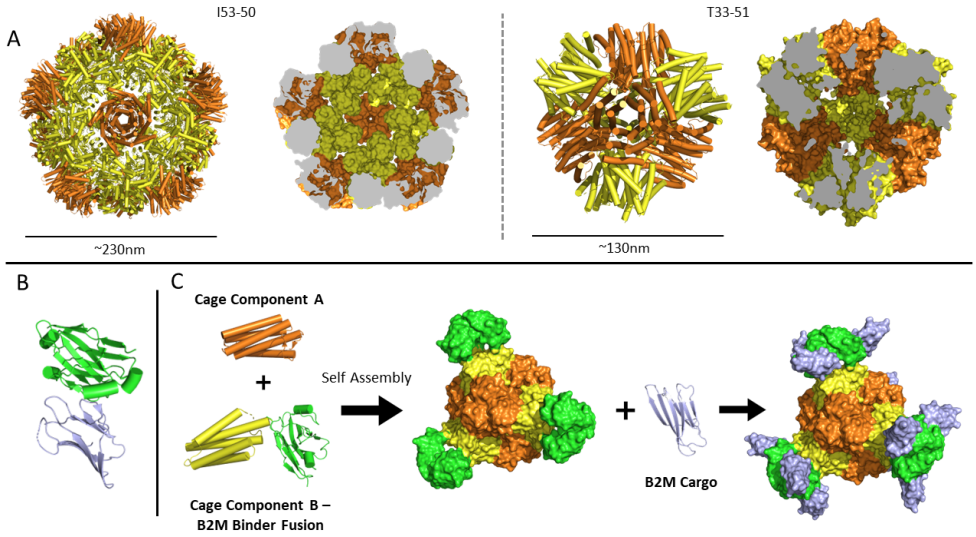


Figure 5.1. Design schema and building blocks of beta-2 microglobulin (B2M) adsorbent cage (BAC) nanoparticles. A: Structural models of the I53-50 icosahedral protein cage and the T33-51 tetrahedral protein cage tested as scaffolds in several BAC designs. The protein structures are depicted as secondary structure cartoons (left) and as slices through space-filling models (right). (B) Structure of the anti-B2M nanobody (green) used in the design of BACs. B2M (purple) is shown in its bound configuration. (C) Design of a BAC from protein cage components (yellow and orange), with component B genetically fused to the anti-B2M nanobody. A fully assembled BAC shown in apo form as well as bound to B2M ligands (right).

Testing and Design Optimization on Cages -

The proteins designed and analyzed here were over-expressed in E. coli cells and purified by affinity chromatography (see Methods). Note that our designed BACs are intended to self-assemble and to therefore appear in their assembled forms upon expression and purification. For binding tests, our BACs were mixed with purified B2M protein and assessed for binding using size exclusion chromatography (SEC). In the event of a stable interaction between the BAC and B2M, co-elution of proteins would be expected. If no interaction occurs, B2M would be expected to elute much later due to its significantly smaller size relative to the BAC assembly.

For experiments on both of our initial BAC designs (nbBAC1_SL and nbBAC2), we noted visible co-elution of B2M with the proteins making up each of the BAC assemblies (Figure 2). However, closer examinations led to two important observations on the first round of nbBAC designs: (1) nbBAC1_SL displayed limited stability in its assembled state, precipitating out of solution over time, and exhibited a propensity to rapidly disassemble into smaller subunits (Supplemental data figure 1); and (2) nbBAC2 displayed sub-stoichiometric binding to B2M when visualized on SDS-PAGE. These observations, which would have limited downstream utility, motivated a second round of designs.

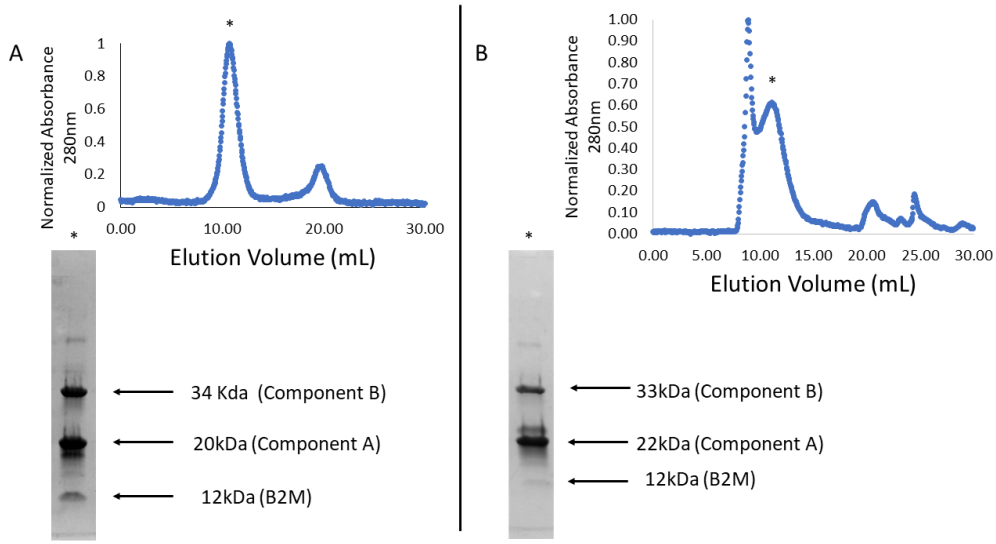


Figure 5.2. Assembled BAC nanoparticles bind to and coelute with B2M cargo. A: nbBAC1_SL mixed with B2M elutes in SEC at a volume consistent with proper cage assembly (denoted with asterisk). SDS-PAGE analysis of the cage peak fraction indicates B2M co-elutes with BAC. B: nbBAC2 mixed with B2M elutes in SEC at a volume consistent with proper cage assembly (denoted with asterisk). SDS-PAGE analysis of the cage peak fraction indicates B2M co-elutes with BAC.

We rationalized that the issues with nbBAC1_SL stability may have been the result of steric clashes between neighboring nanobody domains projecting from the assembled cage via short linkers. To test that hypothesis, we generated a new BAC design based on nbBAC1, but with a longer 12-residue (GS)6 linker, which we designated nbBAC1_LL (long linker)

(Supplemental Data Table 1). We went on to test the assembly and binding properties of nbBAC1_LL using size exclusion chromatography (SEC) and noted improved stability of this BAC as well as a higher yield of protein eluting at the expected elution volume for a tetrahedral cage assembly (Figure 3). Densitometric analysis of B2M co-eluting with assembled BAC on SEC revealed an approximate ratio of 2:1 stoichiometry of B2M binders to B2M protein. In other words, mixing excess B2M with BAC proteins resulted in roughly half occupancy of our BAC binders, corresponding to about six B2M molecules bound to an individual nanoparticle. This improved nanoparticle was evaluated more extensively in further experiments (discussed subsequently).

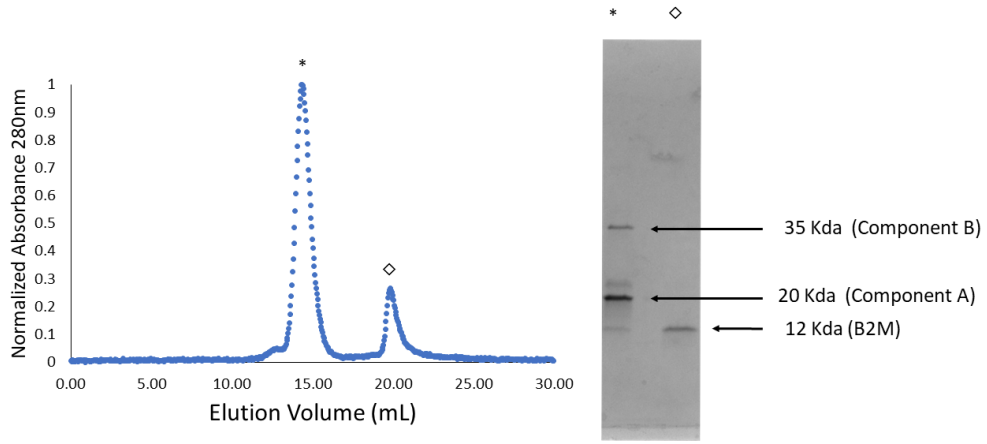


Figure 5.3. Purification and B2M binding of an improved B2M-binding nanoparticle, nbBAC1_LL. The improved nbBAC1_LL, mixed with B2M, elutes in SEC (left) at a volume consistent with proper cage assembly (denoted by an asterisk). SDS-PAGE analysis (right) of the cage peak fraction indicates B2M co-elutes with BAC.

The larger BAC design, nbBAC2 based on the I53-50 cage, suffered from a limited capacity for B2M binding. In this case, we hypothesized that the binding site for the nanobody may have been partially occluded by the scaffold, so we made use of the presence of both N- and C-termini on the outside of the I53-50 protein cage scaffold to generate an alternative nbBAC where the

nanobody was fused through its N-terminus rather than the C-terminal fusion used for nbBAC1 and nbBAC2, thus generating nbBAC2N (Supplemental Data Table 1). However, this alternative construct displayed limited expression and solubility and no visible assembly into icosahedral geometry. Considering its apparent inability to support the fusion of exterior nanobody domains, the I53-50 cage was abandoned as a basis for our BACs.

Assays on nbBAC1_LL for B2M Removal –

To characterize the ability of our leading BAC candidate, nbBAC1_LL, to bind and sequester B2M, we performed a size-based retention assay utilizing a centrifugal filter with a molecular weight cutoff of 100kDa, which falls between the size of B2M and our BAC designs. When centrifuged, a stable interaction between B2M and BAC should prevent B2M from passing through the filter, whereas B2M alone should pass through the filter unimpeded (Figure 4). That is, adding the BAC to a solution of B2M should prevent any protein material from passing through the size filter. Our experiments confirmed that behavior. Using a BSA assay, we measured the protein concentration in the flowthrough of both experiments, testing 100 nM B2M plus or minus addition of nbBAC1_LL at 1 uM. The resulting protein concentration in the flow through was determined to be at least 5 times lower when nbBAC1_LL was added (above the centrifugal filter) compared to when it was absent. The BAC effectively sequesters the B2M in a large complex that is unable to pass through the filter membrane. Owing to the detection limit of the BSA assay, the precise depletion factor for removal of B2M in this size-based experiment is likely considerably higher than the limiting value of 5 noted above. Indeed, B2M was undetectable by immunoblot in the flow-through fraction; chemiluminescent signal is only visible on the blot containing flowthrough from the experiment without nbBAC1_LL added to the B2M (Supplemental Data Figure 2).

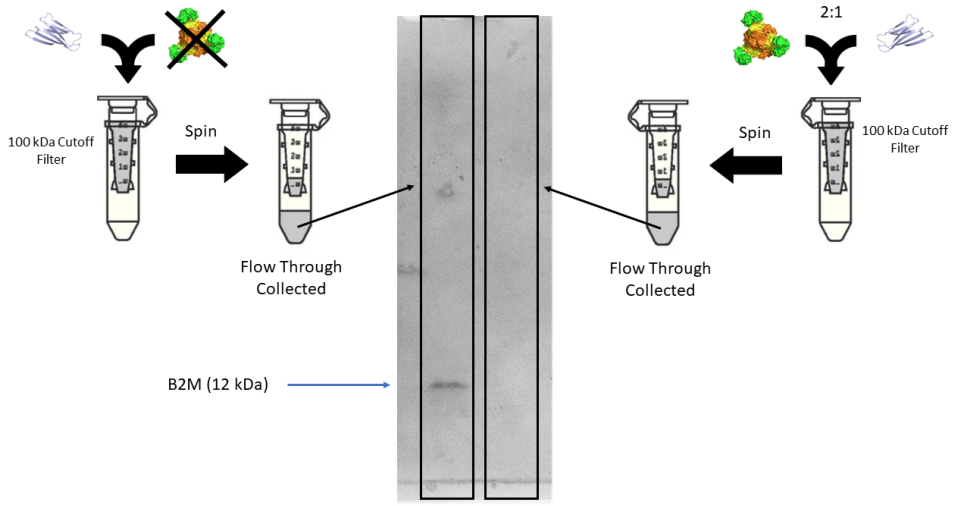


Figure 5.4. Binding of B2M by nbBAC1_LL in a B2M retention assay. nbBAC1_LL binds to soluble B2M and retains it in the supernatant of a concentrator column with a molecular weight cutoff of 100kDa (right). When no nbBAC is added to the column, B2M flows through the membrane unimpeded (left). B2M has a MW of 13.8 kDa and nbBAC has a MW of 662 kDa. An SDS-PAGE gel shows the composition of the flow through in each experiment. B2M is only present in the gel run on the flow through of the experiment without nbBAC added, showing that B2M is fully retarded by the interaction between nbBAC1_LL and B2M.

We investigated the binding of B2M to nbBAC1_LL using biolayer interferometry (BLI). In order to avoid potential artifacts -- e.g. cage disruption and/or occlusion of binding sites -- that might occur by attaching the BAC to the BLI sensor tip, we elected instead to attach B2M to the tip and monitor the binding of nbBAC1_LL following standard protocols (see Methods). We obtained binding kinetic curves under a series of nbBAC1_LL concentrations, ranging from 1 nM to 1000 nM. For concentrations in the range of 1 nM to 100 nM, we obtained excellent fits to theoretical curves for binding to identical and independent binding sites (Supplemental Data Figure 3). In contrast, for concentrations of mbBAC1_LL higher than 100 nM, we obtained binding curves inconsistent with simple binding phenomena. In particular, attempts to fit response curves to a model based on identical and independent binding sites illuminated a tighter binding for occupation of initial sites and weaker binding for occupation of subsequent sites. We interpreted this negative cooperative effect to steric occlusion in binding multiple BACs to the

sensor tip; the ligand (nbBAC1_LL in this case) is much larger than in typical BLI experiments. From the data at moderate concentrations of nbBAC1_LL, the binding affinity (Kd) was measured to be 4.2 nM (+/- 0.3 nM). This presents an improvement by about an order of magnitude over the published Kd value of 58 nM for the same nanobody when used in isolation [30]. We further assessed whether this designed BAC would remain stable under conditions relevant for dialysis applications. We incubated nbBAC1_LL at 37 °C in human serum for 4 hours and then assayed its oligomeric state. After pulling down nbBAC1_LL based on its poly-histidine tag using Ni-NTA resin, and then eluting with imidazole, we analyzed the major species by SEC and established that the size corresponded to that of the natively assembled BAC nanoparticle (Supplemental Data Figure 4). The designed nbBAC1_LL nanoparticle therefore appears stable to those conditions.

In clinical dialysis applications, the slow speed of filtering proteins by size-selective membranes can be limiting, motivating applications based on selective removal of proteins flowing over a stationary binding matrix. Accordingly, to evaluate its capacity as a component of a binding matrix, we immobilized nbBAC1_LL on a metal affinity (His-trap) column and then flowed human serum supplemented with 50 ug/ml B2M over the matrix and tested the flowthrough for B2M. We also tested the column eluate after washing off the BAC from the column using imidazole. We analyzed samples for B2M using western blotting, and observed B2M only in the BAC elution fraction, with no detectable amounts in the flow-through (Figure 5). This shows that, under the conditions tested, which were intended to mimic the serum of afflicted dialysis patients, B2M is practically removed to completion by a single passage over the nbBAC1_LL-bound material.

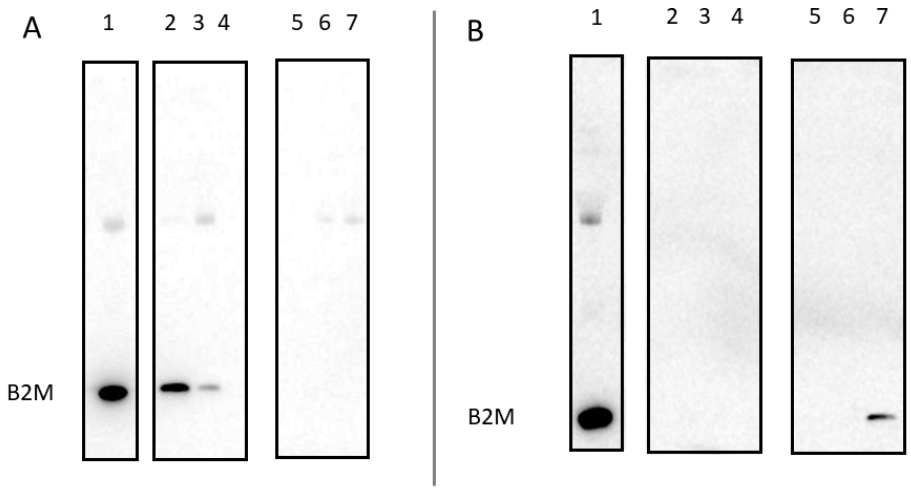


Figure 5.5. Nanoparticle nbBAC1_LL bound to a stationary matrix fully removes B2M from human serum. SDS-PAGE gels are analyzed by Western blotting with an anti-B2M antibody. A: Analysis of sequential flow through fractions of serum supplemented with B2M flowed over resin, either with nbBAC1_LL bound to the column (lanes 5-7) or without (lanes 2-4). Lane 1 is 20x diluted serum supplemented with B2M. B: Analysis of resin-bound components from serum supplemented with B2M, eluted from the resin using 1M imidazole buffer, either with nbBAC1_LL bound to the column (lanes 5-7) or without (lanes 2-4).

5.5 Discussion

A key aim of the present study was to design and evaluate a new type of protein-based polyvalent binding material that might be suitable for depleting human plasma of certain proteins, particularly B2M, which accumulates to toxic levels and then forms pathogenic amyloid deposits in long-term dialysis patients. We chose designed protein cages as a framework for creating such materials, to exploit their size and their polyvalent avidity when modified with binding domains. Among protein cages that have been created to date, a few have proven amenable to further modification to display fused proteins on their outer surfaces[25,27,31–33]. That is, they have outwardly disposed chain termini and retain their stability and self-assembling properties after fusion of additional protein domains; many protein cages do not meet those criteria. In the present study, we found it necessary to examine multiple candidate protein

cages and fusion strategies (e.g. linker lengths) to obtain a robust nanoparticle with favorable B2M binding properties. The specific construct demonstrated here may be a useful starting point for clinical testing.

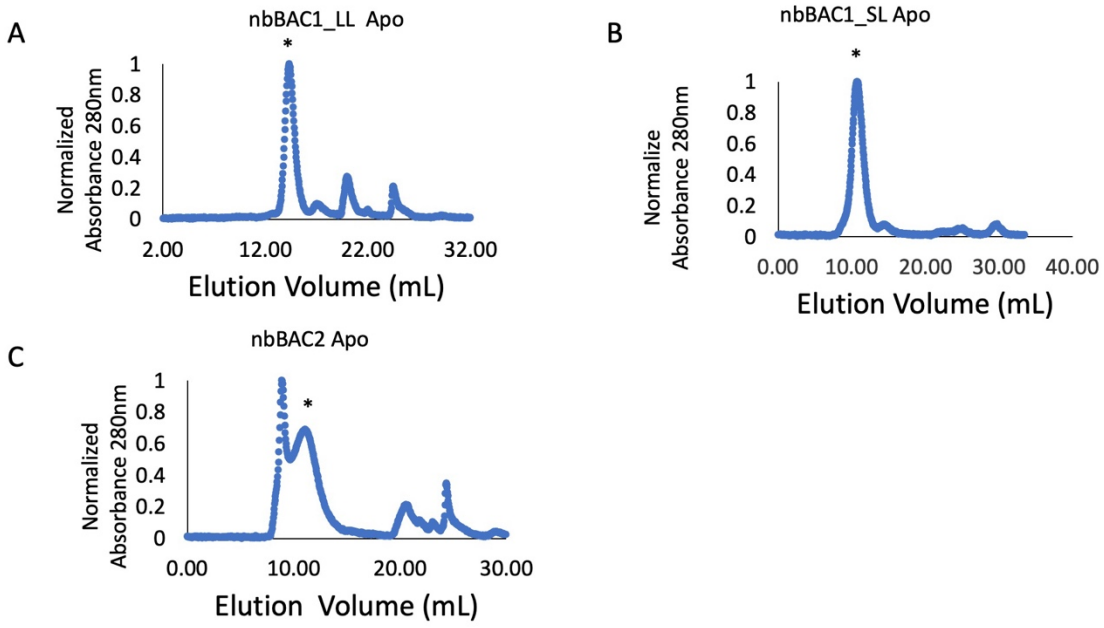
The best performing nanoparticle, nbBAC1_LL, complexes with B2M to produce an approximately 732 kDa assembly with a demonstrated binding capacity of about six copies of B2M on its exterior. We showed that the nanoparticle has the ability to remove B2M in a size-based separation, i.e. by employing a semi-permeable membrane filter that would allow smaller proteins (e.g. albumin) to pass. For possible flow-based applications, we envisage that this nanoparticle could be affixed to a stationary matrix. Note that in the experiments here we affixed our BAC to a nickelated matrix based on polyhistidine tails; for clinical applications, a more specific attachment might be important. A straightforward option would be to attach a biotinylated version of our BAC to a polymeric matrix functionalized with streptavidin. Streptavidin has been employed extensively as a platform for chemical biology and medical device applications[34]. Such a medium could allow blood to pass through, but would retain B2M bound to the stationary phase. This could allow it to be used in series with existing dialyzers with relatively minor modification of existing techniques. Like other immunosorbents, the column could be reused by eluting BAC nanoparticles, followed by sterilization and regeneration with fresh BAC nanoparticles. The experiments in the present study emphasize the effectiveness of our BACs in removing B2M potentially with only one passage of serum over the medium. It is notable that most existing size-based filtration methods involve passage of blood over filtration membranes up to 20 times, while realizing considerably less reduction in B2M concentrations. The materials described here might therefore offer advantages for lowering B2M levels and helping to prevent DRA amyloidosis in patients on chronic dialysis.

Regarding limitations, as demonstrated so far, our BAC materials are only semi-renewable. Whereas a stationary matrix should ideally be fully regenerative, we have thus far not demonstrated a method to dissociate B2M, tightly bound to BACs, without sacrificing the integrity of the BAC nanoparticles. Such an improvement could be critical, given the large quantities of B2M (on the range of hundreds of milligrams) present in one blood volume of an afflicted patient. In summary, with suitable modifications to address issues of regeneration, the novel nanoparticle platform described here could see utility, in conjunction with existing dialysis protocols, for alleviating DRA amyloidosis in large populations subject to long-term dialysis.

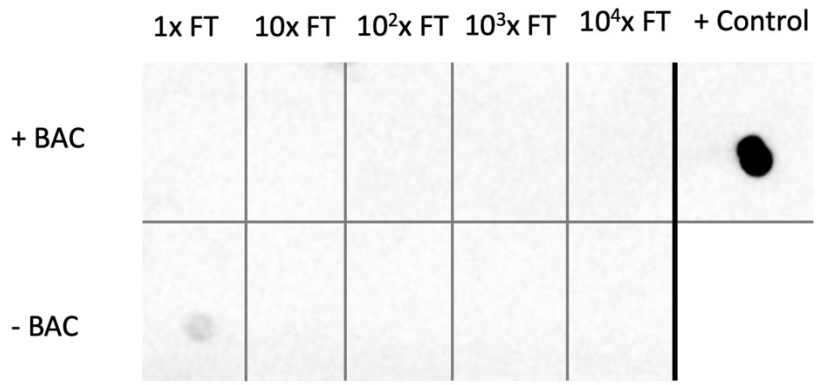
5.6 Acknowledgements

The authors thank Pascal F. Egea for helpful conversations and use of reagents and Genesis Falcon for assistance with the BLI instrument. This work was funded by National Institutes of Health, National Institute of Allergy and Infectious Diseases grant R01AI081146 (TOY) and NIH training grant T32 GM067555 (JEM), and UCLA Graduate Division Dissertation Year Fellowship (JEM).

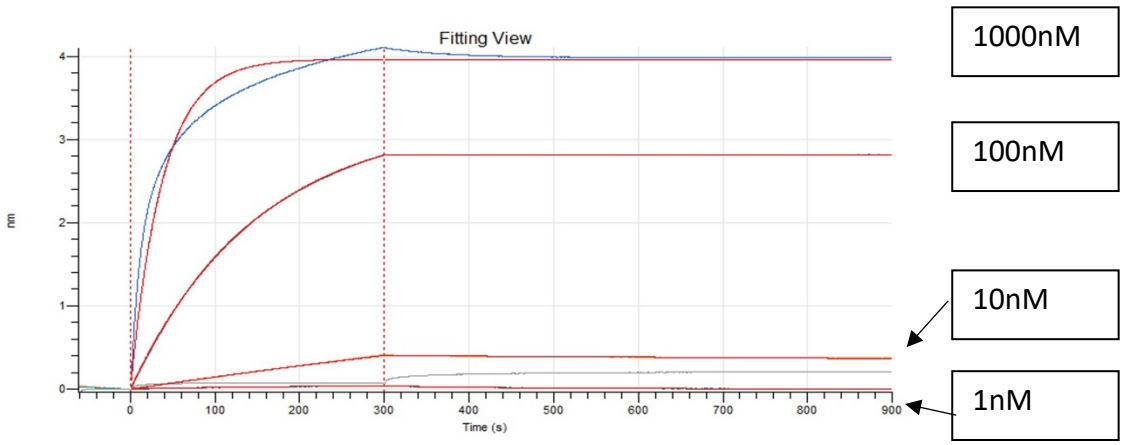
5.7 Supplemental Data



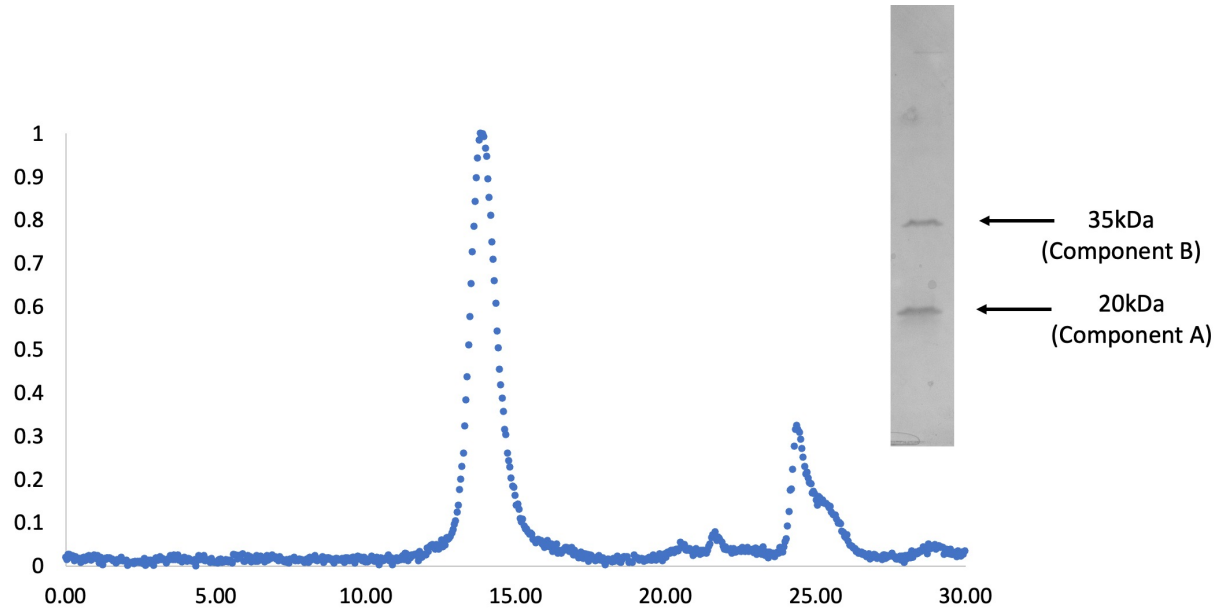
Supplemental Data Figure S5.1. SEC elution profiles of BAC nanoparticles. SEC chromatograms for nbBAC1_LL (A), nbBAC1_SL (B), and nbBAC2_SL (C). Assembled BAC nanoparticles are denoted with asterisks.



Supplemental Data Figure S5.2. Immunoblot analysis of flowthrough from the B2M retention assay. Serial dilutions of flow through from the size-filtration retention assay either with (top) or without (bottom) nbBAC1_LL added to supernatant were analyzed for B2M via immunoblot. A positive control (1uM B2M) is depicted in top right.



Supplemental Data Figure S5.3 Measurement of binding affinity of B2M to nbBAC1_LL. Biolayer interferometry was used to estimate the equilibrium dissociation constant of nbBAC for B2M. Based on experiments with B2M concentrations between 1 nm and 100 nm, the estimated Kd value is 4.2 nM (+/- 8 nM). Curves corresponding to B2M concentrations of 1000nm, 100nm, 10nm, and 1nm are shown.



Supplemental Data figure S5.4 nbBAC1_LL is stable in human serum. nbBAC1_LL incubated in serum is purified and analyzed by SDS-PAGE and by SEC indicating nbBAC1_LL maintains correct assembly geometry in dialysis conditions.

nbBAC1_SL	>A component MFTRRGDQGETDLANRARVGKDSPVVEVQGTIDELNSFIGYALVLSR WDDIRNDLFRIQNDLFVLGEDVSTGGKGRTVTMDMIIYLIKRSVEMKA EIGKIELFVVPGGSVESASLHMARAVSRRLERRIKAASELTEINANVLL YANMLSNILFMHALISNKRLNIPEKIWSIHRVSLE >B component MRITTKVGDKGSTRLFGGEEVWKDDPIIEANGTLDELTSFIGEAKHYV DEEMKGILEEIQNDIYKIMGEIGSKGKIEGISEERIKWLAGLIERYSEMV NKLSFVLPGGTLESAKLDVCRTIARRAERKVATVLREFGIGTLAAIYLA LLSRLLFLLARVIEIEKNKLKEVRSGGSQVQLQESGGGSVQAGGSLRL SCAASGYTDSRYCMAWFRQAPGKEREWVARINSGRDITYYADSVKG RFTFSQDNAKNTVYLQMDSLEPEDTATYYCATDIPLRCRDIVAKGGD GFRYWGQGTQVTVSSHHHHHH
nbBAC1_LL	>A component MFTRRGDQGETDLANRARVGKDSPVVEVQGTIDELNSFIGYALVLSR WDDIRNDLFRIQNDLFVLGEDVSTGGKGRTVTMDMIIYLIKRSVEMKA EIGKIELFVVPGGSVESASLHMARAVSRRLERRIKAASELTEINANVLL YANMLSNILFMHALISNKRLNIPEKIWSIHRVSLE >B component MRITTKVGDKGSTRLFGGEEVWKDDPIIEANGTLDELTSFIGEAKHYV DEEMKGILEEIQNDIYKIMGEIGSKGKIEGISEERIKWLAGLIERYSEMV NKLSFVLPGGTLESAKLDVCRTIARRAERKVATVLREFGIGTLAAIYLA LLSRLLFLLARVIEIEKNKLKEVRSGGSGGGSGGGGSQVQLQESGGG SVQAGGSLRLSCAASGYTDSRYCMAWFRQAPGKEREWVARINSGR DITYYADSVKGRFTFSQDNAKNTVYLQMDSLEPEDTATYYCATDIPLR CRDIVAKGGDGFRYWGQGTQVTVSSHHHHHH
nbBAC2	>A component MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVPDAD TVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQF CKEKGVFYMPGVMTPTELVKAMKLGHDILKLFPGEVVGPQFVKAMK GPFPNVKFVPTGGVNLDNVCKWFKAGVLAVGVGKALVKGKPDEVRE KAKKFVKKIRGCTE >B component MNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRFA VDVFDVPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASA VINGMMNVQLNTGVPVLSAVLTPHNYDKSKAHTLLFLALFAVKGMEA ARACVEILAAREKIAAGSGGSGGSGGSQVQLQESGGGSVQAGGSLR LSCAASGYTDSRYCMAWFRQAPGKEREWVARINSGRDITYYADSVK GRFTFSQDNAKNTVYLQMDSLEPEDTATYYCATDIPLRCRDIVAKGG DGFRYWGQGTQVTVSSAHHSEDPHHHHHH

nbBAC2N	>A component MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVPDAD TVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQF CKEKGVFYMPGVMTPTELVKAMKLGHDILKLFPGEVVGPQFVKAMK GPFPNVKFVPTGGVNLDNVCKWFKAGVLAVGVGKALVKGKPDEVRE KAKKFVKKIRGCTE >B component MQVQLQESGGGSVQAGGSLRLSCAASGYTDSRYCMAWFRQAPGK EREWVARINSGRDITYYADSVKGRFTFSQDNAKNTVYLQMDSLEPED TATYYCATDIPLRCRDIVAKGGDGFRYWGQGTQVTVSSGGSGGSGG SNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRFAV DVFDVPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAV INGMMNVQLNTGVPVLSAVLTPHNYDKSKAHTLLFLALFAVKGMEAA RACVEILAAREKIAAGSHHHHHH

Supplemental Data Table S5.1: BAC sequences.

5.8 References

1. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. & Radford, S. E. Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell* **41**, 161–172 (2011).
2. Iadanza, M. G. *et al.* The structure of a β2-microglobulin fibril suggests a molecular basis for its amyloid polymorphism. *Nat. Commun.* **9**, 4517 (2018).
3. Winchester, J. F., Salsberg, J. A. & Levin, N. W. Beta-2 microglobulin in ESRD: an in-depth review. *Adv. Ren. Replace. Ther.* **10**, 279–309 (2003).
4. Ohashi, K. Pathogenesis of β_2 -microglobulin amyloidosis. *Pathol. Int.* **51**, 1–10 (2001).
5. Floege, J. *et al.* Clearance and synthesis rates of beta 2-microglobulin in patients undergoing hemodialysis and in normal subjects. *J. Lab. Clin. Med.* **118**, 153–165 (1991).
6. Barton, K. T. *et al.* Using the newer Kidney Disease: Improving Global Outcomes criteria, beta-2-microglobulin levels associate with severity of acute kidney injury. *Clin. Kidney J.* **11**, 797–802 (2018).
7. Portales-Castillo, I., Yee, J., Tanaka, H. & Fenves, A. Z. Beta-2 Microglobulin Amyloidosis: Past, Present, and Future. *Kidney360* **1**, 1447–1455 (2020).
8. Furuyoshi, S. *et al.* New adsorption column (Lixelle) to eliminate beta2-microglobulin for direct hemoperfusion. *Ther. Apher. Off. J. Int. Soc. Apher. Jpn. Soc. Apher.* **2**, 13–17 (1998).
9. Masakane, I. & Sakurai, K. Current approaches to middle molecule removal: room for innovation. *Nephrol. Dial. Transplant.* **33**, iii12–iii21 (2018).
10. Daniels, C. M., Woolverton, E. M., Sprague, S. M. & Ameer, G. A. Assessment of the stability of an immunoadsorbent for the extracorporeal removal of Beta-2-microglobulin from blood. *Blood Purif.* **23**, 287–297 (2005).

11. Zhang, L., Zang, B., Huang, C., Ren, J. & Jia, L. One-step Preparation of a VHH-based Immunoadsorbent for the Extracorporeal Removal of β2-microglobulin. *Mol. Basel Switz.* **24**, 2119 (2019).
12. Huang, C., Ren, J., Ji, F., Muyldermans, S. & Jia, L. Nanobody-Based high-performance immunosorbent for selective beta 2-microglobulin purification from blood. *Acta Biomater.* **107**, 232–241 (2020).
13. Kang, S. *et al.* Removal of intact β2-microglobulin at neutral ph by using seed-conjugated polymer beads prepared with β2-microglobulin-derived peptide (58-67). *Biotechnol. Prog.* **27**, 521–529 (2011).
14. Grovender, E. A. *et al.* Single-chain antibody fragment-based adsorbent for the extracorporeal removal of beta2-microglobulin. *Kidney Int.* **65**, 310–322 (2004).
15. Cannon, K. A., Ochoa, J. M. & Yeates, T. O. High-symmetry protein assemblies: patterns and emerging applications. *Curr. Opin. Struct. Biol.* **55**, 77–84 (2019).
16. Edwardson, T. G. W., Mori, T. & Hilvert, D. Rational Engineering of a Designed Protein Cage for siRNA Delivery. *J. Am. Chem. Soc.* **140**, 10439–10442 (2018).
17. Butterfield, G. L. *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420 (2017).
18. Edwardson, T. G. W., Tetter, S. & Hilvert, D. Two-tier supramolecular encapsulation of small molecules in a protein cage. *Nat. Commun.* **11**, 5410 (2020).
19. Divine, R. *et al.* Designed proteins assemble antibodies into modular nanocages. *Science* **372**, eabd9994 (2021).
20. McConnell, S. A. *et al.* Designed Protein Cages as Scaffolds for Building Multienzyme Materials. *ACS Synth. Biol.* **9**, 381–391 (2020).

21. Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102–106 (2013).
22. Marcandalli, J. *et al.* Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **176**, 1420-1431.e17 (2019).
23. Cannon, K. A. *et al.* Design and structure of two new protein cages illustrate successes and ongoing challenges in protein engineering. *Protein Sci.* **29**, 919–929 (2020).
24. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
25. Domanska, K. *et al.* Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic β2-microglobulin variant. *Proc. Natl. Acad. Sci.* **108**, 1314–1319 (2011).
26. Brouwer, P. J. M. *et al.* Lassa virus glycoprotein nanoparticles elicit neutralizing antibody responses and protection. *Cell Host Microbe* **30**, 1759-1772.e12 (2022).
27. Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.* **10**, 1864 (2019).
28. Edwardson, T. G. W. *et al.* Protein Cages: From Fundamentals to Advanced Applications. *Chem. Rev.* **122**, 9145–9197 (2022).
29. Dundas, C. M., Demonte, D. & Park, S. Streptavidin-biotin technology: improvements and innovations in chemical and biological applications. *Appl. Microbiol. Biotechnol.* **97**, 9343–9353 (2013).
30. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

31. Trinh, C. H., Smith, D. P., Kalverda, A. P., Phillips, S. E. V. & Radford, S. E. Crystal structure of monomeric human β-2-microglobulin reveals clues to its amyloidogenic properties. *Proc. Natl. Acad. Sci.* **99**, 9771–9776 (2002).
32. McParland, V. J. *et al.* Partially Unfolded States of β_2 -Microglobulin and Amyloid Formation in Vitro. *Biochemistry* **39**, 8735–8746 (2000).

EPILOGUE

Conclusions

E1.1 Main Text

In this thesis we describe work related to design and modification of protein cages for several biotechnology applications. While many of the demonstrations represent important developments in the field of protein design, much improvement could be made by developing bottom-up methods for protein cage design; in Nature, protein assemblies evolve from functional proteins, thereby incorporating functions of building blocks into the supramolecular assembly. This contrasts with the work described in this thesis where cages are re-engineered i.e., tailored after the fact, to don functionality. This top-down approach comes with the fairly significant limitation that one cannot alter the cage so much such that the assembly is compromised. The degree to which a cage is engineerable depends greatly on the cage itself, and it is not always evident from the outset of a study if the modifications requisite to an application will be tolerated.

In this thesis work we attempt to work within the confines of the current tools in protein design to develop platforms for applications in biotechnology. Our application-driven work in chapters three-five was entirely based upon cage scaffolds that had been engineered previously. In each of the studies, different cage scaffolds were screened for viability in the intended applications. In chapter three, we noted that an entire class of protein cages: those designed using a technique that involves redesigning oligomeric protein sequences to form symmetric interfaces between them, was not amenable to the demonstrated technique. In chapter four, we noted the limitation that cages used as foundations for the design work needed specific termini requirements that are only found on a select few cages. In our fifth chapter we noted that the nanoparticles we designed were not stable in conditions needed to regenerate them of beta-2

microglobulin binding, making them essentially one-time-use devices and much less cost effective.

To this point much of the work on protein engineering has been based on one of two separate strategies[1]; design of inert assemblies/material, *or*[2] repurposing assemblies for intended applications. We envisage that significant advances in efficacy could come from beginning forays into application-driven protein design from the outset. We are beginning to see the payoff of efforts to do this[1]. In this demonstration, a protein cage was engineered from antibodies as building blocks, rather than inert cyclic oligomers used as building blocks in previous cage design efforts. This work incorporating antibodies into cages contrasts with has been demonstrated by other groups, including some of the work in this thesis, attempting to display binders on protein cages. The result is an assembly with characteristics (including size, symmetry, and occupancy) dictated in advance. Much of this work has been enabled by advances in the tools used to design function protein molecules, particularly by the application of machine learning methods to protein design and structure prediction. It is now routine in protein design labs to screen design candidates by AlphaFold to improve confidence in intended folds. The next step will be to continue with design methods demonstrated by Divine et al. (2021) by designing cages made from other functional moieties such as antigens for vaccine development or enzymes for metabolic engineering. As these methods continue to advance and become accessible to a wider swath of researchers, the avenues for application of protein cages and other protein materials will expand considerably.

References E1.2

1. Divine, R. *et al.* Designed proteins assemble antibodies into modular nanocages. *Science* **372**, eabd9994 (2021).

APPENDIX ONE

STRUCTURAL CHARACTERIZATION OF BACTERIAL PROTEIN OF UNKNOWN FUNCTION USING ALPHA-FOLD ASSISTED MOLECULAR REPLACEMENT

A1.1 Abstract

Macromolecular crystallography requires the recovery of missing phase information from diffraction data to reconstruct an electron density map of the crystallized molecule. Most recent structures have been solved using molecular replacement as a phasing method; this technique relies on the existence of an a priori structure that is closely related to the target protein to serve as a search model. When no such search model exists, molecular replacement is typically not feasible. New advances in computational machine learning methods, however, have resulted in major advances in protein structure predictions from sequence information. Notably, AlphaFold predicts the structure of many proteins to high degrees of accuracy, even those without homologues of known structure, providing a potentially powerful approach to molecular replacement. Taking advantage of these advances, we applied AlphaFold predictions to enable structure determination of a bacterial protein of unknown function (UniprotKB Q63NT7), based on diffraction data that had evaded phasing attempts by MIR and anomalous scattering methods. We were able to solve the structure of the main fragment of the protein – the domain for which AlphaFold predicted a three-dimensional fold with high reliability – using X-ray and micro-electron (microED) diffraction data. This provides an early example of a potentially general path for protein structure determination by diffraction methods.

A1.2 Introduction

New variations on traditional x-ray crystallography are expanding the power of diffraction methods for macromolecular structure determination[1–3]. Two ongoing developments are notable for their potential scope. First, recent algorithmic advances in protein structure prediction have made it possible, in many cases, to generate three-dimensional models that are

accurate enough for molecular replacement protocols[4–6]. Such cases ultimately allow for an experimental structure to be elucidated, without the need for experimental phasing (*i.e.* heavy atom or anomalous approaches), and without prior (experimentally validated) knowledge of a similar protein structure. Second, on the side of experimental advance, electron-based diffraction is attracting attention as a potential approach suitable for very small crystals[7]. These two lines of exploration intersect. Heavy atom and anomalous scattering methods of phasing do not transfer readily to electron diffraction, elevating the importance of molecular replacement for that method, including with predicted models. More case studies are needed to demonstrate the utility, and the challenges, of these new structure determination approaches.

The subject of the present study is a bacterial protein of unknown structure and function, UniprotKB Q63NT7. It was chosen for structural investigation based on its unusual genomic presentation. The tendency of the protein family PF08898 (represented by ProteinID, gene name IPR014994 / NCBI locus tag BPSS0212) to be encoded as repeated paralogs within individual operons suggested that it might form part of a larger self-assembling protein complex[8] (Figure 1a). Biochemical and structural studies were therefore undertaken to investigate that idea. Difficulties in obtaining large crystals led to expanded efforts, including structure determination from small crystals by electron diffraction, and molecular replacement using predicted models. The findings also illuminate interesting cases of crystal formation in the presence of partial

structural disorder.

Figure 1

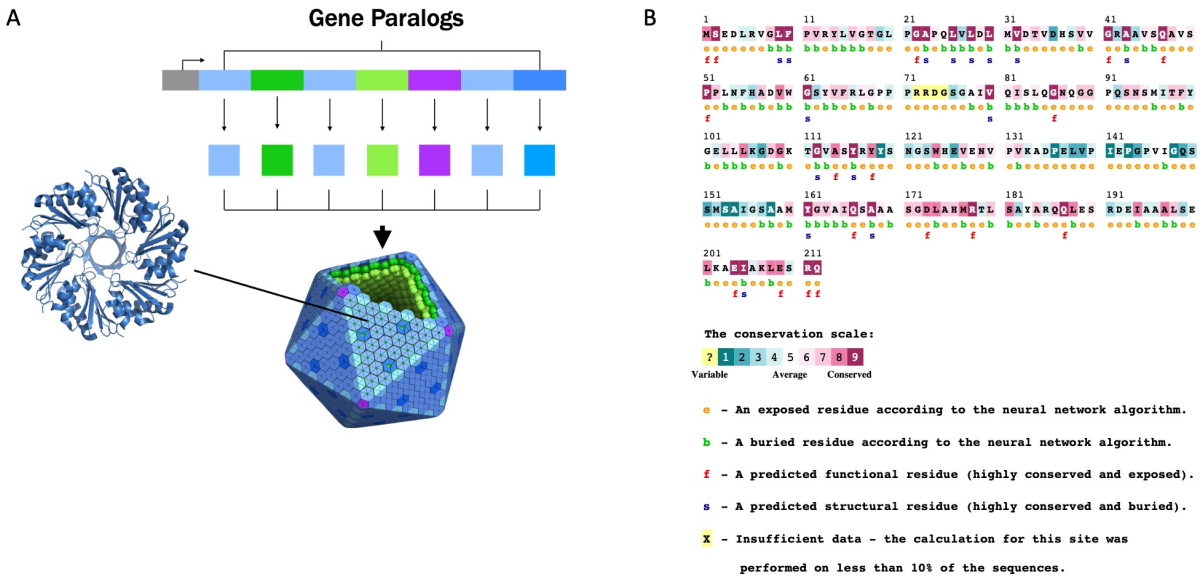


Figure A1.1: Representation of paralogous genes within the operon containing Q63NT7. (A) Graphical representation of the rationale for structural investigations on Q63NT7. Multiple paralogous proteins are shown in shades of blue/green/purple and the resulting assembly of those paralogous proteins into a supramolecular assembly represented by a cartoon depiction of a bacterial microcompartment. (B) Consurf graphical representation of per-residue conservation of Q63NT7[17].

A1.3 Results

Protein Expression and Purification

The Q63NT7 protein from species *Burkholderia pseudomallei* is 212 amino acids long (MW 22.5kDa). It contains two predicted domains: an N-terminal domain of unknown function (14.5kDa) (DUF1842), and the aforementioned C-terminal domain (DUF1843) with paralogs in neighboring proteins (5.4kDa). We ordered sequences encoding the Q63NT7 sequence with C-terminal 6xHistidine tags. We expressed the protein recombinantly in BL21 (DE3) *E. coli* cells

using autoinduction. Biochemical characterization of this protein suggested the protein is monodisperse and likely monomeric in solution (Figure A1.2A).

Protein Crystallization and Crystal Forms

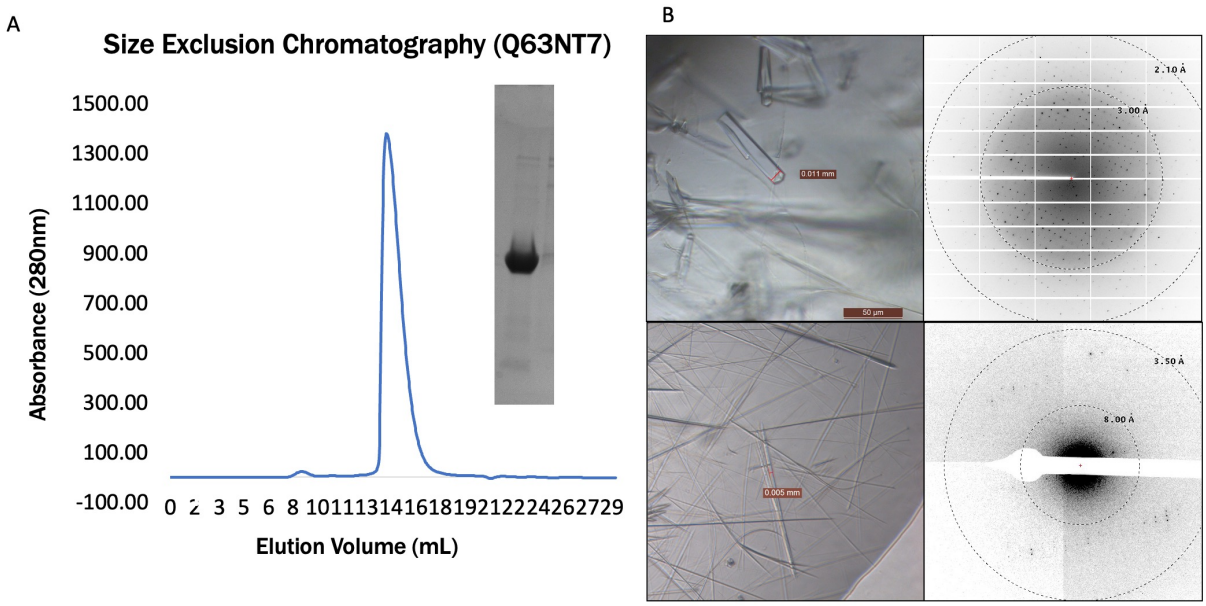


Figure A1.2: Biochemical characterization of Q63NT7: (A) SEC and SDS-Page reveal homogeneity and high purity of purifity Q63Nt7protein. (B) Form 1 (top) and Form 2 (bottom) crystals and representative diffraction data collected from X-ray source or electron microscope respectively.

Encouraged by the purity of our protein sample we attempted to solve the structure crystallographically. Q63NT7 presented a challenge for obtaining large, well-ordered crystals. This led us to explore multiple distinct crystal forms with the goal of improving diffraction quality and, as discussed later, to attempt to visualize a substantial region of the protein that could not be resolved in density maps.

Initial crystallization trials yielded abundant needles across many crystallization conditions, but attempts to obtain x-ray diffraction data were unsuccessful. We also observed inconsistent crystal formation across our replicated crystal trays. Even so, we were ultimately

(after approximate 6-months) able to optimize these conditions and grow larger rectangular shaped crystals that diffracted beyond 3Å on a synchrotron microfocus beamline (Figure A1.2B). We collected datasets on these crystals, which we refer to as form 1. Diffraction data indexing revealed the space group as P21 (Table A1.1). The highest resolution resulted from data collected on a single crystal.

In parallel with efforts to phase form 1 crystals, we continued to optimize crystallization of Q63NT7. Appreciating the possible suitability of needles for micro-electron diffraction Micro-ED methods, we investigated whether our crystal drops contained microcrystals suitable for diffraction in the electron microscope. To investigate this, we pipetted drops that appeared to contain showers of smaller needles onto formvar carbon electron microscopy grids and stained them with uranyl acetate. We observed crystals with visible lattice lines, but the extent of diffraction was somewhat limited (Figure 2). We went on to freeze crystal drops from similar conditions for cryo-EM Micro-ED. We merged datasets from four crystals for processing. We refer to these as form 2 crystals. Unfortunately, the crystals which appeared to be ribbon-shaped at high magnification suffered from preferred orientation problems. We were unable to collect diffraction at high tilt angles, and therefore only achieved 60% completeness in a final diffraction data set. Furthermore, it was difficult to confidently assign a space group due to a substantial missing cone of reflections (Figure A1.3). Noting the poor quality of datasets obtained by merging across multiple crystal specimens, we elected to use a dataset obtained from a single specimen; this did not substantially lower the data completeness, since the regions of reciprocal space missing from distinct data sets were largely overlapping. These datasets indexed

most effectively in space group P212121.

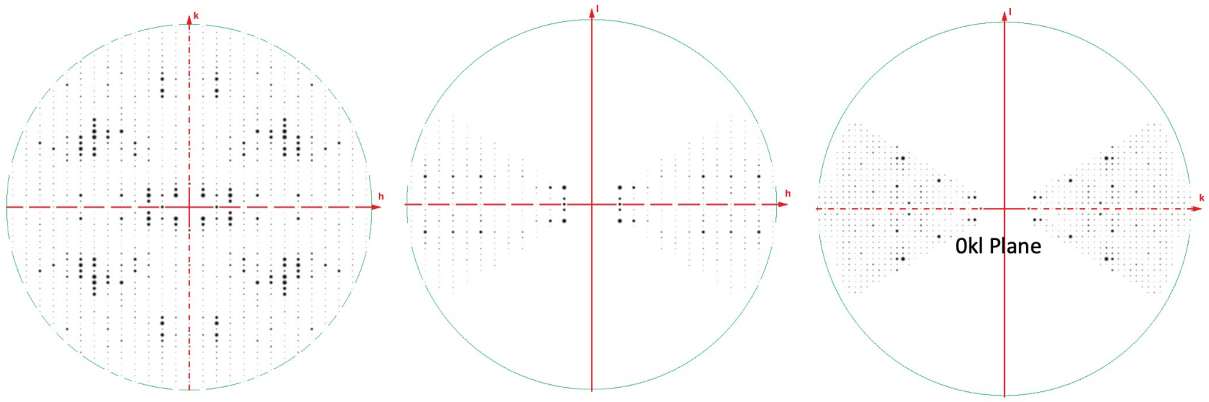


Figure A1.3 slices through reciprocal space show the missing cone present in MicroED data collected on form 2 crystals. Planes in h,k,l are shown to illustrate the missing cone of data due to preferred orientation of crystals on the grid.

Finally, we continued optimization of crystals and identified a condition that grew diffracting needle-form crystals suitable for data collection on the synchrotron micro-focus beamline. We refer to these as form 3 crystals. We were able to collect a complete dataset from a single crystal that indexed in P212121.

Molecular replacement using AlphaFold models

Initial efforts to phase the highest quality dataset (form 1) were futile, with experimental techniques failing; selenomethionine labeled protein crystals did not diffract, and there was no heavy atom signal from crystals soaked in $CsCl_2$ or KI.

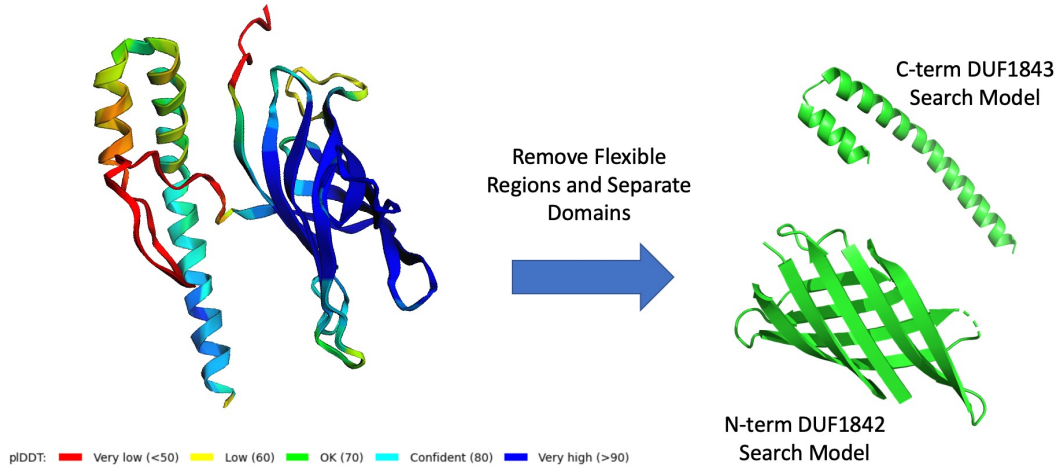


Figure A1.4: AlphaFold model of Q63NT7 used as molecular replacement search model. plDDT gives a per-residue metric of confidence in model prediction

Inspired by studies that had used AlphaFold models to phase datasets with little a priori information, we used the software to generate a model of Q63NT7 (Figure A1.4). AlphaFold identified two domains in the protein, along with a long linker between the two. The N-terminal domain was predicted to form a beta-sandwich fold composed of 8 antiparallel strands. AlphaFold predicted this domain with a high degree of confidence based on per-residue plDDT scores. The C-terminal domain was predicted to form a small helical bundle with modest plDDT confidence metrics (Figure A1.5). Applying existing molecular replacement methods to our

AlphaFold-based molecular replacement efforts, we separated the two domains into separate pdb

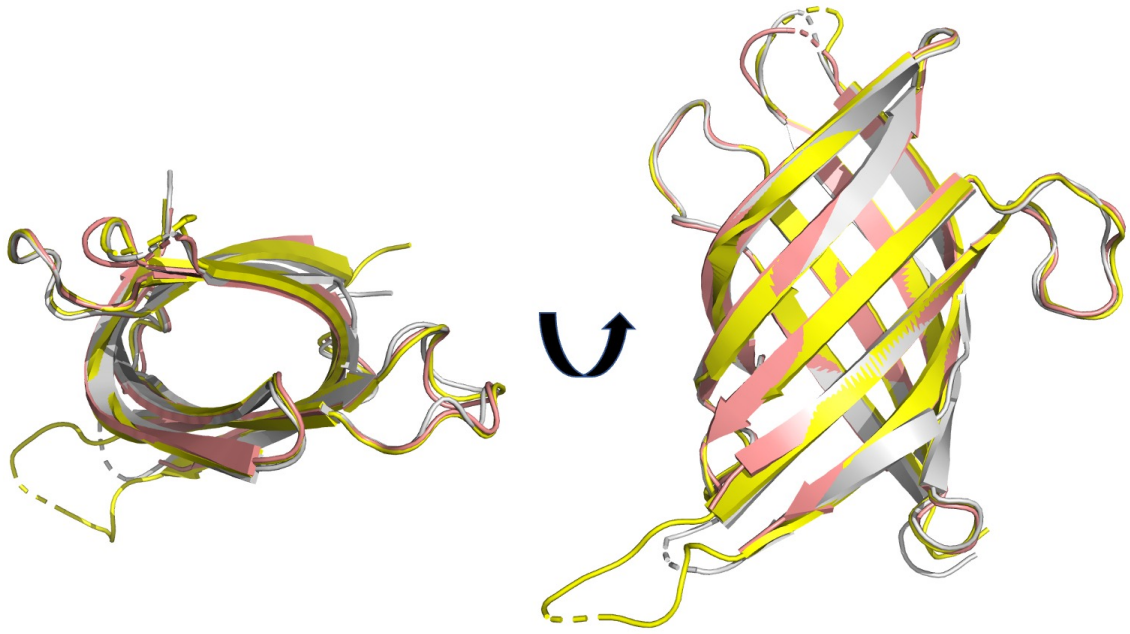


Figure A1.5 Structural comparison of monomers from three crystal forms: Cartoon representation of the structure solved from form 1 crystals (pink), form 2 crystals (yellow), and form 3 crystals (white).

files, and removed extended loops (including the long domain-spanning linker). We used these two files as search models for molecular replacement using the program Phaser[9]. Remarkably, datasets from all three crystal forms gave solutions that passed Phaser's metrics for a correct solution using the N-terminal beta-sheet rich domain, and phasing using models missing residue H125 from the search model revealed a region of positive density in an Fo-Fc omit map that fit the missing residue well (Supplemental Data Figure 1). All three crystals forms identified two copies of the beta barrel domain in the asymmetric unit. Form 1 crystals gave a combined LLG value of 719, form 2 crystals gave an LLG value of 394, and form 3 crystals gave an LLG value of 305. None of the crystals could be phased using the C-terminal alpha helical domain. Given

the small contribution of scattering attributed to this domain because of its small size, this was not altogether surprising.

We next investigated whether similar structures in the PDB existed, and whether they too could have served as search models for molecular replacement with our data. To do this, we submitted the structure obtained from the form 1 crystal dataset (after molecular replacement and preliminary refinement) to the DALI server and identified the top five closest matching protein folds (i.e. those with the highest Z-scores) in the PDB (Supplemental Data Figure 1)[10]. We attempted molecular replacement with these five models, but none of them produced molecular replacement statistics indicative of a correct solution, and attempts to generate electron density maps did not support the models.

Refinement of Atomic Structures

Because form 1 crystals gave the highest resolution diffraction data (from x-rays), a model for the beta domain was refined against that data and then subsequently used as the starting point for model refinements in the other crystal forms (x-ray and electron). This helped especially in the case of the Micro-ED data, which suffered from low completeness and poor I/sigma and Rmerge statistics.

Importantly, Phaser statistics for molecular replacement solutions using the refined form 1 crystal structure were much improved over those from AlphaFold predicted models; form 2 datasets gave a Phaser LLG value of 624 while form 3 crystal datasets gave a phaser LLG value of 662. We used these solutions as starting points for atomic refinements.

During refinement, we paid close attention to the C-terminal region of resulting density maps to observe whether density expected for the C-terminal domain would become visible. In all three forms, large solvent channels are present adjacent to the C-terminus of the beta barrel domain

(Figure 5); Unfortunately, in all crystal forms, we observed no meaningful positive density in Fo-Fc maps in the regions that would have to be occupied by the C-terminal domain (Supplemental Data Figure 3). We hypothesize this could be due to proteolysis, as we observed degradation products on SDS-PAGE gels, and mass spectra from dissolved crystalline samples (Supplemental Data Figure SA1.4 & unpublished data). Final refinement for the form 1 model gave an R factor of 25.1% and R_{free} of 28.7%. The form 2 model had an R factor of 28.5% R_{free} 30.7%. The form 3 model had an R factor of 27.4% and R_{free}: 33.3%. The structure of the N-terminal beta-rich domain was strongly conserved across all crystal forms and asymmetric units; no protein chain from any of the three crystal forms had a backbone RMSD above 0.55 Å compared to any other chain (Figure 5). There was also close agreement between the refined structures and the AlphaFold prediction. Backbone RMSD values between the experimental structures and the AlphaFold model were 0.35Å for form 1, 0.49Å for form 2, and 0.46Å for form 3.

Analyzing the non-crystallographic symmetry of the three crystal forms revealed different molecular packing interfaces across the three crystal forms, indicating that the observed dimeric associations are likely crystallization-induced and not biologically relevant (Figure 6).

A1.4 Discussion and Conclusion

In several cases, microED has proven to be an important tool for structural biologists, enabling the extraction of high-resolution structural information from tiny crystals that are unusable using X-ray diffraction. The earliest demonstration of the method on protein crystals was seminal work on crystals of lysozyme[11]. Important early work from Rodgriguez, Eisenberg and colleagues advanced on these studies and demonstrated the utility of microEDin solving the structures of small peptides[12]. Other work has demonstrated the method's utility in solving

structures of proteins in cases where structures are already known for proteins that are closely or even distantly related, including ligand or drug-bound forms of proteins[13-15]. Nevertheless, experimental methods for phasing MicroED data have been elusive, limiting broader applications of the method. The work presented in this paper adds to the relatively small number of electron diffraction structures of proteins. Notably, as far as we are aware, it represents the first folded (globular) protein structure solved by MicroED whose structure was not already effectively known in advance. We also note that collection of MicroED data was challenged by preferred orientation of the crystals on the EM grid, leading to an incomplete dataset, and to less than ideal statistics. This led to some initial uncertainties in assigning a space group and subsequent structure determination.

We also present the structure of a new small protein fold, and the first from protein family DUF1842. Notably, efforts to obtain structural information on the C-terminal domain from our maps were unsuccessful. We note the presence of a ~25 amino acid long linker predicted to form a loop with low sequence conservation across homologues (Figure 1b)[15]. This could contribute to flexibility of the entire C-terminal region of the protein in the context of the crystal. We also observed several instances of proteolysis in our crystal trays, both with and without the sterilizing agent sodium azide added to the crystal drops (Figure 5). Degradation products appear to be composed of prominent fragments of 4-5kDa and 17-18kDa based on SDS-PAGE (Supplemental Data Figure 2). This could place the cut-site directly N-terminal to the C-terminal domain that did not appear in our crystal structures. The tendency of the protein to undergo proteolysis lends strong support to the hypothesis that the C-terminal region of the protein was absent from all three crystal forms, explaining the absence of detectable density in all cases.

Despite our initial predictions, based on genomic patterns, that Q63NT7 would be involved in oligomerization via its C-terminal domain, we were unable to observe any evidence of higher-order oligomer formation in solution. Our biochemical studies did not support that the protein of unknown function self-assembles into larger architectures under the conditions tested. Nonetheless the appearance of a flexible linker to a terminal domain that was unresolved by crystallography is reminiscent of studies on bacterial microcompartment shell proteins[16] whose genomic patterns were the impetus for the original genomic investigation that identified the IPR014994 domain as a target in the current study.

A1.5 Materials and Methods

Gene Synthesis

Codon-optimized gene sequences were ordered from Integrated DNA Technologies or Twist Biosciences with overlapping sequences corresponding to flanking regions around the hindIII and ndeI restriction sites in the pET-22b expression vector. Intergenic sequences for two-component designs were taken from a pETDuet-1 expression plasmid and ordered as a single gene fragment.

Protein Expression and Purification

Designs were cloned into pET-22b expression vectors using Gibson assembly. Correct cloning of gene was verified using Sanger sequencing. Small-scale expression was performed in BL21(DE3) cells grown in 200 mL of cultures using auto-induction media grown for 24 h at 25 C. Cells were lysed in 50mm Tris-HCl pH 8.0, 250mm NaCl supplemented with 5 mM 2-mecaptoethanol and EDTA-free protease inhibitor tablets (Thermo Fisher Scientific) using an Emulsiflex C3 homogenizer and affinity purified using Ni-NTA agarose resin (Thermo Fisher Scientific) in a gravity flow column. Protein was washed with lysis buffer +100 mM imidazole

and eluted in lysis buffer +500 mM imidazole. Eluted protein was dialyzed against imidazole overnight at 4 °C. Samples were run on SDS-PAGE to purity before SEC using a Superdex-75 column (Cytiva Life Sciences) attached to an Acta FPLC (Cytiva Life Sciences). Sodium Azide was then added to SEC elution fractions at a concentration of 0.05% as well as EDTA at a concentration of 5mM.

Crystallization

96 well crystal screens were set up using a TTP Labtech Mosquito in hanging-drop vapor-diffusion format. Trays were allowed to incubate at 22C until crystal were observed.

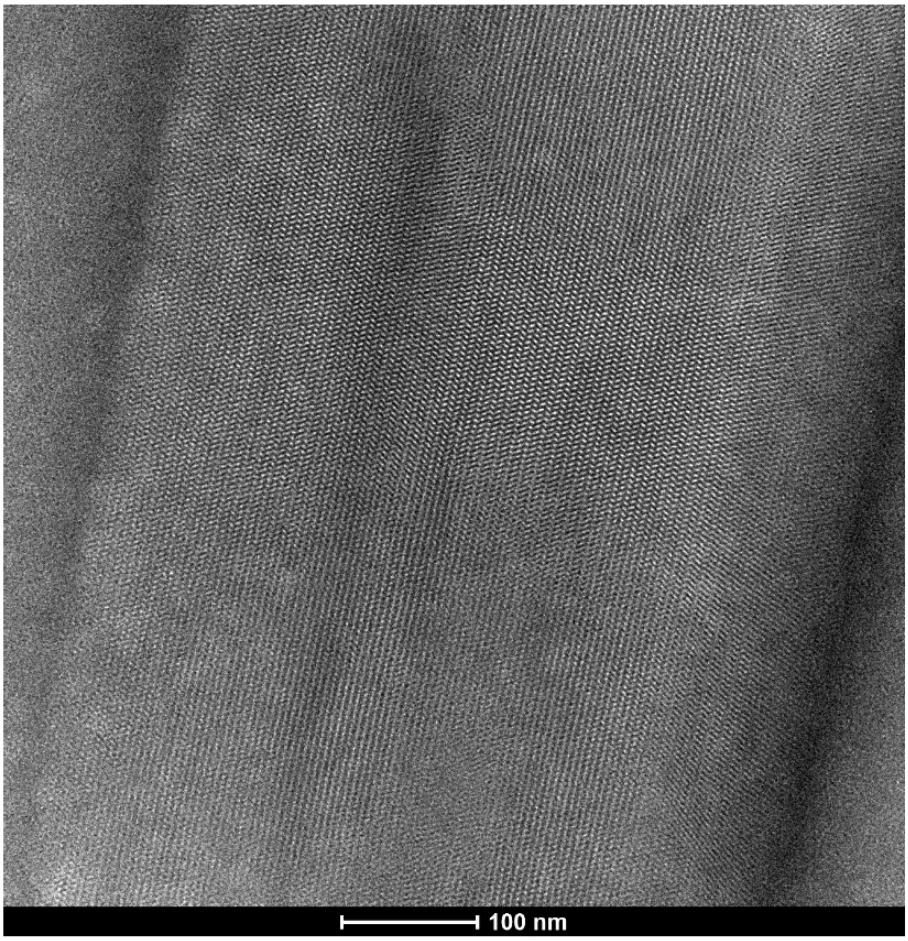
Data Collection and Processing

X-ray diffraction datasets were collected at the Advanced Photo Source on beamlines NE-CAT 24-ID-C equipped with an EIGER 16M detector and 24-ID-E equipped with Dectris PILATUS 6M-F detector. MicroED data collection was performed on a ThermoFisher Tecnai TF30 Microscope equipped with a TVIPPS camera. The XDS software package was used to index diffraction data.

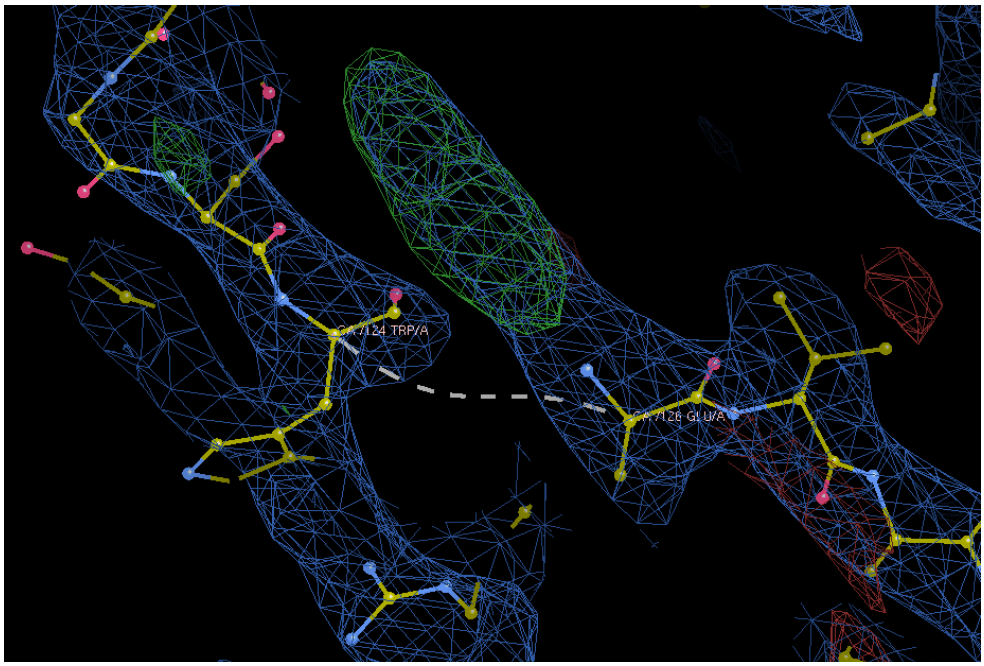
Molecular Replacement and Structure Refinement

The Phaser program was used for molecular replacement. The AlphaFold program was used to generate molecular replacement search models. After refining this initial AlphaFold search model on the basis of form 1 diffraction data, we used the refined structure to phase form 2 and form 3 crystals, as it gave the best statistics and resulted in the best electron density maps. The Coot program was used for model building, and refinement was performed using Phenix.

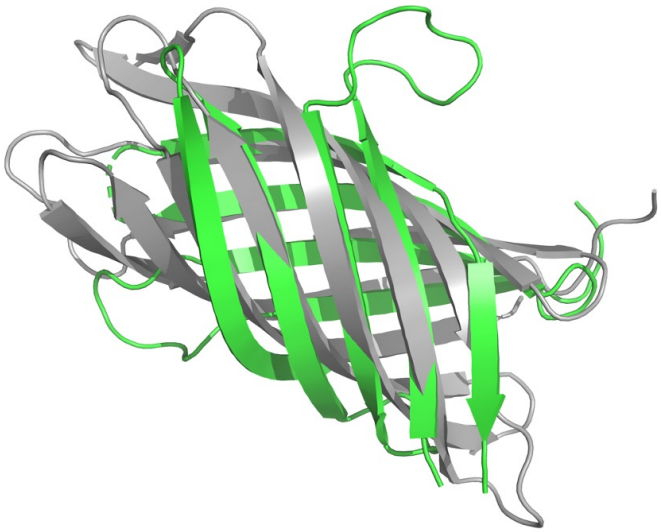
A1.6 Supplemental Data



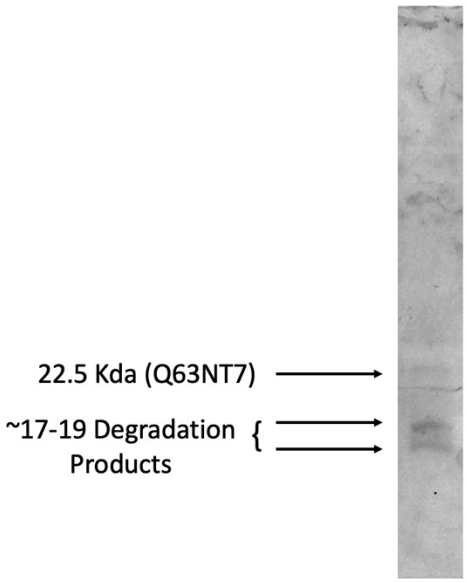
Supplemental Data figure A1.1: Negatively stained Q63NT7 crystal visualized on a T12 electron microscope.



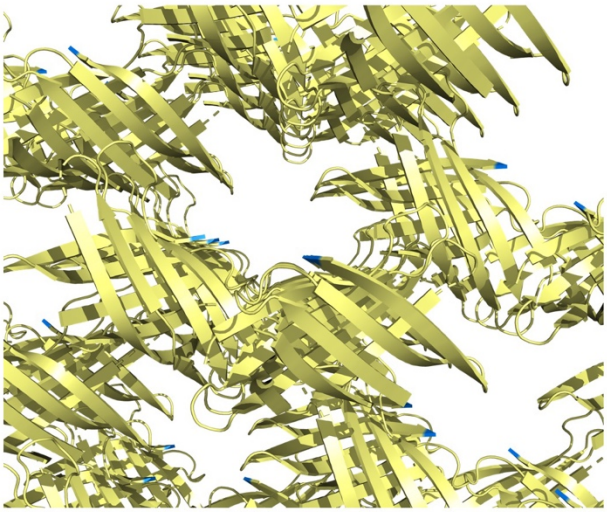
Supplemental Data Figure A1.2: Micro-ED omit map confirms the correct molecular replacement solution when using the AlphaFold search model on form 2 diffraction data. Histidine 125 was deleted from the search model, and density appears for this residue in an F_o-F_c map calculated using model phases. Molecular replacement search model shown as atomic model, green density corresponds to positive density in F_o-F_c map.



Supplemental Data Figure A1.3: Comparison of the closest identifiable homolog of known structure (PDB 1prt) with the experimental structure of protein Q63NT7. The 1prt structure is shown as a cartoon in grey overlayed with the form 1 crystal structure of Q63NT7 (green).



Supplemental Data Figure A1.4: SDS-PAGE analysis of Form 3 crystals reveals prominent degradation products for the Q63NT7 protein.



Supplemental Data Figure A1.5: Crystal packing for the form 1 crystal of the Q63NT7 protein reveals solvent channels at the C-terminus of the beta-barrel rich domain.

Source Organism	Burkholderia Pseudomallei
DNA Source	Synthetic
Expression vector	Pet 22b (+)
Plasmid Construction method	Gibson assembly
Expression host	Escherichia coli (BL21 (DE3))
Expression details	Autoinduction media
Amino Acid Sequence: MSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPLNFHA DVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKGDGKTGVAS YRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAAMYGVAIQSAAASGD LAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQHHHHHH	

Supplemental Data Table SA1.1: Macromolecule Production

Method	Hanging drop
Plate type	96 well
Temperature (°C)	20
Protein Concentration	20 mg/ml
Buffer composition of protein	100mM BisTris pH 5.5, 25% PEG 3350
Volume and ratio of drop	2:1
Drop setting	SPT LabTech Mosquito
Seeding	No

Supplemental Data Table SA1.1: Crystallization Form 1 Crystals

Method	Hanging drop
Plate type	96 well
Temperature (°C)	20
Protein Concentration	20 mg/ml
Buffer composition of protein	100mM BisTris pH 5.5, 100mM Ammonium Acetate 17% PEG 10,000
Volume and ratio of drop	2:1
Drop setting	SPT LabTech Mosquito
Seeding	No

Supplemental Data Table SA1.2: Crystallization Form 2 Crystals

Method	Hanging drop
Plate type	96 well
Temperature (°C)	20
Protein Concentration	100 mg/ml
Buffer composition of protein	100mM TRIS HCl pH 8.5, 150mM MgCl, 12.5% PEG 8000
Volume and ratio of drop	1:1
Drop setting	SPT LabTech Mosquito
Seeding	No

Supplemental Data Table SA1.3: Crystallization Form 3 Crystals

Crystal Form	Form 1	Form 2	Form 3
PDB code	TBD	TBD	TBD
Resolution	APS 24-ID-C	Technai TF30	APS 24-ID-E
No. of Reflections	1.45860	0.01969	0.97918
Rwork/Rfree	100	100	100
B-factors	199.516	5280	398.246
RMSD Bond Lengths (Å)	0.5	0.85	0.5
RMSD Bond Angles	180	65	70
Ramachandran Outliers	P21	P212121	P212121
	39.47,40.43,78.49	40.59,95,101.46	40.11,70.82,94.5
α, β, γ (°)	90,97.013,90	90,90,90	90,90,90
Completeness (%)	89.1 (82.9)	60.9(50.9)	87.9 (92.2)
$\langle I/\sigma(I)\rangle$	8.85 (1.43)	6.34(2.71)	4.38 (2.33)
$R_{\text{r.i.m.}}$	0.059 (0.701)	0.329 (0.539)	0.272 (.627)

Supplemental Data Table SA1.4: Data Collection and Processing

Crystal Form	Form 1	Form 2	Form 3
PDB code	TBD	TBD	TBD
Resolution Range (Å)	77.90-2.10	35.03-3.02	47.25-3.00
Number of unique reflections	13938	5070	4823
Ramachandran Outliers	0	0	0
Ramachandran Allowed	9	10	7
Ramachandran Favored	222	224	220
Rwork	25.11	28.45	27.35
Rfree	28.68	30.66	33.3

Supplemental Data Table SA1.5: Refinement Statistics

A1.7 References

1. Thompson, M. C., Yeates, T. O. & Rodriguez, J. A. Advances in methods for atomic resolution macromolecular structure determination. *F1000Res* **9**, 667 (2020).
2. Johansson, L. C., Stauch, B., Ishchenko, A. & Cherezov, V. A Bright Future for Serial Femtosecond Crystallography with XFELs. *Trends in Biochemical Sciences* **42**, 749–762 (2017).
3. Martynowycz, M. W. & Gonen, T. From electron crystallography of 2D crystals to MicroED of 3D crystals. *Current Opinion in Colloid & Interface Science* **34**, 9–16 (2018).
4. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
5. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
6. Giorgetti, A., Raimondo, D., Miele, A. E. & Tramontano, A. Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics* **21**, ii72–ii76 (2005).
7. Nannenga, B. L. & Gonen, T. MicroED: a versatile cryoEM method for structure determination. *Emerging Topics in Life Sciences* **2**, 1–8 (2018).
8. Beeby, M., Bobik, T. A. & Yeates, T. O. Exploiting genomic patterns to discover new supramolecular protein assemblies. *Protein Science* NA-NA (2008) doi:10.1002/pro.1.
9. McCoy, A. J. *et al. Phaser* crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
10. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res* **44**, W351–W355 (2016).
11. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 (2013).

12. Rodriguez, J. A. *et al.* Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
13. Martynowycz, M. W. & Gonen, T. Ligand Incorporation into Protein Microcrystals for MicroED by On-Grid Soaking. *Structure* **29**, 88-95.e2 (2021).
14. Martynowycz, M. W. *et al.* MicroED structure of the human adenosine receptor determined from a single nanocrystal in LCP. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106041118 (2021).
15. Xu, H. *et al*. Solving a new R2lox protein structure by micro-crystal electron diffraction. *Science Advances* **5,** eaax4621 (2019).
16. Thompson MC, Yeates TO. A challenging interpretation of a hexagonally layered protein structure. *Acta Crystallogr D Biol Crystallogr*. 2014;70(Pt 1):203-208.
17. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344–W350 (2016).

APPENDIX TWO

ANALYSIS OF PROTEIN INTERFACES TO DETERMINE QUATERNARY STRUCTURAL EFFECTS ON PROTEIN FRAGMENTATION IN NATIVE MASS SPECTROMETRY

A2.1 Preamble

In this work, I contributed to a study investigating the effects of protein-protein interfaces on mass spectrometry-induced protein fragmentation. I provided computational analaysis of protein-protein interfaces to allow for stratification of protein fragments by proximity to interfacial regions.

A2.2 Abstract

Native mass spectrometry (MS) of proteins and protein assemblies reveals size and binding stoichiometry, but elucidating structures to understand their function is more challenging. Native top-down MS (nTDMS), i.e., fragmentation of the gas-phase protein, is conventionally used to derive sequence information, locate post-translational modifications (PTMs), and pinpoint ligand binding sites. nTDMS also endeavors to dissociate covalent bonds in a conformation-sensitive manner, such that information about higher-order structure can be inferred from the fragmentation pattern. However, the activation/dissociation method used can greatly affect the resulting information on protein higher-order structure. Methods such as electron capture/transfer dissociation (ECD and ETD, or ExD) and ultraviolet photodissociation (UVPD) can produce product ions that are sensitive to structural features of protein complexes. For multi-subunit complexes, a long-held belief is that collisionally activated dissociation (CAD) induces unfolding and release of a subunit, and thus is not useful for higher-order structure characterization. Here we show not only that sequence information can be obtained directly from CAD of native protein complexes but that the fragmentation pattern can deliver higher-order structural information about their gas- and solution-phase structures. Moreover, CAD-generated internal fragments (i.e., fragments containing neither N-/C-termini) reveal structural aspects of protein complexes.

A2.3 Results and Discussion

Native top-down mass spectrometry (nTDMS) of gas-phase proteins yields product ions that can provide information on amino acid sequence[1,2], sites of modifications[3–5], and even higher-order structure[6]. Performing nTDMS with electron-based techniques such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) [7–11] and photon-based techniques such as infrared multiphoton dissociation (IRMPD) and ultraviolet photodissociation (UVPD) [8,12–14] is generally favored, as it fragments the complex directly without disrupting the overall complex structure. In contrast, it has been generally assumed that collision-based fragmentation does not reveal higher-order structural information, as unfolding and ejection of monomer subunits (and ligands) occurs. However, we have found that direct fragmentation of native protein complexes with Orbitrap-based high-energy C-trap dissociation (HCD)[15], a collision-based fragmentation technique performed with higher energy on a faster time scale than conventional collisionally activated dissociation (CAD), can uncover aspects of protein higher-order structure. For a variety of protein complexes, we show here that HCD can generate b-/y-type product ions that provide information on solvent-exposed regions and subunit interfaces.

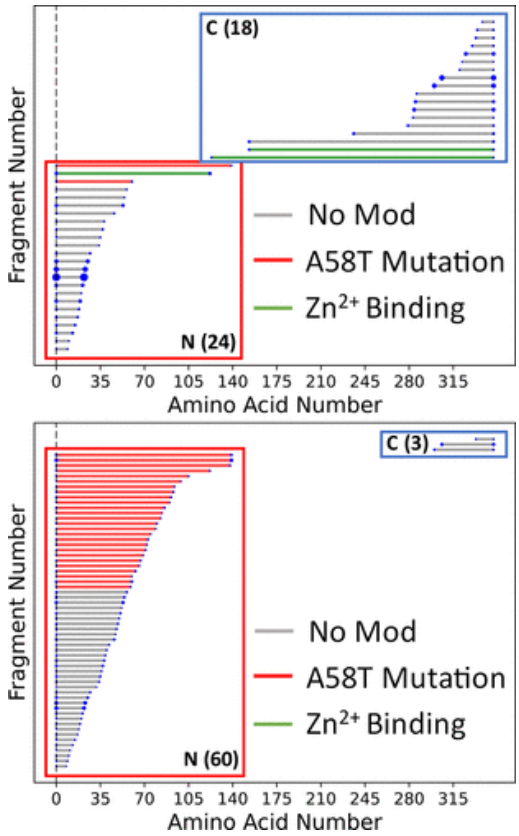
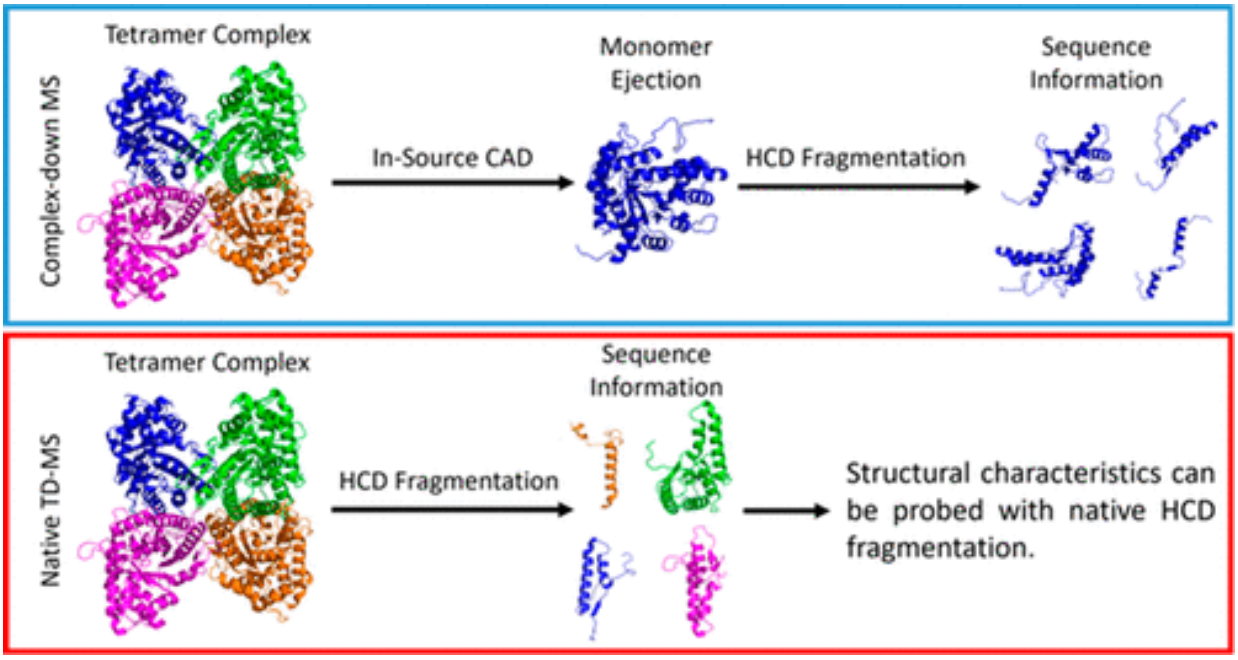


Figure A2.1. Fragment location maps for ADH representing *b*-/*y*-product ions measured by (top) complex-down MS and (bottom) nTDMS with HCD. Red lines indicate V58T mutation, green lines indicate Zn^{2+} binding, the vertical dotted line indicates N-terminal acetylation, and the size of the blue dots indicates the relative intensity of each fragment. Numbers in parentheses indicate the number of product ions detected.

To investigate HCD fragmentation of protein complexes, (16) complex-down MS (pseudo-MS3) [17,18] and nTDMS (Scheme A2.1) of yeast alcohol dehydrogenase (ADH) homotetramer (147 kDa) were compared. Complex-down MS was performed by using in-source CAD to detach a monomer from the tetramer and to subsequently activate the 12+ charged monomer with HCD. The resultant MS/MS spectrum revealed both N-terminal b-fragments and C-terminal y-fragments of ADH (Figure A2.S1A); 24 b-fragments and 18 y-fragments resulted in 11.8% total sequence coverage (Figure A2.1A). The fragmentation pattern also revealed the presence of N-terminal acetylation, a V58T proteoform, and Zn2+ binding. The presence of near

equal numbers of abundant b- and y-fragments from the complex-down MS workflow suggests that both termini of the ADH monomer subunit are easily accessed by HCD fragmentation, i.e., the in-source CAD process releases a low-structured monomer such that subsequent HCD products yield little information about the 3D structure of the native tetramer.



Scheme A2.1 Complex-Down MS and nTDMS Workflows Used in This Study

For comparison, nTDMS results from HCD of the 25+ charged ADH tetramer were examined. Primarily b-products and surprisingly few peaks corresponding to released ADH monomers (Figure A2.1B, Figure A2.S1B) were detected. We speculate that monomers were not ejected from the tetramer complex prior to covalent bond cleavage, i.e., the tetramer fragmented directly. To further support this claim, broadband fragmentation (of all ADH tetramer charge states) with a range of HCD energies did not yield significant levels of released monomer signals (Figure A2.S2). nTDMS of ADH yielded 60 N-terminal b-fragments, but only three C-terminal y-fragments (17.6% sequence coverage) (Figure A2.1B). Numerous abundant N-terminal fragments produced by HCD resemble nTDMS products from electron-based (7,8) and

photodissociation techniques[8,12]. Mapping the fragments onto the crystal structure of ADH shows that the N-terminal region is more solvent exposed than the C-terminal region, with the latter forming subunit–subunit interfaces of the complex (Figure A2.S3). Our analysis indicates that fragments that cut at the interface of the tetramer (residues 240–310) accounted for only 8% of the fragment ion current.

To further examine how collision-based fragmentation can reveal structural information from protein complexes, intact (rabbit) aldolase homotetramer (157 kDa) was fragmented with HCD. Much like ADH, aldolase did not release monomers upon HCD, but rather y-fragments including an especially abundant y74 ion (2+ to 5+ charged) (Figure A2.S4). At low HCD energies, a large complementary fragment corresponding to the mass of the intact tetramer losing a y74-fragment, i.e., (4M – y74), was observed (Figure A2.S5 and Table A2.S1), indicating direct fragmentation of the tetramer. nTDMS yielded 35 C-terminal y-fragments but only eight N-terminal b-fragments (11.0% sequence coverage) (Figure A2.2). This result differs from the complex-down mass spectrum of aldolase, which shows a nearly equal proportion of N-terminal b-fragments[19] and C-terminal y-fragments[16] (Figure A2.S6).

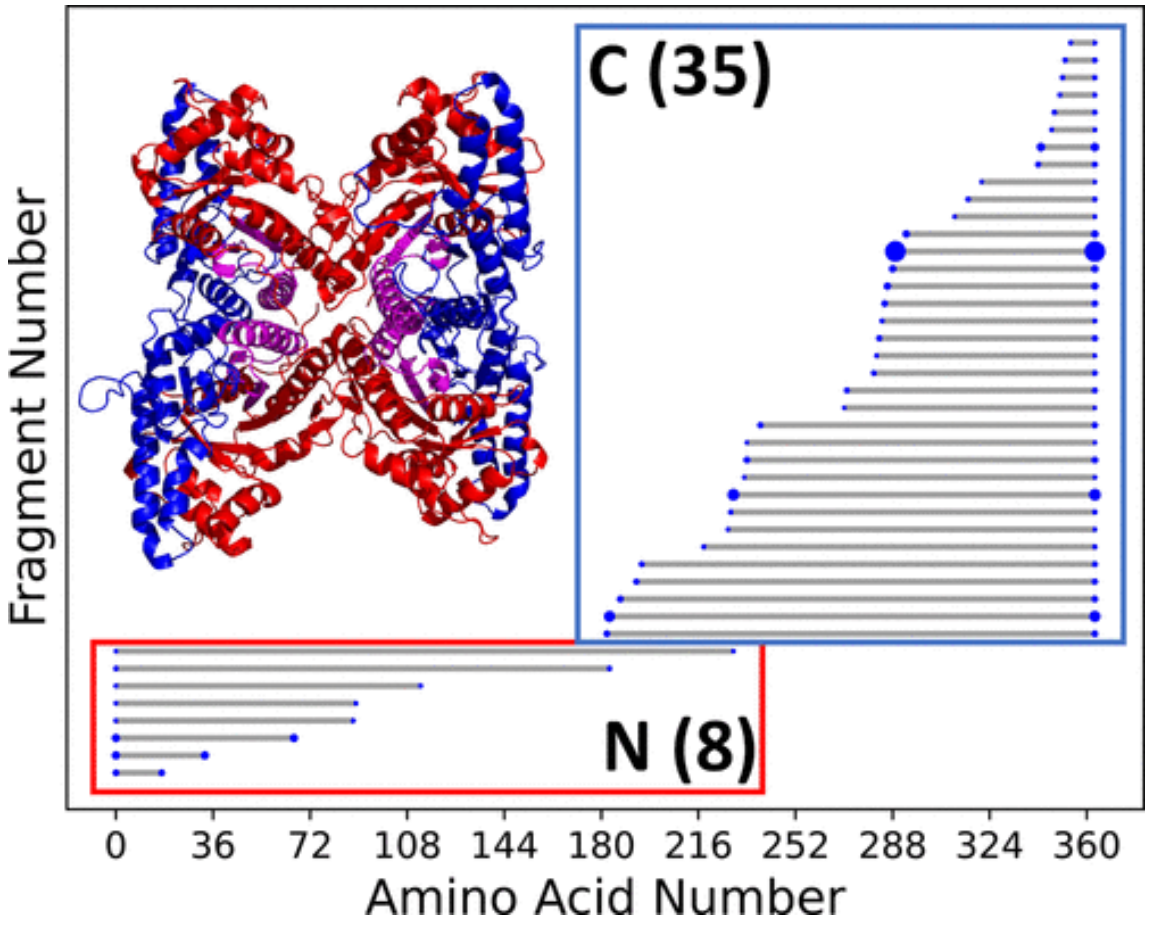


Figure A2.2. Fragment location map for nTDMS products of the 25+ charged precursor of aldolase homotetramer, with the size of the blue dots corresponding to the relative intensity of the fragments The crystal structure shows that most cleavage sites lie on the solvent-exposed C-terminus (blue), rather than the interface forming N-terminus (red). The purple region is covered by both N-terminal and C-terminal fragments.

The HCD fragments from the aldolase tetramer mainly cover the solvent-exposed C-terminus and are absent from the interface forming N-terminus (Figure A2.2). Our analysis indicates that fragments that cut at the interface of the tetramer (residues 110–224) accounted for only 1% of the fragment ion current. The relatively high proportion of C-terminal fragments present in the native HCD spectrum of aldolase is similar to that measured by ECD previously[19] and further suggests that direct HCD fragmentation of some protein complexes can reveal regions of solvent accessibility.

nTDMS with HCD was performed on several other protein complexes. Complex-down fragmentation of the glutathione S-transferase A1 (GSTA1) dimer revealed 25 N-terminal b-fragments and 20 C-terminal y-fragments (Figure A2.S7A). In contrast, the native fragmentation spectrum of GSTA1 reveals five N-terminal b-fragments and 19 C-terminal y-fragments (Figure A2.S7B), consistent with the GSTA1 crystal structure showing that the C-terminus is more solvent exposed than the N-terminus (Figure A2.S7B). For the yeast enolase dimer, 27 b-fragments along with 18 y-fragments were measured by complex-down MS (Figure A2.S8A). nTDMS revealed 48 N-terminal b-fragments along with 51 C-terminal y-fragments without the appearance of abundant monomer ions (Figure A2.S8B). The crystal structure of enolase (Figure A2.S8B) indicates that both N-/C-termini are solvent exposed and are not involved in forming the dimer interface, consistent with the near equal proportion of b-/y-products measured by nTDMS.

Some complexes did not release monomers from in-source CAD for complex-down fragmentation; however, HCD of the native complexes still returned structural information. Native HCD of the creatine kinase dimer revealed nine b- and 38 y-fragments, which suggests that the C-terminus is solvent exposed and the N-terminus forms the interface of the dimer; this aligns well with the crystal structure of creatine kinase (Figure A2.S9). Similarly, HCD of 6-phosphogluconate dehydrogenase (GND1) dimer generated 23 b-fragments but only six y-fragments, consistent with the GND1 crystal structure showing the N-termini to be solvent exposed and the C-termini forming the dimer interface (Figure A2.S10).

There are some exceptions to this pattern of b-/y-product formation directly from intact native complexes under HCD. For example, HCD of the native membrane protein, aquaporin Z (AqpZ) homotetramer[20–22], yielded abundant monomer, dimer, and trimer products released from

the intact complex (Figure A2.S11A). This observation can be attributed to the weak hydrophobic binding interface between the monomer subunits of the AqpZ tetramer. Complexes such as aldolase and ADH are stabilized somewhat by salt bridges that strengthen greatly in the gas phase[23,24], potentially preventing monomer ejection during HCD (Figure A2.S12). That monomer products are released when HCD is applied to native AqpZ complexes suggests that structural information (such as the locations of solvent-exposed regions and the tetramer interface) cannot be inferred from the resulting b-/y-fragments, at least assuming that the monomers likely eject before covalent bonds cleave. This suggestion is supported by the fact that the nTDMS fragmentation pattern of AqpZ tetramers (65 b-fragments, 62 y-fragments, 38.4% sequence coverage; Figure A2.S11A) does not differ significantly from the complex-down fragmentation pattern of isolated monomers (63 b-fragments, 60 C-terminal y-fragments, 34.6% sequence coverage; Figure A2.S11B). Although HCD fragmentation of native AqpZ does not reveal significant higher-order structural information, it does suggest that the interaction between complex monomers in the gas phase is relatively weak.

Monomer releases during HCD are not limited to membrane protein complexes. HCD fragmentation of the hemoglobin (Hb) tetramer revealed monomer and trimer peaks in addition to 10 b-fragments and eight y-fragments from the α-subunit and seven b-fragments and seven y-fragments from the β-subunit (Figure A2.S13A). Fragmentation of the Hb dimer also revealed released monomer peaks in addition to 11 b-fragments and 16 y-fragments from the α-subunit and 10 b-fragments and four y-fragments from the β-subunit (Figure A2.S13B). A similar HCD fragmentation pattern can be observed from complex-down MS of individual subunits (nine b- and 10 y-fragments and six b- and five y-fragments from the α- and β-subunits, respectively) (Figure A2.S13C). Similarly, nTDMS of human transthyretin (TTR) tetramers by HCD releases

monomer products in addition to two b- and 38 y-fragments (Figure A2.S14A). The relative proportion of b-/y-product ions between the tetramer and monomer TTR is similar, with complex-down of the TTR monomer yielding three b-fragments and 41 y-fragments (Figure A2.S14B). The HCD results for all of the complexes included in the study are listed in Table A2.S2.

Lastly, we investigated the utility of internal fragments (i.e., product ions containing neither N-/C-termini that result from at least two bond cleavage events)[1,25–30] for structure determination of protein complexes. Preliminary data show that HCD fragmentation of ADH tetramers reveals numerous internal fragments spanning residues 178–236 (Figure A2.S15A), which correspond to a solvent-exposed region (Figure A2.S15B). More work will extend this concept further, but it demonstrates that HCD-derived internal fragments can deliver structural information on protein assemblies.

Although other studies have noted the detection of b-/y-products with concurrent subunit release from CAD[31] and HCD[16] of protein complexes, we have found that collision-based fragmentation with HCD can reveal higher-order structure information for several multi-subunit protein complexes that appear to be stabilized through the presence of salt bridges[23]. These complexes fragment directly by HCD without significant monomer release. The resulting products map to solvent-exposed areas, while regions delivering fewer fragments likely comprise subunit interfaces. Other weak gas-phase complexes eject monomers upon HCD. Nonetheless, it is currently unclear what differences between HCD and other beam-type CAD experiments are responsible for the unique fragmentation behavior.

An assumption carried over from small-molecule dissociation studies to macroion decompositions is that, on the experimental time scale, activation from collisions always

randomizes fully to steer collision-induced decompositions along the lowest energy pathways. However, those assumptions fail to consider that entropically demanding, slow rearrangements might be essential to releasing a subunit, e.g., to reposition salt bridges tethering one subunit to others[23]. In cases where the number of collisions and/or energy per collision are insufficient to stumble on the rare configuration ejecting a subunit within the experimental time frame, alternative rearrangements to eject smaller polypeptide fragments (with fewer tethers) may be competitive. Nevertheless, we show that HCD can be a powerful biophysical tool to probe the structure of proteins without the need for other electron- and photon-based activation/dissociation methods.

A2.4 Supplemental Materials

Supporting Information

Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order Structure Information for Protein Complexes

Carter Lantz[1], Benqian Wei[1], Boyu Zhao[1], Wonhyeuk Jung[1], Andrew K. Goring[1], Jessie Le[1], Justin Miller,[4] Rachel R. Ogorzalek Loo[1,3,4], Joseph A. Loo[1,2,3,4]*

[1]Department of Chemistry and Biochemistry, University of California-Los Angeles, Los Angeles, CA, 90095; [2]Department of Biological Chemistry, University of California-Los Angeles, Los Angeles, CA, 90095; [3]UCLA-DOE Institute, University of California-Los Angeles, Los Angeles, CA, 90095; and [4]Molecular Biology Institute, University of California-Los Angeles, Los Angeles, CA, 90095

Materials and Methods:

Commerically available protein samples were obtained from Sigma Aldrich (St. Louis, MO, USA), and dissolved in 200mM ammonium acetate, and desalted with 10K Amicon filters from Sigma Aldrich. Aquaporin Z (AqpZ; from Pascal Egea, UCLA)[1] and human hemoglobin (Hb; from Robert Clubb, UCLA)[2] were isolated and prepared as described previously. The samples were then diluted to 10μM and sprayed on a Thermo UHMR (Thermo Fisher Scientific, San Jose, CA) with voltages of 1-2kV. To fragment native complexes, HCD energies of 125-280V were applied. Lower voltages were applied for select applications. Complex-down MS experiments were performed by applying 5-150V of in-source CAD or -60V of desolvation voltage to eject monomers and then applying 100V-177V of HCD energy to subsequently activate those monomers. For internal fragment analysis, the ADH the tetramer was fragmented with 215V of collision energy with argon as the collision gas.

Deconvolution was performed with BioPharma Finder 3.2 and the resulting deconvoluted peak list was run through ClipsMS.[3] *b*- and *y*-fragments were matched to protein sequences with an error tolerance of 5ppm and unlocalized modifications including the addition of a hydrogen atom and the abstraction of a water molecule were added to the theoretical masses. For ADH, additional modifications including an N-terminal acetylation, a V58T mutation, and a Zn^{2+} ion were added to theoretical fragments. *by* internal fragments of ADH were searched with ClipsMS with an error tolerance of 5ppm. To deconvolute large complementary fragments, UniDec was used.[4] Fragments were mapped onto crystal structures of protein complexes using Pymol 2.5.4. The ADH pymol code used was 4W6Z, the aldolase pymol code used was 1ADO, the enolase pymol code used was 1EGB, the GSTA1 pymol code used was 1GSD, and the creatine kinase pymol code used was 1U6R, and the aquaporin Z pymol code used was 1RC2.

References:

1. Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G., Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 183-193.

2. Spirig, T.; Malmirchegini, G. R.; Zhang, J.; Robson, S.; Sjodt, M.; Liu, M.; Kumar, K. K.; Dickson, C. F.; Gell, D A.; Lei, B.; Loo, J. A.; Clubb, R. T., *Staphylococcus aureus* uses a novel multidomain receptor to break apart human hemoglobin and steal its heme. *J. Biol. Chem.* **2013**, *288*, 1065-1078.

3. Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A., ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021**, *20* (4), 1928-1935.

4. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V., Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **2015**, *87* (8), 4370-4376.

Supplementary Figures:

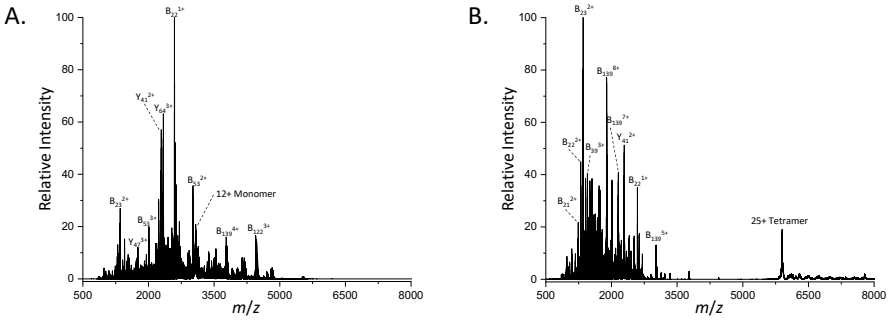


Figure A.S1: (A) Complex-down fragmentation mass spectrum and (B) native TDMS spectrum of ADH.

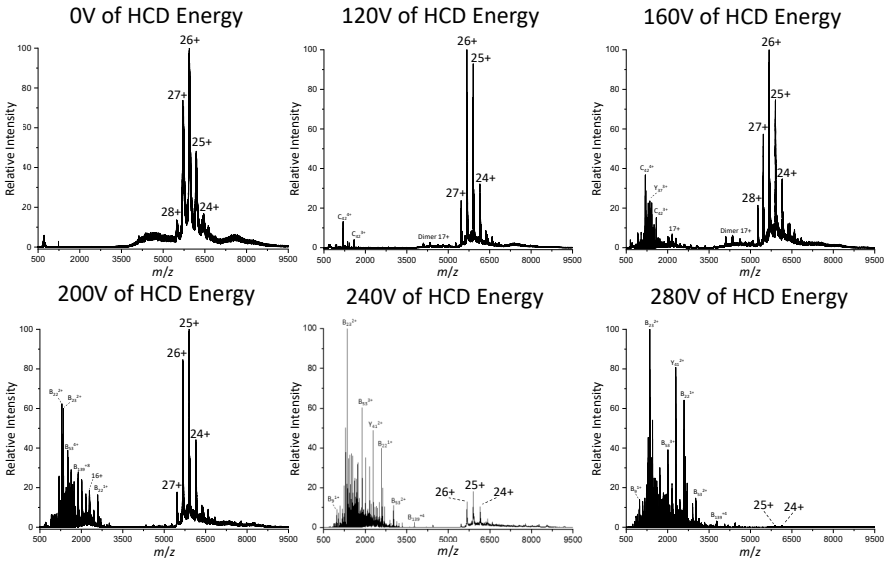


Figure A2.S2: Broadband nTDMS spectra of ADH at various HCD energies.

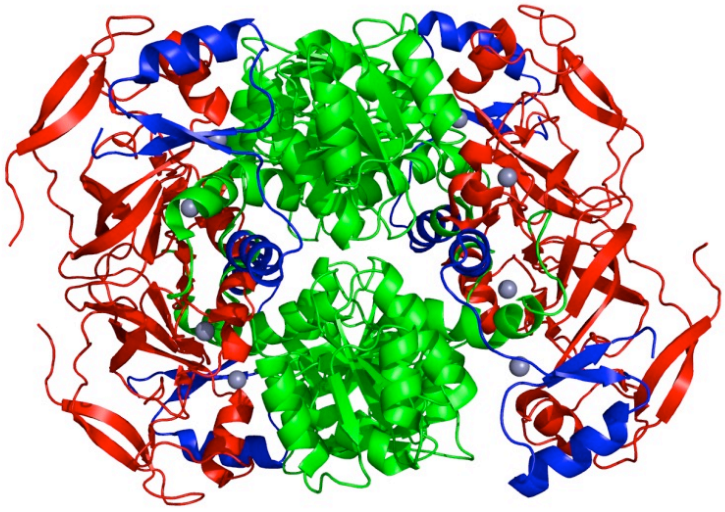


Figure A2.S3: The structure of ADH with the region covered by the N-terminal fragments labeled in red and the region covered by the C-terminal fragments labeled in blue. Fragmentation occurs in the solvent exposed regions and does not occur in the subunit-subunit interface region (green).

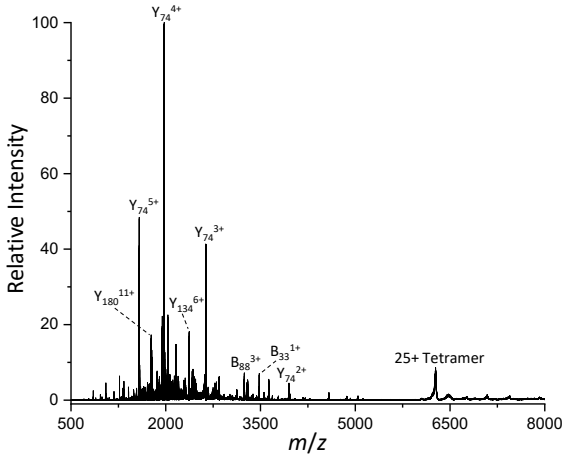


Figure A.S4: Native top-down mass spectrum of the 25+ charged aldolase homotetramer.

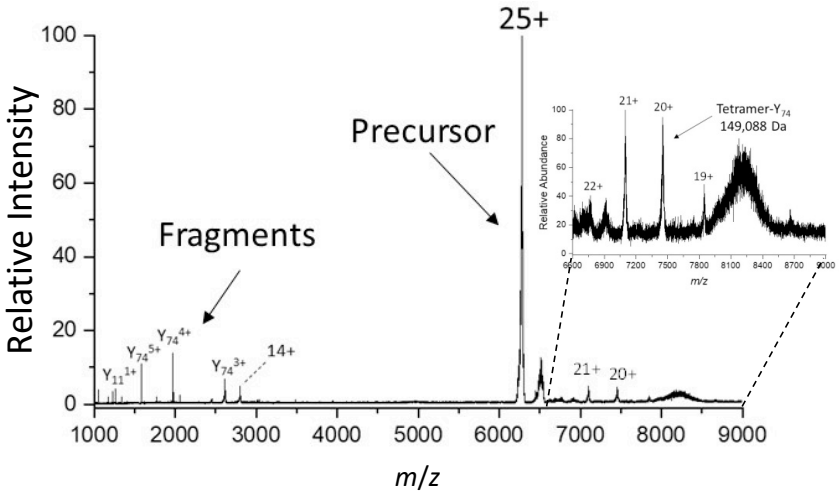


Figure A2.S5: Native top-down mass spectrum of the 25+ charge state of the aldolase homotetramer showing multiple charge states of an abundant y_{74} fragment and high *m/z* peaks corresponding to charge states of the (4M - y_{74}) product ion.

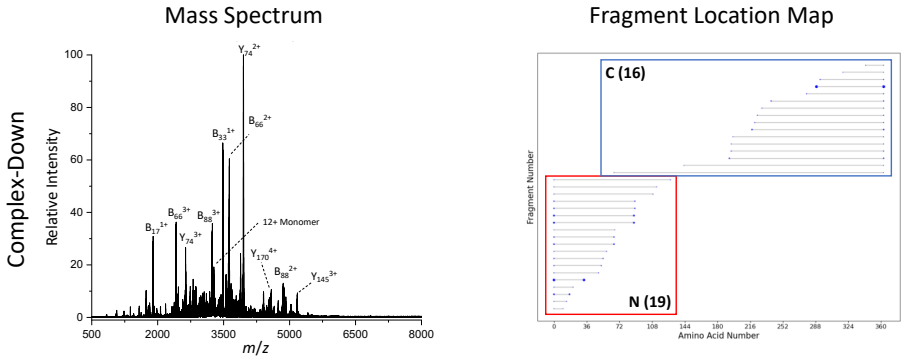


Figure A2.S6: A complex-down mass spectrum (12+ monomer) and the corresponding fragment location map for aldolase.

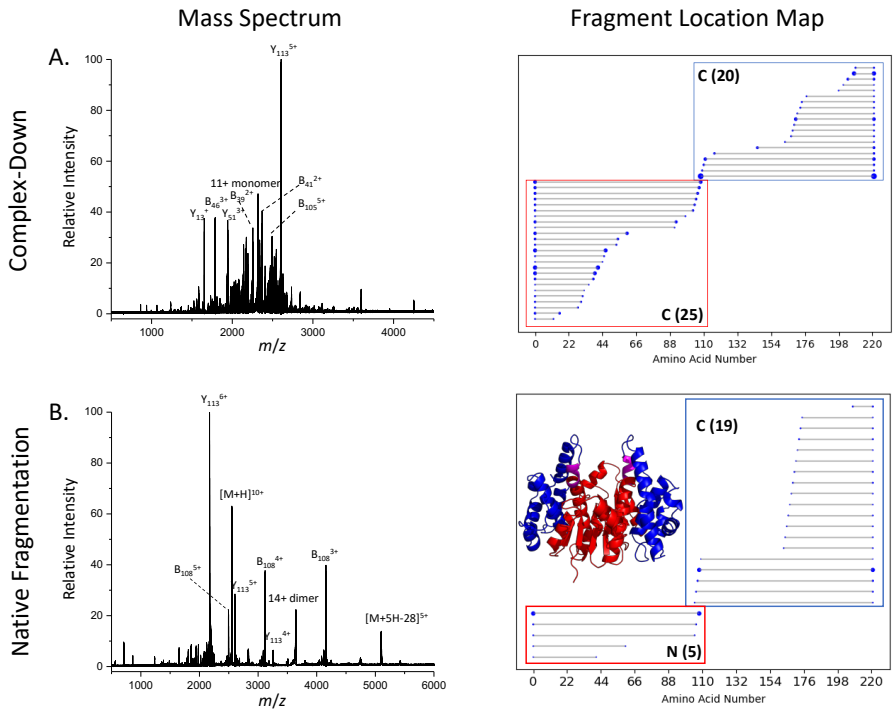


Figure A2.S7: (A) A complex-down mass spectrum (11+ monomer) with the corresponding fragment location map and (B) a native top-down mass spectrum with the corresponding fragment location map for the 12+ charged human GSTA1 dimer. The inset shows the structure of GSTA1 with the region covered by N-terminal fragments labeled in red, the region covered by C-terminal fragments labeled in blue, and the region covered by N- and C-terminal fragments labeled in purple.

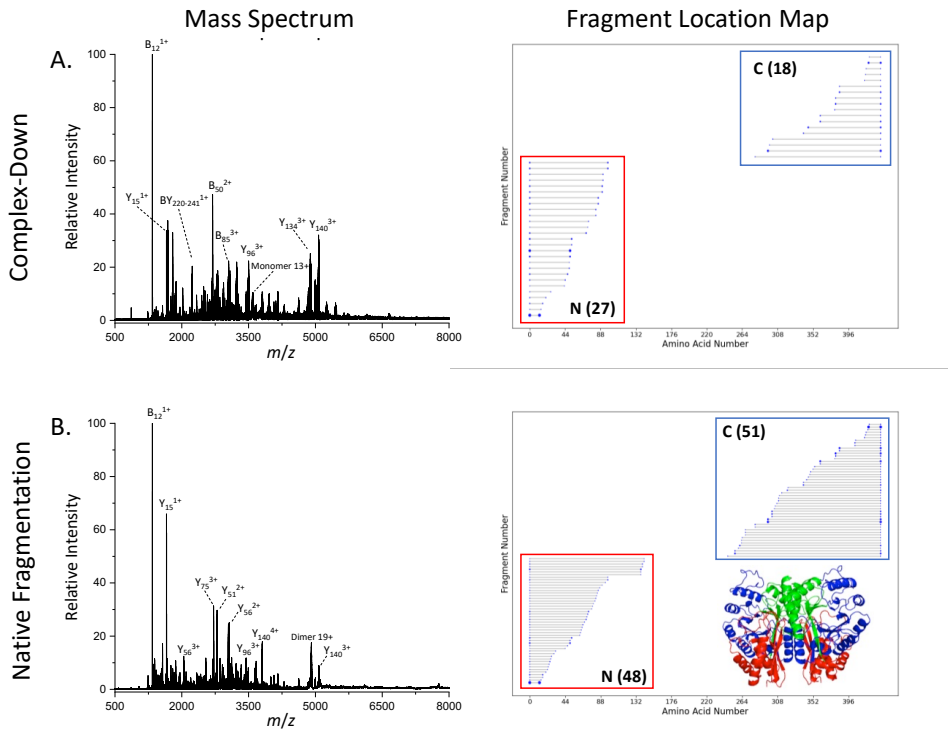


Figure A2.S8: (A) A complex-down mass spectrum (13+ monomer) with the corresponding fragment location map and (B) a native top-down mass spectrum with the corresponding fragment location map for the 19+ charged enolase dimer. The inset shows the structure of enolase with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

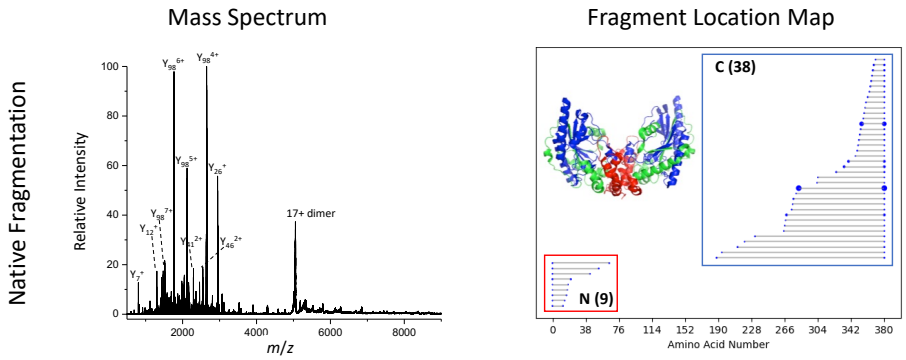


Figure A2.S9: A native top-down mass spectrum with the corresponding fragment location map for 17+ charged creatine kinase dimer. The inset shows the structure of creatine kinase with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

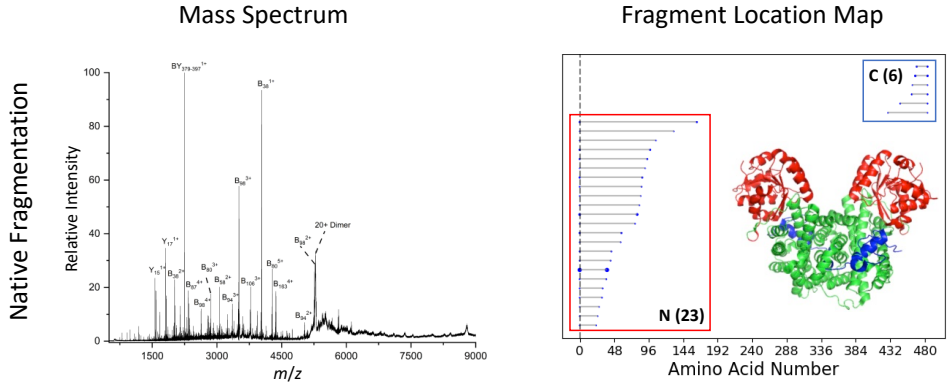


Figure A2.S10: A native top-down mass spectrum with the corresponding fragment location map for the 20+ charged GND1 dimer with the vertical dotted line representing N-terminal acetylation. The inset shows the structure of GND1 with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

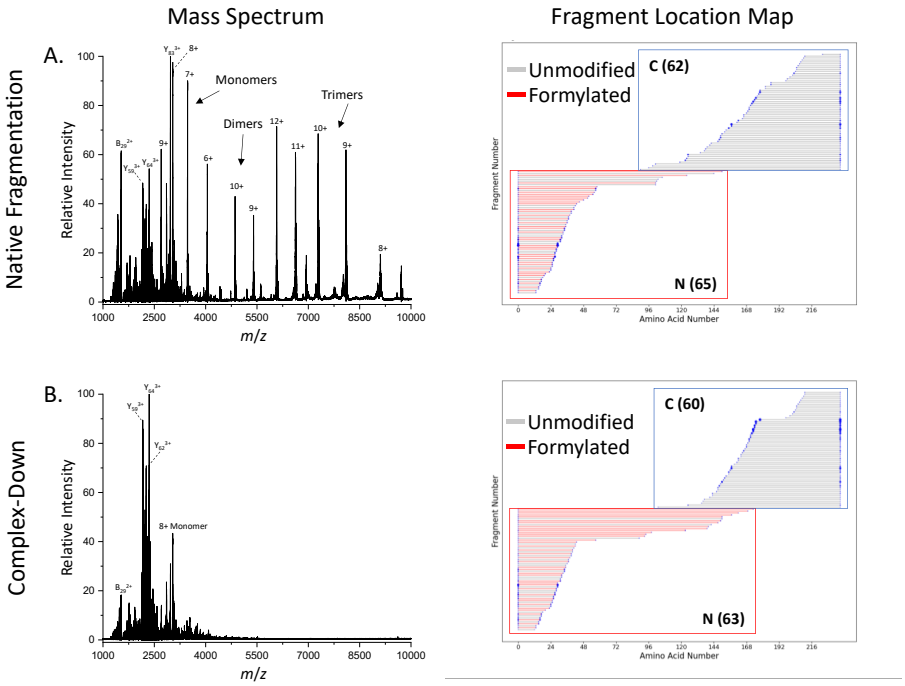


Figure A2.S11: (A) A native top-down mass spectrum (18+ charged tetramer) with the corresponding fragment location map and (B) a complex-down mass spectrum (8+ monomer) with the corresponding fragment location map for AqpZ.

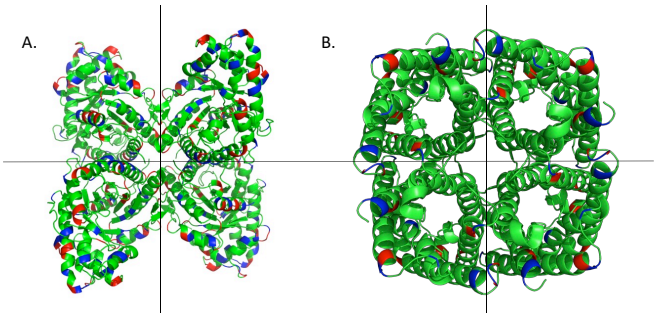


Figure A2.S12: (A) The crystal structures of rabbit aldolase and (B) aquaporin Z with positively charged amino acids (Lys and Arg) labeled in blue and negatively charged amino acids (Glu and Asp) labeled in red. The black lines indicate the complex interface. Aldolase contains many charged residues at the interface of the protein complex and aquaporin Z does not, which may explain why aquaporin Z releases monomers and aldolase does not when HCD is applied to the intact complex.

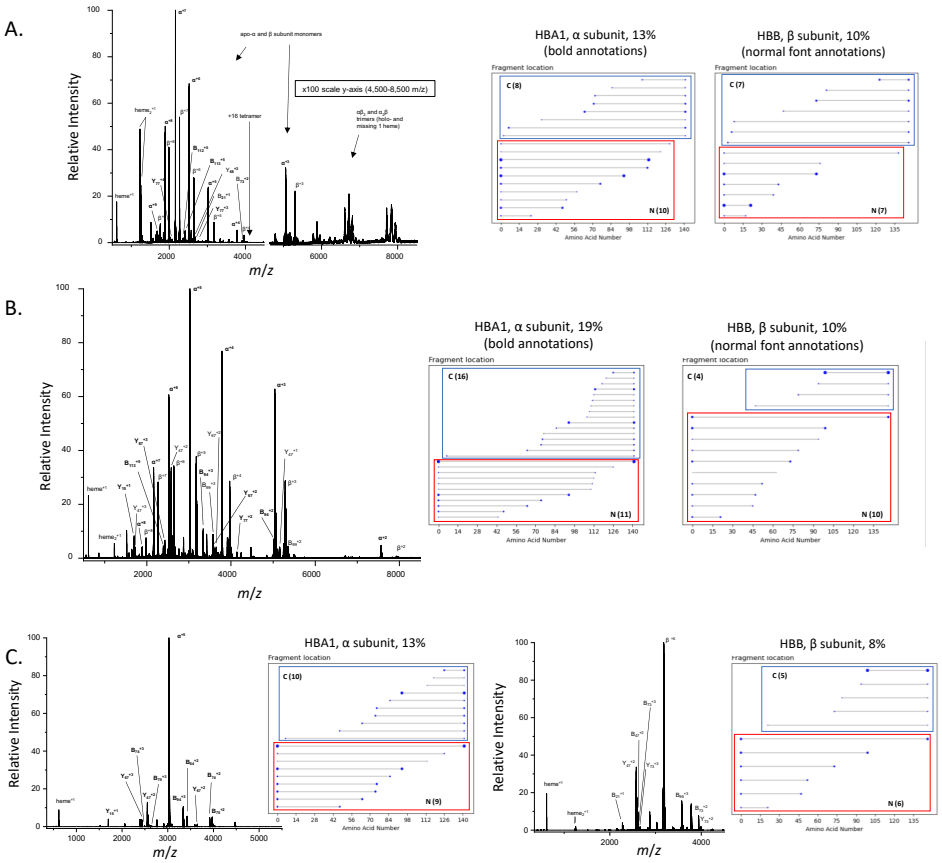


Figure A 2 . S13: (A) nTDMS spectrum of the 16+ charged hemoglobin tetramer with the corresponding fragment location maps for the α-subunit and β-subunit, (B) a nTDMS mass spectrum of the 12+ charged hemoglobin dimer with the corresponding fragment location maps for the α-subunit and β-subunit, and (C) complex-down fragmentation mass spectra and the corresponding fragment location maps for the 6+ charged α-subunit and β-subunit.

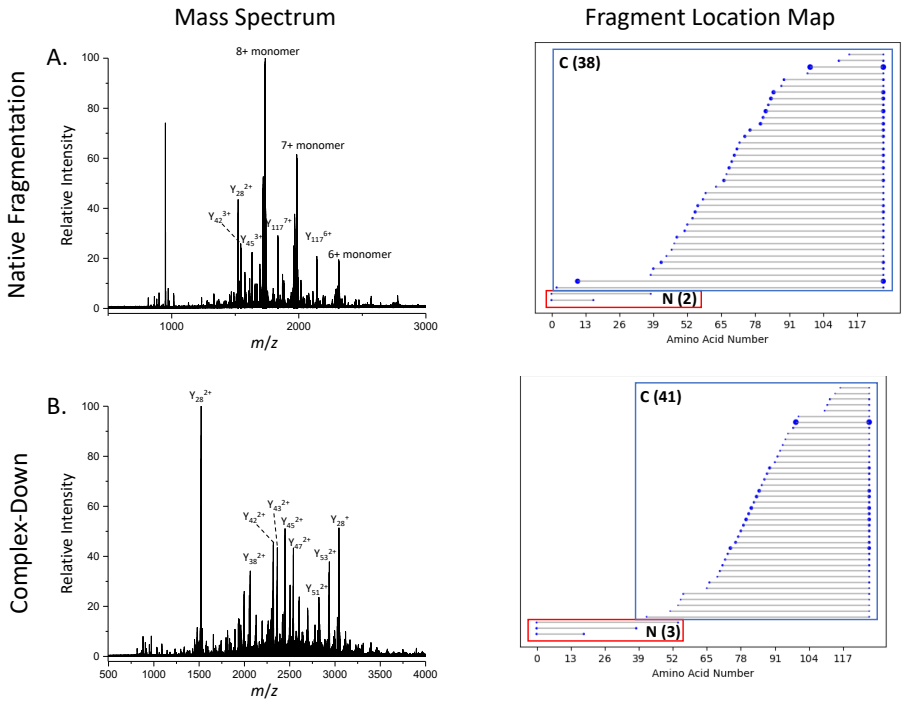


Figure A2.S14: (A) Native top-down mass spectrum (15+ tetramer) with the corresponding fragment location map and (B) a complex-down mass spectrum (6+ monomer) with the corresponding fragment location map for the TTR tetramer.

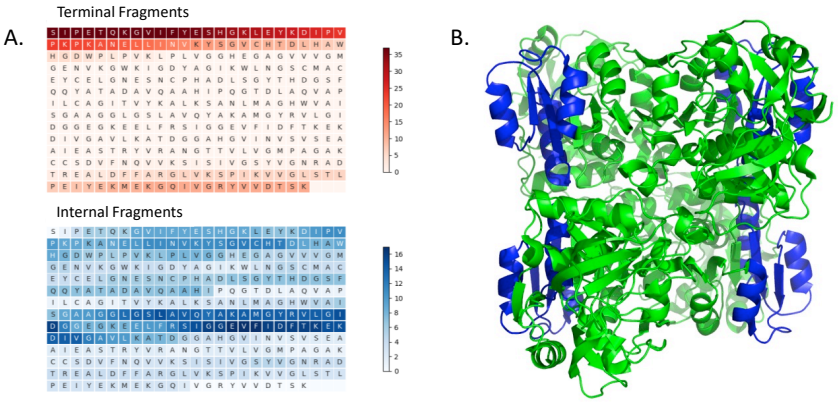


Figure A2.S15: (A) A heatmap representing terminal and internal fragment analysis of the 24+ charged ADH tetramer and (B) the structure of the ADH tetramer with an internal fragment hotspot (residues 178-236) highlighted in blue. Notice how this region of the ADH tetramer is solvent exposed.

Supplementary Tables:

Table S1: Molecular weights of species present in a low HCD energy spectrum of aldolase. The high m/z ions in the spectrum correspond to the aldolase tetramer-y74.

Species	MW (Da)
Measured Molecular weight of Aldolase	156,982
Measured Molecular Weight Higher m/z Peaks	149,088
Molecular Weight Difference	7,894
Theoretical Mass of Y74	7,896

Table S2: Information on the complexes analyzed in this study

General Inforamtion				Native TD-MS Fragmentation				Complex-Down Fragmentation			
Complex	Complex Type	MW Complex (Da)	MW Monomer (Da)	Monomer Release w/ HCD	# N-Term Frags	# C-Term Frags	Sequence Coveage	Monomer Release w/ CAD	# N-Term Frags	# C-Term Frags	Sequence Coverage
ADH	Tetramer	147,472	36,738	No	60	3	18%	Yes	24	18	12%
Aldolase	Tetramer	156,748	39,187	No	8	35	11%	Yes	19	16	8%
GSTA1	Dimer	51,000	25,500	No	5	19	11%	Yes	25	20	21%
Enolase	Dimer	93,312	46,656	No	48	51	18%	Yes	27	18	8%
Creatine Kinase	Dimer	86,224	43,112	No	9	38	13%	No	NA	NA	NA
GND1	Dimer	106,003	52,957	No	23	6	6%	No	NA	NA	NA
AqpZ	Tetramer	97,076	24,269	Yes	65	62	38%	Yes	63	60	35%
Hemoglobin	Tetramer	61,986	α=15,126 β=15,867	Yes	α=10 β=7	α=8 β=7	α=13% β=10%	Yes	α=9 β=6	α=10 β=5	α=13% β=8%
Hemoglobin	Dimer	30,993	α=15,126 β=15,867	Yes	α=11 β=10	α=16 β=4	α=19% β=10%	Refer to row above	Refer to row above	Refer to row above	Refer to row above
Transthyretin	Tetramer	55,044	13,761	Yes	2	38	33%	Yes	3	41	36%

A2.5 References

(1) Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* 2021, *20* (4), 1928−1935.

(2) Vimer, S.; Ben-Nissan, G.; Morgenstern, D.; Kumar-Deshmukh, F.; Polkinghorn, C.; Quintyn, R. S.; Vasil'ev, Y. V.; Beckman, J. S.; Elad, N.; Wysocki, V. H. Comparative structural analysis of 20S proteasome ortholog protein complexes by native mass spectrometry. *ACS Cent. Sci.* 2020, *6* (4), 573−588.

(3) Nshanian, M.; Lantz, C.; Wongkongkathep, P.; Schrader, T.; Klärner, F.-G.; Blümke, A.; Despres, C.; Ehrmann, M.; Smet-Nocca, C.; Bitan, G. Native top-down mass spectrometry and ion mobility spectrometry of the interaction of tau protein with a molecular tweezer assembly modulator. *J. Am. Soc. Mass Spectrom.* 2019, *30* (1), 16−23.

(4) Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A. Top-down ESI-ECD-FT- ICR mass spectrometry localizes noncovalent protein-ligand binding sites. *J. Am. Chem. Soc.* 2006, *128* (45), 14432−14433.

(5) O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S. Characterization of native protein complexes using ultraviolet photodissociation mass spectrometry. *J. Am. Chem. Soc.* 2014, *136* (37), 12920−12928.

(6)Zhou,M.;Lantz,C.;Brown,K.A.;Ge,Y.;Pasă-Tolic,́L.;Loo,J. A.; Lermyte, F. Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. *Chem. Sci.* 2020, *11* (48), 12918−12936.

(7) Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native electrospray and electron-capture dissociation in FTICR mass spectrometry provide top-down sequencing

of a protein component in an intact protein assembly. *J. Am. Soc. Mass Spectrom.* 2010, *21* (12), 1966−1968.

(8) Li, H.; Wongkongkathep, P.; Van Orden, S. L.; Ogorzalek Loo, R. R.; Loo, J. A. Revealing ligand binding sites and quantifying subunit variants of noncovalent protein complexes in a single native top-down FTICR MS experiment. *J. Am. Soc. Mass Spectrom.* 2014, *25* (12), 2060−2068.

(9) Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D.; Loo, J. A. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nature Chem.* 2018, *10* (2), 139−148.

(10) Lermyte, F.; Sobott, F. Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes. *Proteomics* 2015, *15* (16), 2813−2822.

(11) Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F. Top-down characterization of denatured proteins and native protein complexes using electron capture dissociation implemented within a modified ion mobility-Mass spectrometer. *Anal. Chem.* 2020, *92* (5), 3674−3681.

(12) Zhou, M.; Liu, W.; Shaw, J. B. Charge Movement and Structural Changes in the Gas-Phase Unfolding of Multimeric Protein Complexes Captured by Native Top-Down Mass Spectrometry. *Anal. Chem.* 2020, *92* (2), 1788−1795.

(13) Cammarata, M. B.; Thyer, R.; Rosenberg, J.; Ellington, A.; Brodbelt, J. S. Structural characterization of dihydrofolate reductase complexes by top-down ultraviolet photodissociation mass spectrom- etry. *J. Am. Chem. Soc.* 2015, *137* (28), 9128−9135.

(14) Cammarata, M. B.; Brodbelt, J. S. Structural characterization of holo-and apo-myoglobin in the gas phase by ultraviolet photo- dissociation mass spectrometry. *Chem. Sci.* 2015, *6* (2), 1324−1333.

(15) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* 2007, *4*, 709−712.

(16) Hale, O. J.; Cooper, H. J. Native Mass Spectrometry Imaging of Proteins and Protein Complexes by Nano-DESI. *Anal. Chem.* 2021, *93*, 4619−4627.

(17) Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A. Top or middle? Up or down? A standard lexicon for protein top-down and allied mass spectrometry approaches. *J. Am. Soc. Mass Spectrom.* 2019, *30*, 1149−1157.

(18) Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Vale, L. H. F. D.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentic, K.; Kelleher, N. L.; Compton, P. D. Top-down character- ization of endogenous protein complexes with native proteomics. *Nature Chem. Biol.* 2018, *14*, 36−41.

(19) Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A. Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 2014, *86* (1), 317−20.

(20) Campuzano, I. D. G.; Nshanian, M.; Spahr, C.; Lantz, C.; Netirojjanakul, C.; Li, H.; Wongkongkathep, P.; Wolff, J. J.; Loo, J. A. High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: From Inorganic Salt Clusters to Antibody Conjugates and Beyond. *J. Am. Soc. Mass Spectrom.* 2020, *31*, 1155−1162.

(21) Laganowsky, A.; Reading, E.; Hopper, J. T.; Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nature Protoc.* 2013, *8*, 639−651.

(22) Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* 2018, *29*, 183−193.

(23) Ogorzalek Loo, R. R.; Loo, J. A. Salt Bridge Rearrangement (SaBRe) Explains the Dissociation Behavior of Noncovalent Complexes. *J. Am. Soc. Mass Spectrom.* 2016, *27*, 975−990.

(24) Yin, S.; Loo, J. A. Elucidating the Site of Protein-ATP Binding by Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2010, *21*, 899−907.

(25) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. Analyzing Internal Fragmentation of Electrosprayed Ubiquitin Ions During Beam-Type Collisional Dissociation. *J. Am. Soc. Mass Spectrom.* 2015, *26*, 782−787.

(26) Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* 2021, *93*, 6355−6362.

(27) Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A. Towards Understanding the Formation of Internal Fragments Generated by Collisionally Activated Dissociation for Top-Down Mass Spectrometry. *Anal. Chim. Acta* 2022, *1194*, 339400.

(28) Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Ogorzalek Loo, R. R.; Loo, J. A. Internal fragments generated by electron ionization dissociation enhance protein top-down mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2020, *31*, 1896−1902.

(29) Zenaidee, M. A.; Wei, B.; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A. Internal Fragments Generated from Different Top-

Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* 2021, *32*, 1752−1758.

(30) Harvey, S. R.; Porrini, M.; Konijnenberg, A.; Clarke, D. J.; Tyler, R. C.; Langridge-Smith, P. R.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E. Dissecting the dynamic conformations of the metamorphic protein lymphotactin. *J. Phys. Chem. B* 2014, *118* (43), 12348−12359.

(31) Pagel, K.; Hyung, S.-J.; Ruotolo, B. T.; Robinson, C. V. Alternate Dissociation Pathways Identified in Charge-Reduced Protein Complex Ions. *Anal. Chem.* 2010, *82*, 5363−5372.