

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Design and Functionalization of Artificial Protein Assemblies via Biologically Inspired Interactions

Permalink

<https://escholarship.org/uc/item/90j58399>

Author

Subramanian, Rohit Harihara

Publication Date

2020

Supplemental Material

<https://escholarship.org/uc/item/90j58399#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Design and Functionalization of Artificial Protein Assemblies via Biologically Inspired
Interactions

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

in

Chemistry

by

Rohit Harihara Subramanian

Committee in charge:

Professor F. Akif Tezcan, Chair
Professor Karen Christman
Professor Kamil Godula
Professor Michael Sailor
Professor Jerry Yang

2020

©

Rohit Harihara Subramanian, 2020

All rights reserved

The Dissertation of Rohit Harihara Subramanian is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

To my friends and family for your continued support and encouragement

EPIGRAPH

“Principles aren’t principles when you pick and choose when you’re gonna follow them”

Chidi Anagonye

“What I cannot create, I do not understand.”

Richard Feynman

TABLE OF CONTENTS

SIGNATURE PAGE	iii
DEDICATION	iv
EPIGRAPH.....	v
TABLE OF CONTENTS.....	vi
LIST OF ABBREVIATIONS.....	x
LIST OF FIGURES	xiii
LIST OF TABLES	xvi
LIST OF SUPPLEMENTARY VIDEOS	xvii
ACKNOWLEDGEMENTS.....	xviii
VITA.....	xxiii
ABSTRACT OF THE DISSERTATION	xxv
Chapter 1: Design and Construction of Functional Protein Assemblies.....	1
1.1 Introduction.....	1
1.2 Employing native biological interactions to create novel protein assemblies	3
1.2.1 Fusion of natively oligomeric proteins	3
1.2.2 Exploiting protein-ligand interactions	6
1.2.3 Nucleic acid base pairing as a molecular template for supramolecular assembly.....	8
1.3 Employing native biological interactions to create novel protein assemblies	10
1.3.1 Design of noncovalent protein-protein interactions.....	10
1.3.2 Metal-mediated protein self-assembly	12
1.3.3 Disulfide-mediated protein self-assembly	15
1.4 Developing hybrid protein-based materials	17
1.5 Dissertation objectives	20
1.6 References.....	21
Chapter 2: Enzyme-directed Modification of Two-Dimensional Crystalline Protein Lattices	29
2.1 Abstract.....	29
2.2 Introduction.....	29
2.3 Results and Discussion	31
2.3.1 Design and characterization of chemically modified RIDC3+ybbR arrays	31
2.3.2 Generation of RIDC3+CoA lattices.....	43

2.3.3 Design and characterization of genetically modified RIDC3-ybbR and RhuA-ybbR.....	46
2.4 Conclusions.....	52
2.5 Materials and Methods.....	56
2.5.1 Protein expression and purification.....	56
2.5.2 Preparation of ybbR-N ₃ peptide	58
2.5.3 Preparation of modified-CoA conjugates	59
2.5.4 Chemical conjugation of self-assembled RIDC3 crystals	59
2.5.5 Preparation of RIDC3-ybbR and RhuA-ybbR crystals.....	60
2.5.6 Gel digestion and MS-MS analysis of RIDC3+ybbR crystals.....	61
2.5.7 Enzymatic labeling procedure.....	62
2.5.8 Confocal microscopy	63
2.5.9 Negative-stain transmission electron microscopy (ns-TEM)	63
2.5.10 Atomic force microscopy (AFM)	64
2.5.11 Fluorescence microplate reader measurements	64
2.6 Acknowledgments.....	64
2.7 References.....	65
Chapter 3: Self-assembly of a Designed Nucleoprotein Architecture Through Multimodal Interactions.....	69
3.1 Abstract.....	69
3.2 Introduction.....	69
3.3 Results and Discussion	71
3.3.1 Design and self-assembly of DNA-protein chimeras	71
3.3.2 Structural characterization of RIDC3-10a/b lattices.....	80
3.3.3 Construction of a 3D structural model for RIDC3-10a/b lattices	84
3.3.4 Computational analysis of protein-DNA interactions within the RIDC3-10a/b architecture.....	95
3.3.5 Effect of DNA length and sequence on RIDC3-DNA self-assembly.....	97
3.4. Conclusions.....	103
3.5 Materials and Methods.....	103
3.5.1 General Considerations	103
3.5.2 Site-directed mutagenesis, protein expression, and purification	103
3.5.3 Preparation of RIDC3-DNA conjugates	104
3.5.4 Characterization of RIDC3-DNA conjugates	105
3.5.5 Self-assembly of 2D RIDC3-10a/b lattices.....	106
3.5.6 Negative-stain transmission electron microscopy (ns-TEM) analysis.....	106
3.5.7 Job's analysis of RIDC3-10a/b self-assembly	107
3.5.8 Confocal microscopy	107
3.5.9 Scanning electron microscopy (SEM)	108
3.5.10 Atomic force microscopy (AFM)	108
3.5.11 Monitoring RIDC3-10a/b self-assembly with Small Angle X-ray Scattering (SAXS).....	108
3.5.12 Monitoring thermal disassembly of RIDC3-10a/b lattices with SAXS...	109
3.5.13 Cryo-EM data acquisition and processing.....	109
3.5.14 Structural modeling.....	110

3.5.15 Molecular modeling and simulations.....	111
3.5.16 DNA structure analysis.....	114
3.5.17 Quantification of Zn ²⁺ in assembled arrays.....	114
3.5.18 Sedimentation velocity analytical ultracentrifugation (SV-AUC).....	115
3.6 Acknowledgments.....	115
3.7 References.....	116
 Chapter 4: Constructing Protein Polyhedra via Orthogonal Chemical Interactions.....	 121
4.1 Abstract.....	121
4.2 Introduction.....	122
4.3 Results and Discussion.....	123
4.3.1 Design of bimetallic protein cages.....	123
4.3.2 Reversible assembly of dodecameric cages.....	139
4.3.3 Formation of a hexameric cage.....	146
4.4 Conclusions.....	147
4.5 Materials and Methods.....	149
4.5.1 Synthesis of the IHA ligand.....	149
4.5.2 Protein expression and purification.....	150
4.5.3 Protein labelling and post-labelling purification.....	151
4.5.4 Redesign of CFMC1 interfaces.....	151
4.5.5 Crystallography.....	152
4.5.6 Crystallographic metal content analysis.....	153
4.5.7 Protein cage sample preparation.....	155
4.5.8 Negative-stain transmission electron microscopy.....	156
4.5.9 Oligomerization state determination using AUC.....	156
4.5.10 Preparation of samples involving crystal dissolution.....	156
4.5.11 Calculation of BMC void volumes.....	157
4.5.12 Solution self-assembly, disassembly and thermal stability of BMC3 and BMC4.....	157
4.5.13 Cryo-EM sample preparation.....	158
4.5.14 Cryo-EM data acquisition and image processing.....	158
4.5.15 Model building and refinement.....	159
4.5.16 Encapsulation of rhodamine in BMC3 cages.....	160
4.5.17 Statistics and reproducibility.....	161
4.6 Acknowledgments.....	162
4.7 References.....	163
 Chapter 5: Conclusions.....	 167
5.1 Introduction.....	167
5.2 Construction of functionalized crystalline lattices.....	167
5.3 Design of protein-NA hybrid materials.....	169
5.4 Design of protein cages with emergent functions.....	172
5.5 References.....	175

Appendix 1: Scripts used for generating a structural model of the RIDC3-10a/b nucleoprotein architecture.....	176
A.1 Script 1: Generating unique arrangements of protein and DNA, symmetrizing, and creating projection maps for structural modeling	176
A.2 Script 2: Generating projection maps using dimeric protein-DNA modules.....	184
A.3 References	188
Appendix 2: NMR spectra	189

LIST OF ABBREVIATIONS

2D	two-dimensional
ACP	acyl carrier protein
AcpH	acyl carrier protein hydrolase
AFM	atomic force microscopy
ALS	Advanced Light Source
Arg	arginine
Asp	aspartic acid
AUC	analytical ultracentrifugation
AuNP	gold nanoparticle
BMC	bimetallic cage
β ME	β -mercaptoethanol
CB	conjugation buffer
CCD	charge-coupled device
CCM	cytochrome C maturation
CD	circular dichroism
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
CoA	coenzyme A
ConA	concanavalin A
Cryo-EM	cryogenic electron microscopy
Cyt <i>cb</i> ₅₆₂	cytochrome <i>cb</i> ₅₆₂
Cys	cysteine
DBCO-NHS	dibenzocyclooctyne-N-hydroxysuccinimidyl ester
DCM	dichloromethane
DIPEA	N,N-diisopropylethylamine
DMF	dimethylformamide
dsDNA	double-stranded deoxyribonucleic acid

DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ESI-MS	electrospray ionization mass spectrometry
FPLC	fast protein liquid chromatography
GFP	green fluorescent protein
Glu	glutamic acid
HA	hydroxamic acid
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
HPLC	high-performance liquid chromatography
HSAB	Hard-Soft Acid-Base
HuHF	human heavy-chain ferritin
IHA	iodo-hydroxamic acid
IPTG	isopropyl β -D-1-thiogalactopyranoside
KDPGal	2-keto-3-deoxy-6-phosphogalactonate
LC	liquid chromatography
Lys	lysine
MALDI	matrix-assisted laser desorption/ionization
MBPC1	metal binding protein complex 1
MDPSA	metal-directed protein self-assembly
MES	2-(N-morpholine)ethanesulfonic acid
MeTIR	metal-templated interface redesign
MD	molecular dynamics
MOPS	3-(N-morpholino)propanesulfonic acid
MS/MS	tandem mass spectrometry
NA	nucleic acid

NaOAc	sodium acetate
NAMD	nanoscale molecular dynamics
NaPi	sodium phosphate
ns-TEM	negative-stain transmission electron microscopy
PAR	4-(2-pyridylazo)resorcinol
PDB	protein data bank
Ppant	phosphopantetheine
PPI	protein-protein interaction
PPTase	phosphopantetheinyl transferase
RIDC1	Rosetta interface designed cytochrome 1
RIDC3	Rosetta interface designed cytochrome 3
RhuA	L-rhamnulose-1-phosphate aldolase
rpm	rotations per minute
SAXS	small-angle X-ray scattering
SEM	scanning electron microscopy
sfGFP	superfolder green fluorescent protein
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
ssDNA	single-stranded deoxyribonucleic acid
SSRL	Stanford Synchrotron Radiation Lightsource
sulfo-SMCC	sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate
SV-AUC	sedimentation velocity analytical ultracentrifugation
TEM	transmission electron microscopy
TFA	trifluoroacetic acid
Tris	tris(hydroxymethyl)aminomethane hydrochloride
UA	uranyl acetate
UV-Vis	ultraviolet-visible spectroscopy
VMD	visual molecular dynamics

LIST OF FIGURES

Figure 1.1 Examples of natural protein assemblies	2
Figure 1.2 Fusion strategies for protein self-assembly	5
Figure 1.3 Using protein-ligand interactions to direct self-assembly	7
Figure 1.4 DNA-mediated assembly of proteins	9
Figure 1.5 The use of electrostatic or hydrophobic interactions to mediate protein self-assembly	11
Figure 1.6 Strategies used for metal-mediated protein self-assembly	14
Figure 1.7 Disulfide-mediated protein assemblies.....	16
Figure 1.8 Using 2D protein lattices to generate hybrid biomaterials.....	19
Figure 2.1 Schematic of ybbR addition onto RIDC3 for enzymatic labeling.....	32
Figure 2.2 Characterization and enzymatic labeling of RIDC3+ybbR arrays	34
Figure 2.3 Characterization of synthesized ybbR-N ₃ peptide.....	35
Figure 2.4 TEM characterization of modified RIDC3 crystals.....	35
Figure 2.5 MS/MS analysis of ybbR conjugation onto RIDC3 arrays	37
Figure 2.6 TEM and confocal microscopy characterization of enzymatically labeled RIDC3+ybbR arrays.....	38
Figure 2.7 Generation and characterization of GFP-CoA	39
Figure 2.8 Quantification of fluorescent labeling of RIDC3+ybbR crystals	42
Figure 2.9 Characterization and enzymatic labeling of RIDC3+CoA arrays	45
Figure 2.10 Characterization of genetically modified RIDC3-ybbR.....	48
Figure 2.11 Self-assembly and enzymatic labeling of genetically incorporated RIDC3-ybbR....	49
Figure 2.12 Characterization and enzymatic modification of ^{C98} RhuA-ybbR. (a) Cartoon schematic of genetic incorporation of ybbR to the C-terminus of RhuA	53
Figure 2.13 Confocal microscopy and TEM characterization of RhuA-ybbR control samples...	54
Figure 2.14 Attempts at reversible labeling of RIDC3+ybbR using AcpH.....	55

Figure 3.1 Design of RIDC3-10a/10b.....	71
Figure 3.2 FPLC chromatograms and ESI-MS spectra of RIDC3-DNA conjugates	72
Figure 3.3 TEM characterization of RIDC3-10a/b self-assembly under various solution conditions	76
Figure 3.4 Initial characterization of self-assembled RIDC3-10a/b lattices.....	77
Figure 3.5 Thermal and EDTA-mediated disassembly of RIDC3-10a/b lattices	78
Figure 3.6 Determination of RIDC3-DNA stoichiometry in RIDC3-10a/b assemblies.....	79
Figure 3.7 Fluorescent dye intercalation into RIDC3-10a/b assemblies	80
Figure 3.8 Characterization of the crystalline RIDC3-10a/b architecture	81
Figure 3.9 ns-TEM characterization of RIDC3-10a/b and RIDC3 assemblies	82
Figure 3.10 3D stacking in RIDC3-10a/b assemblies.....	83
Figure 3.11 Structural modeling of the RIDC3-10a/b architecture	85
Figure 3.12 Geometric considerations for initial rigid-body fitting of protein and DNA structures into experimental cryo-EM density	86
Figure 3.13 Generation of candidate structural models and their calculated projection maps	88
Figure 3.14 Determination of DNA orientation using calculated electron density maps.....	89
Figure 3.15 Top four candidate RIDC3-DNA models based on calculated projection maps.....	90
Figure 3.16 TEM analysis of the effects of alanine point mutations on RIDC3-10a/b self-assembly	91
Figure 3.17 Temperature-dependent SAXS profiles of RIDC3-10a/b assemblies.....	93
Figure 3.18 Inhibition of RIDC3-10a/b self-assembly process by the inclusion of complementary ssDNA.....	94
Figure 3.19 pH dependence of the calculated surface charges of RIDC3 dimers	95
Figure 3.20 Molecular dynamics simulations of the RIDC3-10a/b lattice and comparison of calculated and experimental 2D projection maps	96
Figure 3.21 Structural integrity of DNA duplexes in various contexts as determined by MD simulations	98
Figure 3.22 Determination of the oligomerization state of RIDC3-10a/10b in solution by AUC 99	

Figure 3.23 Characterization of various RIDC3-DNA constructs evaluated for self-assembly.	100
Figure 3.24 Role of DNA sequence specificity in RIDC3-DNA self-assembly.....	101
Figure 3.25 Structural model of RIDC3-12a/b lattices.....	102
Figure 4.1 Design of protein cages.....	124
Figure 4.2 Characterization of the IHA ligand and the BMC constructs.....	126
Figure 4.3 Structural comparison of CFMC1 and BMC1 cages.....	128
Figure 4.4 Characterization of BMC2 cages.....	129
Figure 4.5 ns-TEM characterization of BMC constructs.....	130
Figure 4.6 Cavity volumes of BMC cages.....	132
Figure 4.7 Anomalous densities of engineered metal binding sites and conformational flexibility of Cys82-HA site.....	134
Figure 4.8 Characterization of BMC3 cages.....	139
Figure 4.9 Solution characterization of self-assembled BMC3 and BMC4 cages.....	141
Figure 4.10 Cryo-EM analysis of BMC3 cages.....	144
Figure 4.11 Encapsulation of rhodamine inside BMC3 cages.....	145
Figure 4.12 Characterization of BMC4 cages.....	148
Figure 5.1 Formation of ^{C98} RhuA-ACP arrays.....	168
Figure 5.2 Proposed mutations for improving RIDC3-10a/b stability.....	171
Figure 5.3 Generating new hydroxamate-based cages.....	173
Figure A2.1 ¹ H NMR spectrum of 2-chloro-N-hydroxyacetamide.....	189
Figure A2.2 ¹³ C NMR spectrum of 2-chloro-N-hydroxyacetamide.....	190
Figure A2.3 ¹ H NMR spectrum of 2-iodo-N-hydroxyacetamide.....	191
Figure A2.4 ¹³ C NMR spectrum of 2-iodo-N-hydroxyacetamide.....	192

LIST OF TABLES

Table 4.1 Amino acid sequences of protein variants employed in the current research.....	127
Table 4.2 X-ray data collection, processing and refinement statistics.....	131
Table 4.3 Crystallographic quantification of metal content at BMC1 coordination sites	136
Table 4.4 Crystallographic quantification of metal content at BMC2 coordination sites	136
Table 4.5 Crystallographic quantification of metal content at BMC3 coordination sites	137
Table 4.6 Crystallographic quantification of metal content at BMC4 coordination sites	137
Table 4.7 Cryo-EM data collection, processing, and refinement statistics.....	142
Table 4.8 Crystallization conditions for the different variants	152

LIST OF SUPPLEMENTARY VIDEOS

Supplementary Movie 3.1 RIDC3-10ab-lattice	89
--	----

ACKNOWLEDGEMENTS

I would like to thank my advisor, Akif Tezcan, for allowing me to work in his lab and supervising my research throughout my graduate school career. His unending enthusiasm and passion for science was, and remains, inspirational, contagious, and encouraging, especially for a young researcher new to independent academic research. There have been many peaks and valleys throughout my time in graduate school but Akif's unwavering support and reassurance provided me with confidence to continue when faced with scientific and personal challenges. I have grown into a better scientist and person due to his guidance and am grateful for his faith in me during periods when I did not believe in myself.

I have been fortunate to partake in many collaborative projects during my time at UC San Diego. I would like to thank Lorillee Tallorin and Eunice Kim from the Burkart lab and Swagat Sahu and Matthew Thompson from the Gianneschi lab for their invaluable assistance in developing the enzyme functionalization work in Chapter 2. As members of the BioAutoCatalysis grant with Yuta Suzuki and Steven Chabolla, we were able to share research ideas and knowledge from our individual areas of expertise. I would also like to thank Norm Olson, James Bouwer, Timothy Booth, Xiaodong Yan, and Professor Timothy Baker for their assistance in TEM data collection and processing over the years. I have spent countless hours at the TEM facility and it was made easier by working among an excellent group of scientists. For the cryo-EM work in Chapter 3, I was lucky to travel to the Professor Henning Stahlberg's lab in Basel, Switzerland and work with his group for two weeks. I thank Professor Stahlberg and members of his research group, in particular: Ricardo Righetto, Mohamed Chami, Lubomir Kovacik, and Ken Goldie for welcoming me and providing invaluable cryo-EM expertise to aid in completing the work in Chapter 3.

My time in the Tezcan lab has allowed me to work with and amongst remarkable researchers who have made the journey through graduate school much more bearable. They have supported me through tough times, listened to my incoherent rants, and nodded along through my arbitrary fits of rage – they have put up with me and allowed me to develop as a researcher and adult human. I thank Yuta for teaching me the ropes of the Tezcan lab when I first joined and guiding me through my first research project. We did not always see eye-to-eye but he was always tolerant of my ignorant first-year questioning. I will always be inspired by Woon Ju’s work ethic and dedication to science and I am grateful for her advice and encouragement over the years. I have spent countless hours driving around (or rather, being driven around) San Diego to find new places to eat and drink with Pam and our shared interests in music and popular culture provided a necessary respite from the constant challenges of Tezcan lab science.

I owe a huge debt of gratitude to Sarah for her mentorship over the years and I have enjoyed talking science, soccer, and generally complaining to each other about the frustrations of graduate school (and beyond). I had many struggles in the first half of my graduate career and Sarah’s encouragement, scientific advice, and friendship were invaluable to my successes as a scientist. I thank Lewis for welcoming me as a friend when I first moved here and didn’t know anyone and inviting me on crazy Pokemon Go adventures to regroup after a long workday. Jake, Faith and I joined the lab at the same time and while the three of us could not be more different in our personalities and outlook on life, we have managed to help each other through graduate school. I may be the last of us to graduate but, well, let’s move on (it’s not a race). I was fortunate to mentor an undergraduate researcher (let’s call him John) for a few years and we both learned a lot from each other on how to approach a new research project, teach (and learn), and time management. I was fortunate to work with Eyal for the research discussed in Chapter 4 but beyond the science,

Eyal's cheery personality was enough to brighten my mood any day. I thank Julian for his helpful discussions and expertise regarding crystallography and his willingness to discuss and help sort out scientific challenges.

Albert's eclectic and quirky personality makes for an entertaining work environment. His passion for science and excitement when thinking up new, crazy ideas is undeniable and I look forward to reading about his future accomplishments (and his adventures in making fried chicken). I am thankful to Jie for always being open to discussing science and, importantly, basketball and sesame bread. I thank Kenneth for our many shared conversations (some even scientific!) over the banalities of everyday life, coffee and that one Denny's on the drive to and from LA. I envy Yiyi's optimism and cheerful outlook towards life and I continue to learn nuggets of philosophical wisdom from her daily. I have witnessed Ling's transformation from a (relatively) timid first year asking me to help her to the fierce dog-mom (and incredible scientist) she is today. I am grateful for her friendship and her support throughout my years in graduate school. Hannah's passion and enthusiasm for science will never cease to amaze me and her willingness to question and ponder scientific problems to anyone who will listen has taught me a great deal about nitrogenase. I am grateful for her friendship over the past few years and I suppose I must thank her for indirectly introducing me to Katie.

The most recent iteration of the materials subgroup has made for a scientifically fulfilling research atmosphere with a heavy dose of fun personalities. Though Tomo has left the lab, his two years as a visiting researcher were a blessing and his quirky wisdom is sorely missed. It has been very enjoyable to work with Iris over the past year and watch her grow as a talented researcher. I thank Nicole for her encouragement (even if they are coated in a generous helping of pessimism) and fun chats about movies, music, and Adjaruli. I have worked closely with Rob (both

scientifically and literally since we sit back-to-back) for the past 5 years and his passion for science continues to impress me daily. It's possible our constant banter could be viewed as a distraction to those trying to work in the lab but his keen research mind and friendship have kept me sane as a scientist and person during graduate school. Last (and least because she's the smallest member of the lab), I am thankful to Miumiu for always reminding me that strangers (*i.e.* any person who walks by the conference room) must be barked at to assert dominance and providing a great deal of emotional support and joy.

I thank the friends near and far who have supported me throughout the years. I have been on many random adventures in San Diego or Los Angeles with David and he is a constant reminder not to take life too seriously. I have been fortunate to remain close with many of the graduate students who joined the program when I did and de-stressing over a weekend bbq has made the graduate school experience more enjoyable. I am especially grateful and indebted to my girlfriend Katie for her unwavering support and encouragement as I finished my graduate work. Her presence in my life has been a true joy and I will always cherish her quick decision-making skills and her love of olives.

I would like to thank my parents and sister for their support over the years. I may not have always explained what I was doing in lab and why it was (or wasn't) working but their frequent visits to San Diego were a necessary break from the challenges of graduate research and I am thankful for their encouragement throughout my life. I am especially grateful to my grandparents for inspiring me to be better and work harder. Their weekly calls to check in have been a large source of motivation to finish my Ph.D.

This dissertation was funded in part by UCSD's Chemical Biology Interfaces Training Grant (NIH).

Chapter 2 is reproduced, in part, from a manuscript currently being prepared for submission: Subramanian, R. H.; Suzuki, Y.; Tallorin, L.; Sahu, S.; Tezcan, F.A. Enzyme-directed Modification of Two-Dimensional Crystalline Protein Lattices.

The dissertation author is primary author on all reprinted materials.

Chapter 3 is reproduced, in part, with permission, from Subramanian, R. H.*; Smith, S. J.*; Alberstein, R. G.; Bailey, J. B.; Zhang, L.; Cardone, G.; Suominen, L.; Chami, M.; Stahlberg, H.; Baker, T., Tezcan, F.A. Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions, *ACS Cent. Sci.* **2018**, *4* (11), 1578-1586.

The dissertation author is primary author on all reprinted materials.

Chapter 4 is reproduced, in part, with permission, from Golub, E.; Subramanian, R. H.; Esselborn, J.; Alberstein, R. G.; Bailey, J. B.; Chiong, J. A.; Yan, X.; Booth, T.; Baker, T. S.; Tezcan, F. A. Constructing Protein Polyhedra via Orthogonal Chemical Interactions, *Nature* **2020**, *578*, 172-176.

VITA

EDUCATION

- 2013 B.S., Biomedical Engineering, Northwestern University
- 2015 M.S., Chemistry, University of California San Diego
- 2020 Ph.D., Chemistry, University of California San Diego

HONORS AND AWARDS

- 2015 – 2016 Chemical Biology Interfaces Training Grant
- 2013 University of California San Diego Chemical Biology Excellence Award

PUBLICATIONS

Golub, E.; **Subramanian, R. H.**; Esselborn, J.; Alberstein, R. G.; Bailey, J. B.; Chiong, J. A.; Yan, X.; Booth, T.; Baker, T. S.; Tezcan, F. A. Constructing Protein Polyhedra via Orthogonal Chemical Interactions, *Nature* **2020**, *578*, 172-176.

Subramanian, R. H.*; Smith, S. J.*; Alberstein, R. G.; Bailey, J. B.; Zhang, L.; Cardone, G.; Suominen, L.; Chami, M.; Stahlberg, H.; Baker, T., Tezcan, F.A. Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions, *ACS Cent. Sci.* **2018**, *4* (11), 1578-1586.

Zhang, L.; Bailey, J. B.; **Subramanian, R. H.**; Tezcan, F. A. Hyperexpandable, Self-Healing Macromolecular Crystals with Integrated Polymer Networks, *Nature* **2018**, *557*, 86–91.

Bailey, J. B.; **Subramanian, R. H.**; Churchfield, L. A.; Tezcan, F. A., Metal-Directed Design of Supramolecular Protein Assemblies, In *Methods in Enzymology*; Pecoraro, V. L., Ed.; Academic Press, **2016**; Vol. 580.

Smith, S. J.; Radford, R. J.; **Subramanian, R. H.**; Barnett, B. R.; Figueroa, J. S.; Tezcan, F. A. Tunable helicity, stability and DNA-binding properties of short peptides with hybrid metal coordination motifs, *Chem. Sci.* **2016**, *7* (8), 5453-5461

Trivedi, E. R.; Ma, Z.; Waters, E. A.; Macrenaris, K. W.; **Subramanian, R.**; Barrett, A. G.; Meade, T. J.; Hoffman, B. M. Synthesis and Characterization of a Porphyrazine-Gd(III) MRI Contrast Agent and *in vivo* imaging of a Breast Cancer Xenograft Model. *Contrast Media Mol. Imaging* **2014**, *9* (4), 313-322.

FIELDS OF STUDY

Major Field: Chemistry

Studies in Bioinorganic Chemistry, Chemical Biology
Professor F. Akif Tezcan

ABSTRACT OF THE DISSERTATION

Design and Functionalization of Artificial Protein Assemblies via Biologically Inspired
Interactions

by

Rohit Harihara Subramanian

Doctor of Philosophy in Chemistry

University of California San Diego, 2020

Professor F. Akif Tezcan, Chair

Proteins are Nature's fundamental multitools, fulfilling crucial roles in catalyzing complex chemical reactions, mediating cell-cell signaling to coordinate biochemical responses, and providing the structural scaffolding necessary for intracellular transport and cell motility among a myriad of other functions. The functional diversity of proteins is enhanced by the associations of intracellular protein, nucleic acid and small molecule components to generate sophisticated self-assembled architectures. The "bottom-up" construction of biological components is a burgeoning field of study which seeks to generate novel functional assemblies by directing protein interactions

in a controlled fashion. Protein complexes in Nature are driven by an accumulation of weak noncovalent interactions over large interfaces, which ensure specific and stable assembly of the desired architecture. However, such nuanced interactions are difficult to emulate by intuition (or computation), making their designability one of the foremost challenges in protein engineering. Nevertheless, our strategies streamline such design efforts via the integration of well-studied biological motifs into self-assembling protein scaffolds to generate structurally and functionally diverse architectures.

Previous studies have shown that reversible yet specific interactions, such as metal-coordination and disulfide bonding, can be used to programmably assemble both discrete and pseudoinfinite protein oligomers. We first utilized designed crystalline lattices to generate functional materials through post-translational modification of assembled proteins using biological enzymes. Incorporation of a functional peptide substrate onto our protein scaffolds enables the use of phosphopantetheinyl transferase (PPTase) enzymes to site-specifically tailor the surface of crystalline two-dimensional protein materials. In addition to expanding the functionality of existing designed assemblies, we explore the use of biologically relevant motifs to create novel protein-based architectures. Integrated protein and nucleic acid (NA) complexes are among the most complex biological machines in Nature, but the design of a synthetic assembly of protein and NA components via synergistic interactions remains an outstanding challenge in biomolecular design. We create a protein-DNA conjugate via covalent tethering of a monomeric protein and single-stranded DNA to enable the assembly of an artificial three-dimensional nucleoprotein architecture through protein-metal coordination, Watson-Crick DNA base pairing, and DNA-protein interactions. Appropriately balanced thermodynamics of these interactions is necessary to achieve well-ordered self-assembly products instead of disordered protein-NA aggregation (as we

observe when one set of interactions dominates). Finally, we use siderophore-inspired hydroxamate motifs, which selectively bind Fe^{3+} ions very tightly, to construct bimetallic protein cages from a monomeric protein building block. We show that a protein monomer modified with both hydroxamate groups and zinc-binding motifs assembles through concurrent binding of Zn^{2+} and Fe^{3+} ions to form dodecameric and hexameric protein cages. These cages can assemble and disassemble in response to multiple stimuli, and can be used for cargo encapsulation and storage. Overall, we show that the integration of native biological components and protein design strategies enables the construction of novel functional protein assemblies that can serve to guide future protein engineering efforts.

Chapter 1: Design and Construction of Functional Protein Assemblies

1.1 Introduction

Proteins, biological linear polymers constructed from a suite of 20 amino acid building blocks, are ubiquitous in nature and fulfill a vast array of functions that support cell survival. Biological processes ranging from cellular respiration, DNA replication, metabolic catalysis, and structural scaffolding are fulfilled by the compositionally diverse proteome. The complexity of protein function in a crowded cellular environment is, in part, supported by their self-assembly into multimeric complexes, affording new properties difficult to access from a monomer.^{1,2} A multitude of homomeric and heteromeric complexes are formed by the assembly of proteins to facilitate sophisticated chemical reactions (*e.g.* biological nitrogen fixation by nitrogenase³ and facilitating electron cascades to generate O₂ by photosystem II⁴), energy production by ATP synthase through concerted protein motions across a proton gradient⁵, or the regulation of cellular protein concentration through the degradation of excess or damaged proteins by proteasomes⁶ (**Figure 1.1a, b**). In addition to these discrete oligomeric species, proteins can organize into infinite arrays to provide structural support via rapid assembly/disassembly of microtubule filaments⁷ or protective surface coats with crystalline two-dimensional (2D) s-layer proteins that are found on nearly all bacteria and archaea⁸ (**Figure 1.1c**). The aforementioned examples rely primarily on protein-protein contacts but a number of protein complexes critical for proper cellular function require intricate interactions between proteins and nucleic acids (NAs), such as the packaging of genetic information through DNA compaction into nucleosomes⁹ or protecting vital NA cargo by the organization of capsid proteins into protein cages¹⁰ (**Figure 1.1d**). The assembly and ensuing function of natural protein complexes is a source of inspiration and motivation for the design of artificial protein complexes with novel properties that match, or even surpass, those observed in

nature. A fundamental challenge of protein design lies in understanding the properties that govern protein association and harnessing innate chemical diversity to drive the construction of unique supramolecular assemblies.

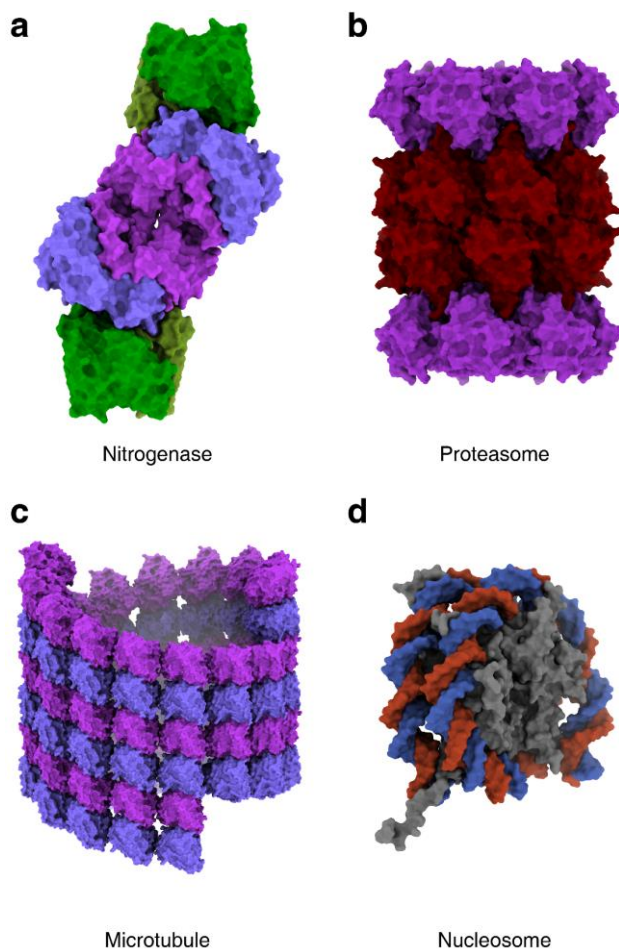


Figure 1.1 Examples of natural protein assemblies. **(a)** Nitrogenase complex (PDB ID: 1n2c) consisting of homodimer Fe protein (shown in green) and heterotetramer MoFe protein (shown in blue and purple). The Fe protein subunit is the only known biological reductant of MoFe protein, which houses the catalytic center for the reduction of N_2 to NH_3 . **(b)** Proteasome core (PDB ID: 1pma) comprised of an outer ring of proteins (α subunit, shown in purple) and an inner ring of proteins (β subunit, shown in red). The α subunits act to gate which proteins enter the β subunit core, which contains protease active sites. **(c)** Microtubule (PDB ID: 3j2u) comprised of α (colored in purple) and β (colored in blue) subunits of tubulin. Microtubules are cytoskeletal elements that provide structural support in cells. **(d)** Nucleosome core (PDB ID: 1a0i) shown in grey with dsDNA coiled around the protein (shown in red and blue). Histone proteins form an octameric complex that enables the coiled organization of DNA, important in the efficient compaction of DNA in the cell.

The manipulation and incorporation of biological materials as nano- and micro-scale building blocks is not a new concept: lipid molecules (which make up the ubiquitous bilayer cell membrane of nearly all organisms) and nucleic acids (the genetic backbone of life itself) have been key players in the “bottom-up” construction of nanoscale biomaterials.¹¹⁻¹³ In particular, the specificity and high fidelity of Watson-Crick base pairing interactions has enabled the creation of elaborate DNA architectures, termed “DNA origami”, by the precise folding of a long single-stranded DNA scaffold using short staple DNA strands.¹⁴ In contrast, the higher chemical complexity of protein side chains, relative to nucleobases, has precluded the creation of a simple set of design rules for precise and controlled protein self-assembly. Proteins adopt complicated 3D folds through a multitude of weak, noncovalent interactions. Whereas these interactions provide a necessary dynamicity in the natural function of proteins (*e.g.* the “walking” of kinesin motor proteins across a microtubule¹⁵), the challenge of accurately designing such interactions remains an active area of research. Therefore, protein engineers have developed an array of biochemical tools and design strategies to direct proteins into supramolecular architectures as discussed below.

1.2 Employing native biological interactions to create novel protein assemblies

1.2.1 Fusion of natively oligomeric proteins

Many proteins exist naturally as oligomers, forming well-defined symmetric structures comprised of multiple copies of a singular protein building block.¹ Creating genetic linkages between pairs (or more) of symmetric proteins has been used as a strategy to precisely position proteins into higher-order architectures. For instance, Noble and coworkers created genetically fused combinations of C_2 , D_2 and C_4 symmetric proteins to yield 1D filaments and 2D crystalline arrays (**Figure 1.2a**).¹⁶ Individual subunits of each symmetric component were fused at N- and C-termini along a shared rotational symmetry axis to provide rigid contacts between each pair of

proteins and disfavor disordered assemblies through flexible linkages. This strategy was also employed by the Yeates group to generate discrete cage-like protein assemblies.¹⁷⁻¹⁹ In their first report, Yeates and coworkers described a general strategy for designing symmetrical protein architectures through the fusion of two symmetric components at a particular geometric arrangement (*i.e.* defined hinge angle between the linked proteins) to yield long filamentous assemblies and a discrete protein cage, the latter of which was later resolved crystallographically to confirm the efficacy of their fusion design.¹⁷ Yeates and coworkers further built on this strategy by altering the angle between the two symmetric components to generate a larger 24-subunit cage (**Figure 1.2b**).¹⁸ This was accomplished by fusing the trimeric 2-keto-3-deoxy-6-phosphogalactonate (KDPGal) aldolase to the dimeric FkpA protein through an alpha-helical linker to yield a protein cube with octahedral symmetry consisting of dimers along the edges and trimers at the vertices. Interestingly, the inherent flexibility of proteins resulted in a distribution of 12-, 18- and 24-mer cages, highlighting the challenge of controlling conformational flexibility in the construction of artificial protein assemblies. The Yeates lab has continued to expand the scope of genetic fusion of symmetric elements in a recent report, describing the design and characterization of a protein icosahedron through the creation of a double fusion protein consisting of dimeric, trimeric and pentameric subunits (**Figure 1.2c**).²⁰

In contrast with the fusion of rigid linkers used in the above examples, Marsh and coworkers used a flexible linker to connect coiled-coil domains to symmetric proteins to generate well-defined protein cages.^{21,22} The fusion of a C_3 -, C_4 -, or C_5 -symmetric coiled-coil elements to the C_3 -symmetric trimeric esterase using a flexible polyglycine linker resulted in tetrahedral, octahedral, and icosahedral cages respectively (**Figure 1.2d**).²² Different linker lengths were tested to find the optimal design for each cage, suggesting that flexibility between the protein domains is

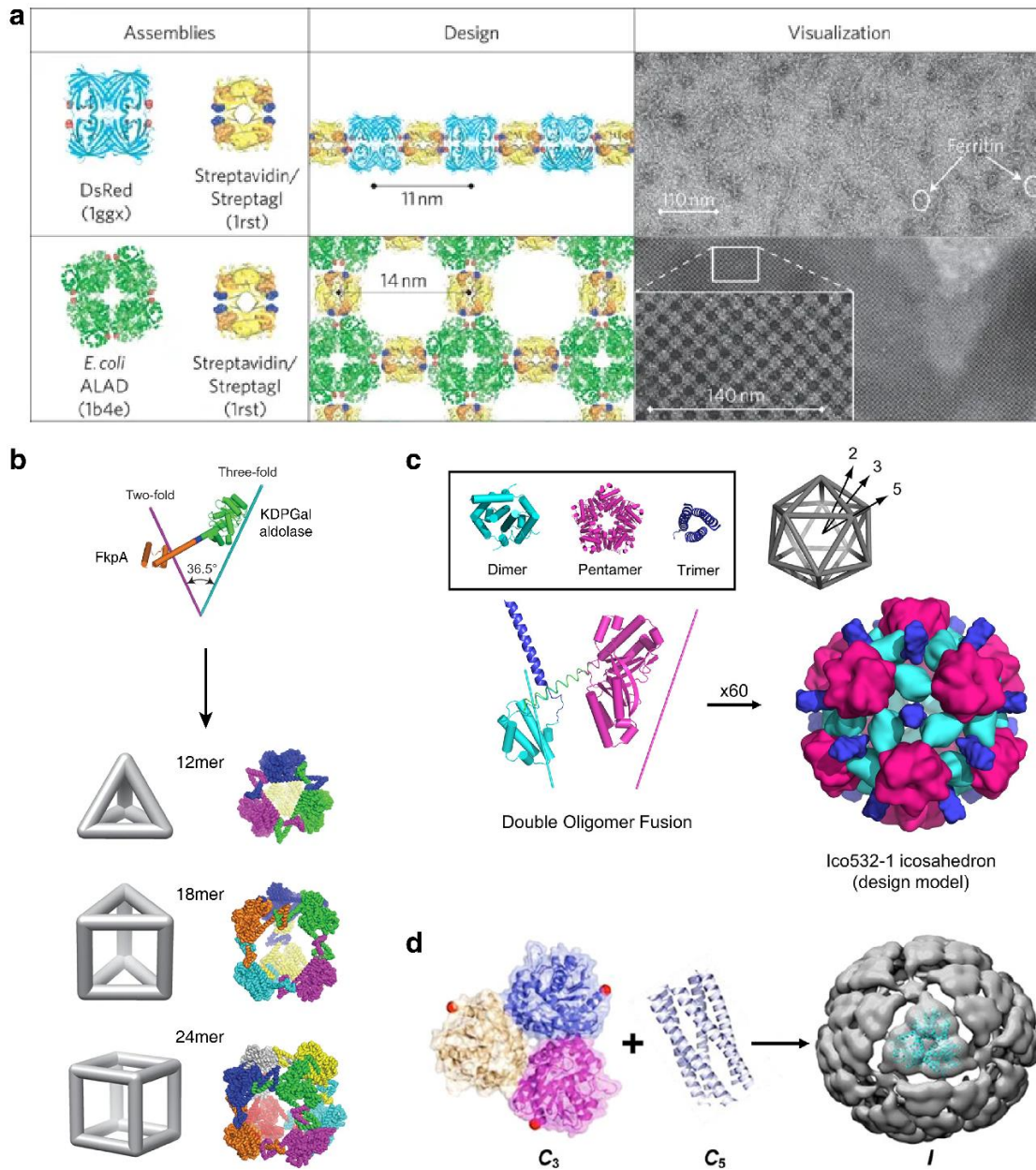


Figure 1.2 Fusion strategies for protein self-assembly. **(a)** Ordered 1D and 2D assemblies formed from the genetic fusion of symmetrical C_2 - and C_4 -symmetric proteins. Adapted from reference 12. **(b)** Fusion of the C_2 -symmetric FkpA and C_3 -symmetric KDPGal aldolase using a rigid α -helical linker to promote a particular angle of curvature upon self-oligomerization. The resultant assemblies were found to contain a mix of 12-, 18- and 24-meric cages. Adapted from reference 18. **(c)** Design of an icosahedral protein by fusing three symmetric components. Upon protein expression and native oligomerization, a 60-subunit icosahedron formed in ideal solution conditions. Adapted from reference 20. **(d)** Genetic fusion of a C_5 -symmetric coiled-coil domain and a C_3 -symmetric protein enables the formation of icosahedral cages with C_5 -symmetric nodes and C_3 -symmetric faces. Adapted from reference 22.

crucial to permit assembly in the specified geometries. The approaches described above demonstrate that coupling symmetric elements, through a rigid or flexible linker, can be used to generate a variety of protein assemblies in a modular fashion.

1.2.2 Exploiting protein-ligand interactions

In addition to harnessing protein-protein interactions to create desired structures, there have been considerable efforts in exploiting specific protein-ligand interactions to drive self-assembly. One important biological interaction is between sugar molecules and their binding partner proteins, lectins.^{23,24} Freeman and coworkers first demonstrated this using a C_2 -symmetric bismannopyranoside linker and the tetrahedral lectin, concanavalin A (ConA) to generate three-dimensional crystalline arrays.²⁵ Hayashi and coworkers used heme-heme protein interactions as their mode of supramolecular polymerization to drive protein assembly.²⁶⁻²⁹ By covalently tethering a heme group at a site-specific location on the four-helix bundle cytochrome b_{562} , 1D supramolecular protein wires were observed mediated by heme-heme pocket interactions (**Figure 1.3a**).²⁶ They further adapted this strategy using the heme binding myoglobin and the high-affinity biotin-streptavidin pair³⁰ to generate two-component protein copolymers using a synthesized heme-bisbiotin small molecule linker (**Figure 1.3b**).²⁸ In this instance, the native proteins did not require any additional modification and construction of the appropriate bisfunctional linker was the driving factor in self-assembly. Biotin can also be chemically conjugated to a symmetric nodule of choice as demonstrated by Schulz and coworkers by linking biotin to C_4 -symmetric L-rhamnulose-1-phosphate aldolase (RhuA) and combining with the D_2 -symmetric streptavidin tetramer to create small two-component protein networks (**Figure 1.3c**).³¹ Overall, this design strategy benefits from using existing protein-ligand interactions with known chemical behavior; however, control over self-assembly is largely dictated by the binding affinity of the ligand and

can lead to kinetic traps or low fidelity assemblies if the interaction strength is not situated in a “Goldilocks” zone.

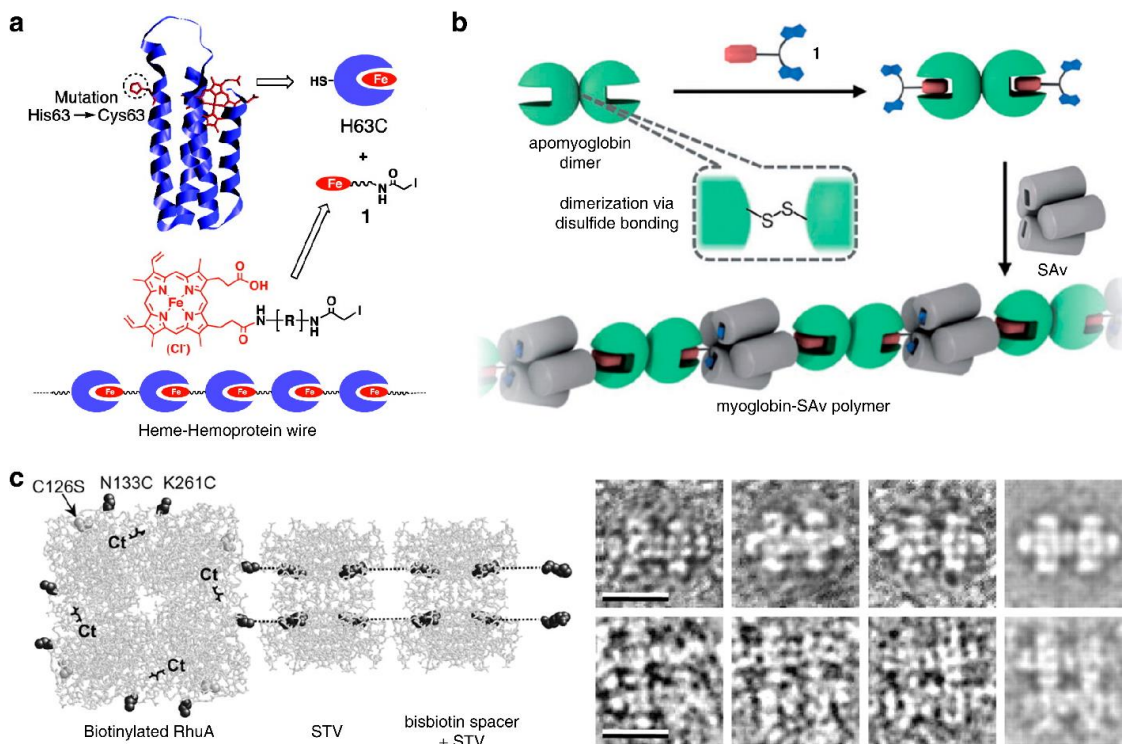


Figure 1.3 Using protein-ligand interactions to direct self-assembly. **(a)** Installation of a Cys residue onto cytochrome b_{562} allows for covalent linkages of the native heme ligand through the reactive iodoacetamide group. Incubation of heme-hemoprotein conjugates results in extended protein wires. Adapted from reference 26. **(b)** Disulfide-mediated dimerization of heme-binding myoglobin enables the co-assembly of two-component protein wires with the tetrameric SAv using a synthesized bisbiotin-heme ligand. Adapted from reference 28. **(c)** RhuA, a C_4 -symmetric protein, is covalently modified with a biotin linker at the specific Cys residues on the protein. When incubated with streptavidin (STV), 2D protein lattices are formed from biotin-avidin interactions. TEM micrographs (right) show the resultant two-component protein networks. Adapted from reference 31.

1.2.3 Nucleic acid base pairing as a molecular template for supramolecular assembly

Sequence-specific Watson-Crick base-pairing has been used as a means to programmably assemble nanoparticle superlattices relying solely on multivalent DNA-DNA bonding to drive crystallization.^{32,33} Mirkin and coworkers demonstrated that proteins can be used as a surrogate for nanoparticles by decorating the surface of catalase with single-stranded oligonucleotides that hybridized with a linker DNA strand, leaving a short single-stranded linker region.³⁴ Complimentary linker strands were used to generate protein-protein or protein-nanoparticle lattices driven by DNA hybridization (**Figure 1.4a, b**).³⁴ The chemical complexity of proteins permitted selective functionalization of amines or thiols with DNA to modulate the lattice parameters of the assembly or facilitate directional bonding along one axis to generate protein nanowires.^{35,36} Aida and coworkers also reported the generation of 1D protein assemblies by modifying the molecular chaperone GroEL with complementary single-stranded oligonucleotides (**Figure 1.4c**).³⁷ The multivalency of the DNA-DNA interactions between GroEL proteins afforded nanotubes high thermal stabilities. In addition, the use of partially complementary strands allowed the generation of “cleavable” nanotubes upon the addition of a fully complementary GroEL-DNA protein conjugate. Mayo and coworkers created a hybrid assembly that required both protein-protein and protein-DNA interactions, engineering a DNA-binding ENH domain³⁸ to form protein dimers which, upon incubation with dsDNA, were able to spontaneously assemble into protein nanowires (**Figure 1.4d**).³⁹ The use of a programmable bonding interaction (DNA hybridization) thus enabled the generation of a structurally diverse set of protein-NA assemblies but necessitates the use of protein-DNA covalent conjugation or exclusively relying on proteins that natively bind NAs.

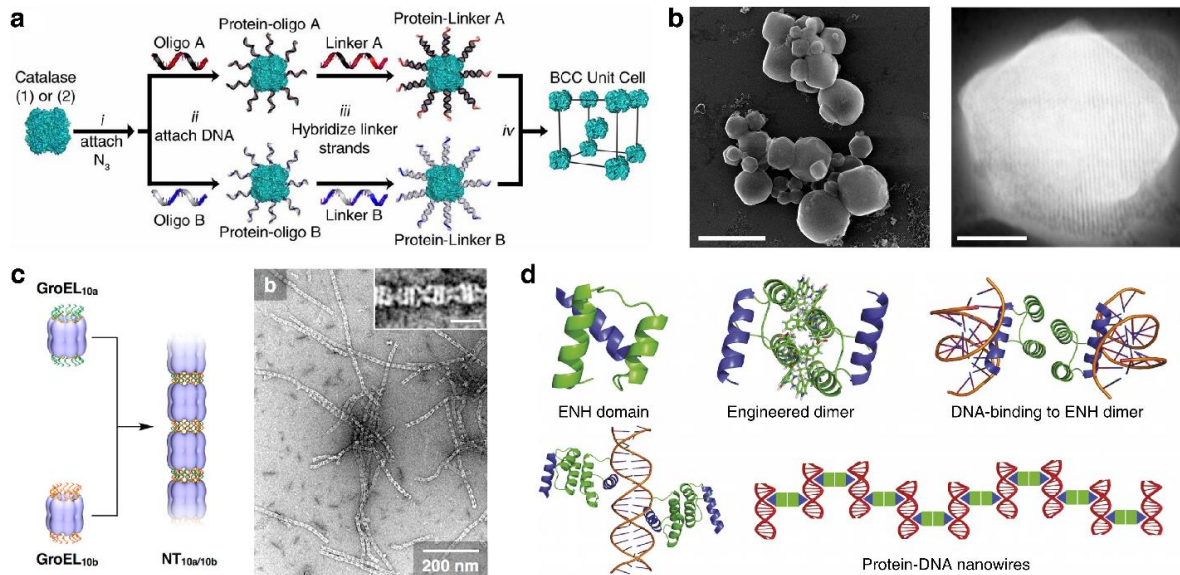


Figure 1.4 DNA-mediated assembly of proteins. **(a)** Indiscriminant labeling of catalase with single-stranded DNA (ssDNA) followed by the hybridization of linker strands creates a coat of DNA duplexes with short sticky ends around the protein. Mixing proteins bearing complementary linker strands can form crystalline lattices **(b)** through DNA-directed organization of proteins. Adapted from reference 34. **(c)** GroEL proteins are covalently modified with ssDNA strands. Incubation of two GroEL proteins bearing complementary strands enables the formation of protein nanotubes (shown on the right) through DNA hybridization. Adapted from reference 37. **(d)** The DNA binding ENH domain is computationally engineered to form a noncovalent dimer. Protein dimers can bind DNA in a pseudo C_2 -symmetric fashion to form extended protein-DNA nanowires. Adapted from reference 39.

1.3 Employing native biological interactions to create novel protein assemblies

1.3.1 Design of noncovalent protein-protein interactions

Natively occurring protein-protein interfaces are often comprised of weak, noncovalent interactions spread over a large area to promote and maintain association.⁴⁰ The precise engineering of surface residues to direct protein assembly is a fundamental challenge of protein design.⁴¹ One design approach taken has been to treat a symmetric protein building block as a nanoparticle and re-engineering its surface with positive or negative charges to generate protein lattices.⁴²⁻⁴⁴ Beck and coworkers used the 24meric human heavy chain ferritin (HuHF) to generate HuHF^{pos} and HuHF^{neg} variants that coassembled into binary protein crystals (**Figure 1.5a**).⁴⁴ X-ray crystallographic analysis revealed extensive salt-bridge contacts at the two-fold symmetric axes of each ferritin cage between pos-neg as well as pos-pos and neg-neg HuHF molecules, facilitated by plasticity at each interface to maximize favorable interactions. Ellington and coworkers used a similar strategy, this time using the monomeric green fluorescent protein (GFP), to favor assembly of well-defined binary protein architectures by mixing a positively (+33 net charge) GFP and a negatively (-17 net charge) GFP (**Figure 1.5b**).⁴⁵ Cryogenic electron microscopy (cryo-EM) characterization of a symmetrical 16-mer assembly revealed an accumulation of weak electrostatic contacts between proteins due to the charge anisotropy of the engineering GFP variants, suggesting that electrostatic interactions can act as a key driver in the construction of a well-structured protein complex.

Advances in computational protein design have enabled the creation of novel protein folds and assemblies with increasingly high fidelity.^{41,46-50} Early examples of designed protein folds relied primarily on mathematical modeling of backbone conformations and side chain packing algorithms to produce solution stable proteins.^{51,52} As computational power increased through

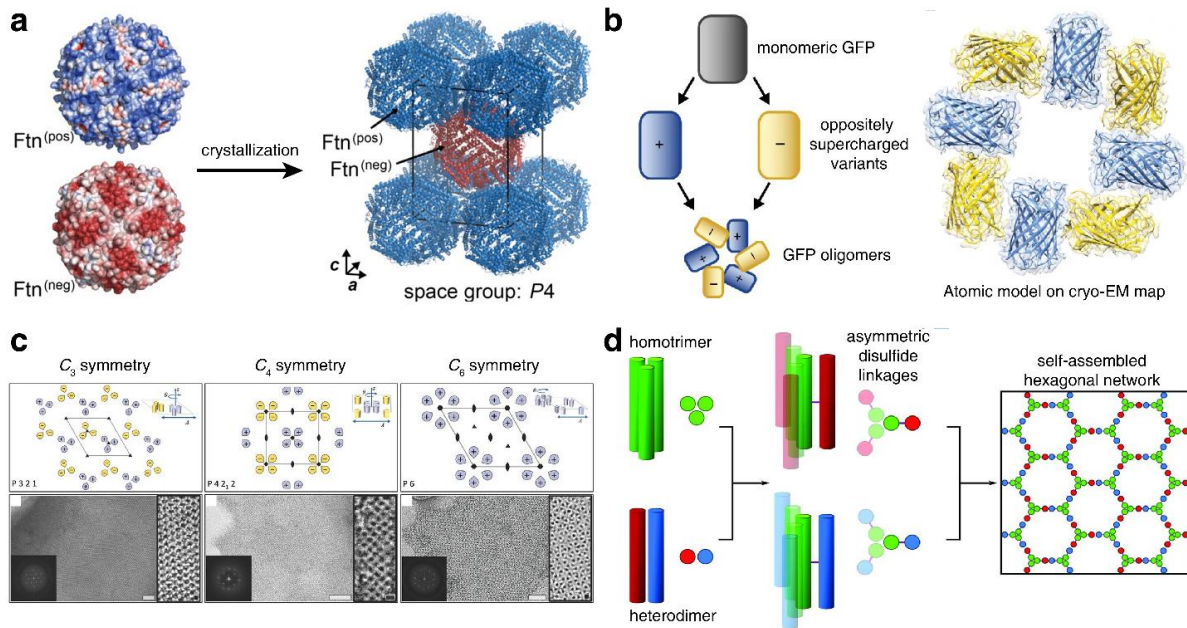


Figure 1.5 The use of electrostatic or hydrophobic interactions to mediate protein self-assembly. **(a)** The surface of human heavy-chain ferritin is genetically engineered to bear a high positive charge (shown in blue) or negative charge (shown in red). Co-crystallization of both proteins results in the formation of binary ferritin lattices mediated by electrostatic interactions between positively and negatively charged ferritins. Adapted from reference 44. **(b)** Monomeric GFP is engineered into positively and negatively charged protein variants to encourage the self-assembly of binary GFP oligomers through electrostatic interactions. Cryo-EM characterization of an octameric GFP oligomer reveals the organization of positive and negative GFP in a ring-like shape. Adapted from reference 45. **(c)** Computational design of protein-protein interfaces between C_3 , C_4 or C_6 -symmetric proteins results in the formation of 2D lattices through hydrophobic contacts between protein oligomers. Adapted from reference 55. **(d)** Coiled-coil domains are used to create self-assembled cages. A homotrimeric coiled-coil is linked to a heterodimeric coiled-coil through asymmetric disulfide linkages to create a hexagonal network that closes into a cage-like architecture. Adapted from reference 61.

technological improvements, more sophisticated sampling of sequence rotamer conformations and energy scoring of protein-protein interactions and folds guided the experimental design of protein assemblies.^{53,54} Baker and coworkers used symmetrical proteins, scoured from an exhaustive search of the Protein Data Bank (PDB), to computationally redesign C_3 , C_4 and C_6 symmetric proteins at homomeric symmetric interfaces to generate extended 2D assemblies (**Figure 1.5c**).⁵⁵ In addition to 2D polymerization, interface redesign using Rosetta scoring and docking algorithms

were able to generate megadalton-scale one-⁵⁶ and two-component⁵⁷ protein cages, the latter of which was even electrostatically modified for RNA encapsulation *in vivo*.⁵⁸ Both the protein cage and extended 2D assemblies were capable of forming *in vivo* due to the large hydrophobic interfaces that facilitated tight protein association. In addition to the creation of hydrophobic patches, Baker and coworkers were also able to design elaborate hydrogen bonding networks to promote the formation of protein homo-oligomers.⁵⁹ Woolfson and colleagues used a balance of computational and rational design strategies for the *de novo* construction of homo- and heteromeric α -helical coiled-coil peptides, which were engineered to form 80-100 nm unilamellar spheres driven entirely by coiled-coil interactions (**Figure 1.5d**).^{60,61} Computational design enables the formation of energetically favorable protein-protein interactions but interfacial flexibility and reorganization, hallmarks of native protein complexes, remains an outstanding challenge in the field.

1.3.2 Metal-mediated protein self-assembly

Metal ions are vital for the proper function of biological systems. Metals can act as cofactors at interfacial sites in protein oligomers or serve crucial structural roles in maintaining a stable protein fold.^{62,63} Indeed, metal-ligand interactions are advantageous properties for use in protein design due to their reversibility, directional bonding and strength. Natively, proteins coordinate transition metal ions most commonly through the electron donating side chains of cysteine (Cys), histidine (His), aspartic acid (Asp) and glutamic acid (Glu) residues.⁶⁴ One strategy developed to take advantage of protein-metal interactions was developed in our lab, termed metal-directed protein self-assembly (MDPSA) in which metal-binding surface residues are installed onto a protein to assemble into an oligomer upon the addition of a transition metal ion.⁶⁵ In initial studies, MDPSA was used on the monomeric 4-helix bundle protein, cytochrome *cb*₅₆₂ by

installing a pair of bis-His clamps at i and $i+4$ positions along a α -helix to yield MBPC1.⁶⁶ MBPC1 was found to assemble into a homotetramer upon Zn^{2+} coordination and further characterization with other first-row transition metals confirmed our hypothesis that the oligomerization state can be dictated by the metal's preferred coordination geometry (**Figure 1.6a**).⁶⁷⁻⁶⁹ Computation design of the MBPC1 scaffold in a process termed metal-templated interface redesign (MeTIR) was employed to create the RIDC1 scaffold, which forms a noncovalent protein dimer in the absence of metal and maintains the ability to bind Zn^{2+} to form a tetramer.⁷⁰ The MeTIR strategy relies on using the metal-templated oligomer as a structural template to aid in computational interface design in lieu of designing desired protein packing from scratch. Further engineering efforts enabled the use of open coordination Zn^{2+} sites to facilitate the formation of 1D helical nanotubes and 2D crystalline arrays^{71,72}, structural and catalytic Zn^{2+} sites within the same scaffold to demonstrate hydrolytic activity on antimicrobial compounds^{73,74}, or strained intermolecular disulfide bonds coupled to Zn^{2+} -binding to create an allosteric assembly^{75,76} (**Figure 1.6b, c**). Judicious use of the MDPSA and MeTIR strategies created a structurally/functionally diverse set of proteins all from a single monomeric building block.

As with previous protein design strategies, symmetric proteins are attractive scaffolds to construct higher order assemblies. In our lab, bis-His clamps were installed on the tetrameric protein RhuA at its four corners (^{H63/H98}RhuA) to form crystalline lattices upon Zn^{2+} binding (**Figure 1.6d**).⁷⁷ An engineered variant of HuHF containing a His residue at its three-fold symmetry axes (^{H122}HuHF) can be loaded with first-row transition metals to create a cubic “metal node” similar to those employed in the construction of small-molecule metal-organic-frameworks (MOFs). Protein-based MOFs were generated by the addition of ditopic bridging linkers to metal-loaded ^{H122}HuHF predictably forming protein-MOFs with cubic or tetragonal symmetry by

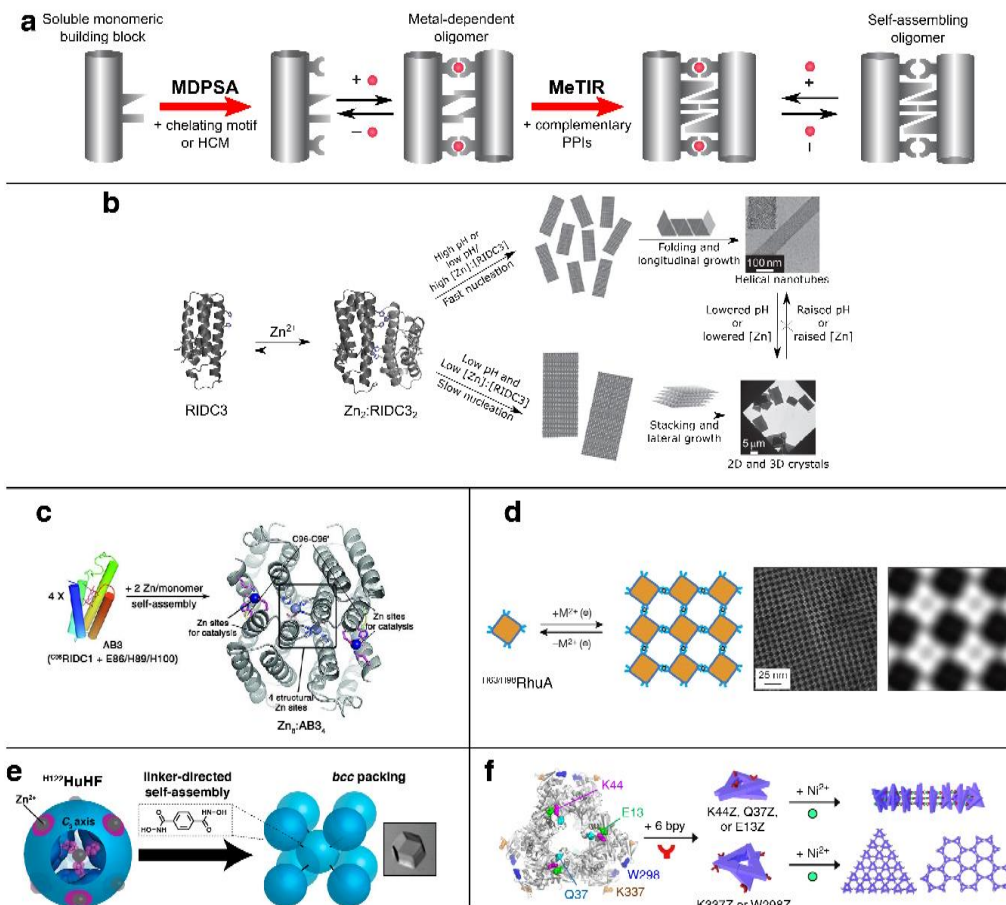


Figure 1.6 Strategies used in metal-mediated protein self-assembly. **(a)** Schematic for the generation of metal-templated protein assemblies. A monomeric building block installed with metal-chelating residues can self-assemble into a metal-dependent oligomer. Computational design of protein-protein interfaces within the oligomer can create a metal-independent protein architecture. Adapted from reference 69. **(b)** The strategy described in **(a)** is used for the generation of a monomer (RIDC3) that self-assembles into 1D helical nanotubes or extended 2D arrays upon incubation with excess Zn^{2+} . Adapted from reference 71. **(c)** A protein monomer (AB3) designed to bear structural and catalytic Zn-sites self-assembles into a tetrameric architecture capable of Zn-mediated hydrolysis of ampicillin. Adapted from reference 73. **(d)** The C_4 -symmetric RhuA is fashioned with metal-chelating motifs ($^{\text{H}63/\text{H}98}\text{RhuA}$) in order to assemble into 2D protein lattices upon incubation with metal. Adapted from reference 77. **(e)** The octameric human heavy-chain ferritin (HuHF) is engineered with a His residue at its three-fold axes of symmetry ($^{\text{H}122}\text{HuHF}$). The addition of a ditopic hydroxamate-based linker results in the metal-mediated assembly of HuHF crystals connected by the short linkers. Adapted from reference 78. **(f)** A bipyridine unnatural amino acid is installed onto the D_3 -symmetric acetyltransferase in a site-specific manner to enable the formation of 1D protein wires or 2D protein lattices upon the addition of Ni^{2+} . Adapted from reference 82.

modulating the linker or metal identity (**Figure 1.6e**).^{78,79} Marsh and coworkers genetically tethered a metal-dependent coiled-coil motif onto the C_3 -symmetric trimeric esterase for the metal-induced assembly of a protein cage.⁸⁰ The coiled-coil peptide motifs trimerized in the presence of Ni^{2+} to form discrete globular structures that contained 12 copies of each protein monomer. In a recent report, Heddle and coworkers used a 11-mer ring-like protein with a surface-exposed Cys that was triggered to self-assemble upon the addition of a gold-triphenylphosphine compound.⁸¹ The resultant snub cube consisted of 24 copies of the protein but interestingly, only 10 of the 11 possible thiols on each protein ring participated in Au-S interactions. This is an advantage of using reversible but strong chemical bonding (metal-protein interactions) that can tolerate the interfacial flexibility required to leave an uncoordinated Cys thiol. The aforementioned examples have all used metal-coordinating ligands native to proteins but advances in unnatural amino acid (UAA) incorporation can facilitate the use of multidentate or more strongly coordinating ligands to direct self-assembly. Song and coworkers incorporated a bipyridyl-alanine (bpy-Ala) residue onto the D_3 -symmetric hexamer acetyltransferase, which readily assembled into 1D protein wires or 2D lattices by varying the location of the bpy: top/bottom faces for 1D assembly and corners for 2D assembly (**Figure 1.6f**).⁸²

1.3.3 Disulfide-mediated protein self-assembly

Cysteine residues reversibly form covalent disulfide bonds in a redox-dependent manner and have biological relevance in stabilizing protein folds through intramolecular crosslinks and increasing tensile strength and stiffness of fibrous proteins like keratin.⁸³⁻⁸⁵ In the context of protein design, incorporation of disulfide bonds requires minimal genetic manipulation while permitting the use of a reversible covalent linkage within a self-assembly scaffold. One example of directing self-assembly via disulfide bond formation is a report from Mougous and colleagues in which two

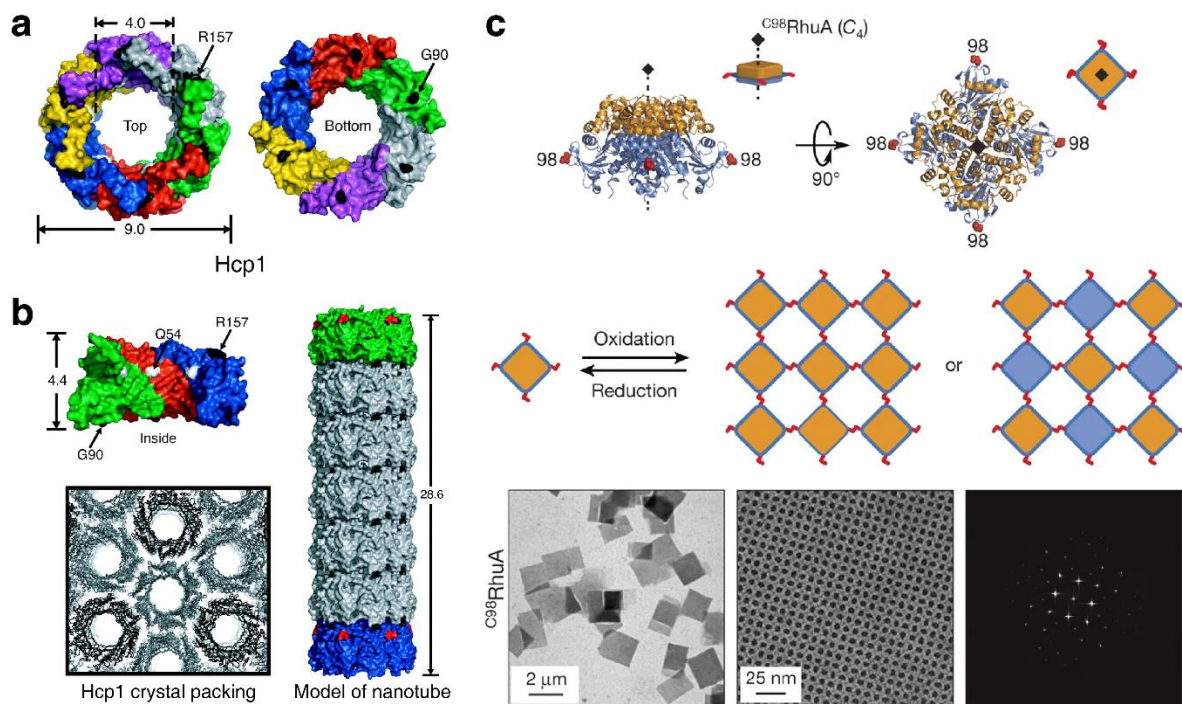


Figure 1.7 Disulfide-mediated protein assemblies. **(a)** The hexameric protein Hcp1 is modified with Cys residues at position 90 and 157. A close inspection of the crystal packing of Hcp1 **(b)** reveals long channels. Protein nanotube could be formed by engineering protein association at the top and bottom faces of Hcp1. The model of a protein nanotube is shown of Hcp1 proteins that could be connected by oxidation of Cys-bearing C^{90}/C^{157} Hcp1. Adapted from reference 86. **(c)** A single Cys residue is installed on the corners of the C_4 -symmetric protein RhuA. Oxidation of C^{98} RhuA in solution forms large 2D protein crystals connected through disulfide linkages. Adapted from reference 77.

Cys residues were installed on opposite faces of the hexameric protein Hcp1 (**Fig. 1.7**).⁸⁶ A close inspection of the crystal packing of Hcp1 hexamers suggested that the ring-ring interface could be stabilized through the use of disulfide bonds, and indeed protein oxidation to promote disulfide bond formation resulted in 1D nanotubes. In this case, structure-guided design aided the placement of Cys residues but disulfide linkages could also be rationally engineered to promote 2D and 3D crystallization.^{77,87} In our lab, an engineered variant of the tetrameric RhuA containing a Cys at its corners (C^{98} RhuA) assembled into μm -sized 2D crystalline lattices under controlled oxidation conditions (**Fig. 1.7**).⁷⁷ Rotational degrees of freedom at the flexible disulfide linkages allowed

RhuA tetramers to coherently rotate within the lattice to create a dynamically porous crystalline framework. In fact, ^{C98}RhuA lattices exhibit a maximal Poisson's ratio of -1 and are auxetic (*i.e.* longitudinal expansion upon transverse stretching), a unique property among self-assembled materials.

1.4 Developing hybrid protein-based materials

A key aspect of protein design is the development of functional assemblies towards potential applications in sensing, catalysis, drug delivery, nanomedicine, therapeutic devices or as molecular display scaffolds.⁸⁸ Considerable efforts have been made to use crystalline S-layers, which natively envelop bacterial and archaeal cells, have been targeted as scaffolds for the bottom-up assembly of functional structures.⁸⁹ Though atomic resolution characterization of S-layer proteins and interactions is limited to a few structures^{90,91}, S-layer proteins can be harvested from bacteria and reassembled using a variety of methods (*e.g.* in solution, on liposomes, at air-water or lipid interfaces, and on solid silicon substrates) enabling their manipulation for further engineering (**Figure 1.8**).⁸ Sletyr and colleagues (among others) performed pioneering research on the study of S-layer protein function and exploring potential applications. For instance, the deposition of S-layer fragments onto a microfiltration membrane showed molecular weight cutoffs at 30-40 kDa show the potential for using S-layer lattices as part of a specialized filtration membrane.^{92,93} Reconstituting S-layer proteins *in vitro* enabled the generation of fusion proteins that used the surface of the S-layer lattice as a molecular display for antigens (towards vaccine development)^{94,95}, streptavidin (to interface with biotinylated compounds for sensing)^{96,97}, or fluorescent proteins for use as biomarkers^{98,99}. More recently, a handful of reports have utilized S-layer lattices as molecular displays through the use of the split-protein SpyTag-SpyCatcher system which forms an irreversible isopeptide bond between the 45-mer SpyTag and its partner protein,

SpyCatcher. In one example, *Caulobacter crescentus* cells displaying engineered S-layer lattices with SpyTag can interact with SpyCatcher partners modified with fluorescent proteins or quantum dot nanoparticles can be selectively and uniformly displayed on the surface (**Fig. 1.8**).¹⁰⁰

Designed protein assemblies have also been further engineered as molecular displays. Metal-mediated 2D RIDC3 crystals were modified with a Zn-porphyrin instead of the native heme to induce the formation of Pt nanoparticles upon light irradiation (**Fig. 1.8**).¹⁰¹ Wang and coworkers reported the use of 2D ^{C98}RhuA lattices as a template for patterning 5 nm gold nanoparticles (AuNPs). The introduction of a second Cys residue (Cys4) on one face of ^{C98}RhuA enabled the precise binding of AuNPs on each tetramer (**Fig. 1.8**).¹⁰² Baneyx and colleagues used a computationally redesigned hexameric TTM scaffold, which forms extended 2D arrays upon Ca²⁺ binding, as a scaffold for the display of polymeric beads, AuNPs and avidin by genetically encoding surface functionalities that promoted ligand binding.¹⁰³ Engineered protein cages are also attractive targets for generating functional assemblies due to the appeal of encapsulation within the closed architecture. *De novo* designed icosahedral cages have been engineered with a positive interior and subsequently placed through directed evolution experiments to generate a cage that captured its native genome, akin to many viral capsids.⁵⁸ Hilvert and coworkers also demonstrated that engineering an existing protein cage, lumazine synthase, with cationic peptides and subsequent evolutionary optimization of the scaffold enabled the encapsulation and protection of the RNA genome *in vivo*.¹⁰⁴ While these are inspiring examples for the potential uses of native and designed protein assemblies, considerable challenges and opportunities remain in the pursuit of dynamic multicomponent architectures that rival native protein complexes.

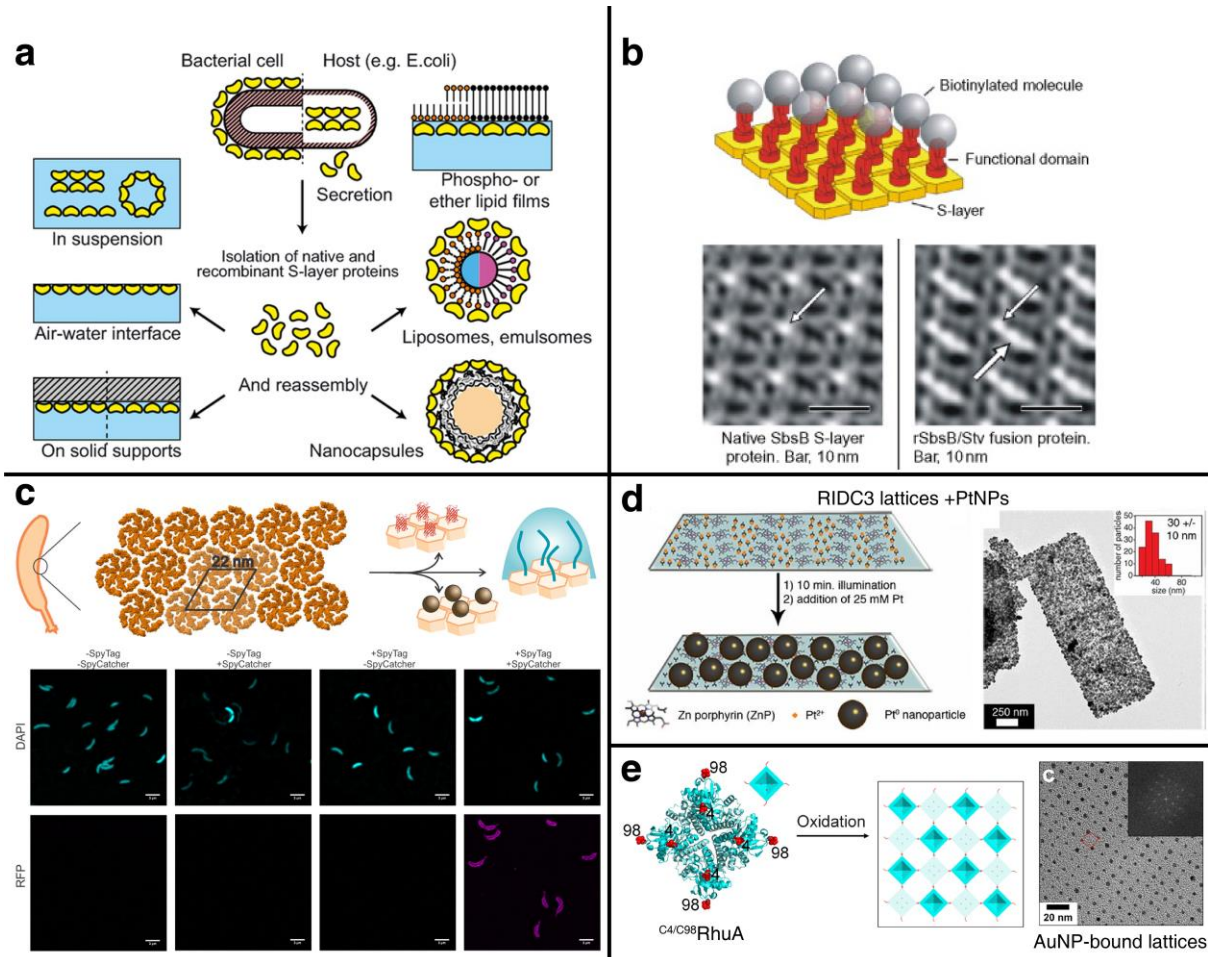


Figure 1.8 Using 2D protein lattices to generate hybrid biomaterials. **(a)** Schematic showing the reassembly of isolated S-layer proteins using different strategies. Adapted from reference 8. **(b)** Cartoon showing self-assembled S-layer fusion domains with streptavidin. 2D reconstructions from TEM micrographs of the native and fusion protein lattices are shown below. Adapted from reference 8. **(c)** A crystalline s-layer lattice of *Caulobacter crescentus* is genetically modified with the split-protein SpyTag. Incubation of cells bearing the engineered S-layer proteins with a fluorescent protein or nanoparticle-bearing SpyCatcher creates in the covalent attachment of the partner to cell surfaces. Adapted from reference 100. **(d)** Metal-mediated 2D RIDC3 lattices bearing a Zn-porphyrin instead of the native heme (Fe-porphyrin) are incubated with Pt^{2+} . Light-mediated reduction of Pt^{2+} through the Zn-porphyrins results in the formation of Pt nanoparticles on crystal surfaces. Adapted from reference 101. **(e)** C_4 -symmetric C^{98} RhuA is fashioned with a second Cys residue at the N-terminal face to facilitate the association of Au nanoparticles onto 2D C_4/C^{98} RhuA lattices. Adapted from reference 102.

1.5 Dissertation objectives

The primary goals of this thesis are twofold:

1. Develop functional protein materials by using enzymes to selectively modify crystalline 2D protein lattices
2. Create novel self-assembled architectures that integrate biologically relevant interaction motifs into a protein scaffold

The use of reversible but strong interactions such as metal coordination and disulfide bonding has enabled the construction of sophisticated protein assemblies with emergent functions (*e.g.* templated growth of nanoparticles, Zn^{2+} -catalyzed hydrolysis, and one-of-a-kind dynamic crystalline lattices) in our lab. There is great interest in developing novel strategies for creating functional protein assemblies, building upon many of the methods discussed above. We aim to show that existing crystalline protein lattices can be readily modified using enzymes as a generalizable strategy for generating functionalized arrays. We also demonstrate how cooperative assembly of nucleic acid and protein elements can be used to create a novel artificial nucleoprotein architecture. These assemblies are one of the only hybrid protein-NA materials that do not solely rely on programmable nucleic acid interactions to drive assembly. Lastly, we integrate a biological chelating group, hydroxamate, into a protein scaffold equipped to bind metal ions to generate a bimetallic protein cage that can readily assemble and disassemble in response to multiple stimuli.

1.6 References

- (1) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551-575.
- (2) Goodsell, D. S.; Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Struct.* **2000**, *29*, 105-153.
- (3) Rutledge, H. L.; Tezcan, F. A. Electron Transfer in Nitrogenase. *Chem. Rev.* **2020**.
- (4) Barber, J. Photosystem II: the engine of life. *Q. Rev. Biophys.* **2003**, *36* (1), 71-89.
- (5) Junge, W.; Nelson, N. ATP Synthase. *Annu. Rev. Biochem.* **2015**, *84* (1), 631-657.
- (6) Groll, M.; Ditzel, L.; Löwe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D.; Huber, R. Structure of 20S proteasome from yeast at 2.4Å resolution. *Nature* **1997**, *386* (6624), 463-471.
- (7) Hess, H.; Ross, J. L. Non-equilibrium assembly of microtubules: from molecules to autonomous chemical robots. *Chem. Soc. Rev.* **2017**, *46* (18), 5570-5587.
- (8) Sleytr, U. B.; Schuster, B.; Egelseer, E. M.; Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **2014**, *38* (5), 823-864.
- (9) Eitoku, M.; Sato, L.; Senda, T.; Horikoshi, M. Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. *Cell. Mol. Life. Sci.* **2008**, *65* (3), 414-444.
- (10) Bol, J. F. Role of capsid proteins. *Methods Mol. Biol.* **2008**, *451*, 21-31.
- (11) Antonietti, M.; Förster, S. Vesicles and liposomes: a self-assembly principle beyond lipids. *Adv. Mater.* **2003**, *15* (16), 1323-1333.
- (12) Richard, C.; Balavoine, F.; Schultz, P.; Ebbesen, T. W.; Mioskowski, C. Supramolecular self-assembly of lipid derivatives on carbon nanotubes. *Science* **2003**, *300* (5620), 775-778.
- (13) Seeman, N. C.; Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068.
- (14) Hong, F.; Zhang, F.; Liu, Y.; Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **2017**, *117* (20), 12584-12640.
- (15) Woehlke, G.; Schliwa, M. Walking on two heads: The many talents of kinesin. **2000**, *1* (1), 50-58.
- (16) Sinclair, J. C.; Davies, K. M.; Venien-Bryan, C.; Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **2011**, *6* (9), 558-562.

- (17) 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. USA* **2001**, *98* (5), 2217-2221.
- (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.
- (19) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andrzejczyk, 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.
- (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**.
- (21) 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. USA* **2016**, *113* (31), 8681-8686.
- (22) 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.
- (23) Cecioni, S.; Imberty, A.; Vidal, S. Glycomimetics versus Multivalent Glycoconjugates for the Design of High Affinity Lectin Ligands. *Chem. Rev.* **2015**, *115* (1), 525-561.
- (24) Rini, J. M. Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 551-577.
- (25) Dotan, N.; Arad, D.; Frolow, F.; Freeman, A. Self-Assembly of a Tetrahedral Lectin into Predesigned Diamondlike Protein Crystals. *Angew. Chem. Int. Ed. Engl.* **1999**, *38* (16), 2363-2366.
- (26) Kitagishi, H.; Oohora, K.; Yamaguchi, H.; Sato, H.; Matsuo, T.; Harada, A.; Hayashi, T. Supramolecular hemoprotein linear assembly by successive interprotein heme-heme pocket interactions. *J. Am. Chem. Soc.* **2007**, *129* (34), 10326-10327.
- (27) Kitagishi, H.; Kakikura, Y.; Yamaguchi, H.; Oohora, K.; Harada, A.; Hayashi, T. Self-assembly of one- and two-dimensional hemoprotein systems by polymerization through heme-heme pocket interactions. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (7), 1271-1274.
- (28) Oohora, K.; Burazerovic, S.; Onoda, A.; Wilson, Y. M.; Ward, T. R.; Hayashi, T. Chemically programmed supramolecular assembly of hemoprotein and streptavidin with alternating alignment. *Angew. Chem. Int. Ed. Engl.* **2012**, *124* (16), 3884-3887.
- (29) Oohora, K.; Onoda, A.; Hayashi, T. Supramolecular assembling systems formed by heme-heme pocket interactions in hemoproteins. *Chem. Commun.* **2012**, *48* (96), 11714-11726.

- (30) Weber, P.; Ohlendorf, D.; Wendoloski, J.; Salemme, F. Structural origins of high-affinity biotin binding to streptavidin. *Science* **1989**, *243* (4887), 85-88.
- (31) Ringler, P.; Schulz, G. E. Self-assembly of proteins into designed networks. *Science* **2003**, *302* (5642), 106-109.
- (32) Macfarlane, R. J.; Lee, B.; Jones, M. R.; Harris, N.; Schatz, G. C.; Mirkin, C. A. Nanoparticle superlattice engineering with DNA. *Science* **2011**, *334* (6053), 204-208.
- (33) Nykypanchuk, D.; Maye, M. M.; van der Lelie, D.; Gang, O. DNA-guided crystallization of colloidal nanoparticles. *Nature* **2008**, *451* (7178), 549-552.
- (34) Brodin, J. D.; Auyeung, E.; Mirkin, C. A. DNA-mediated engineering of multicomponent enzyme crystals. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (15), 4564-4569.
- (35) McMillan, J. R.; Brodin, J. D.; Millan, J. A.; Lee, B.; Olvera de la Cruz, M.; Mirkin, C. A. Modulating Nanoparticle Superlattice Structure Using Proteins with Tunable Bond Distributions. *J. Am. Chem. Soc.* **2017**, *139* (5), 1754-1757.
- (36) McMillan, J. R.; Mirkin, C. A. DNA-Functionalized, Bivalent Proteins. *J. Am. Chem. Soc.* **2018**, *140* (22), 6776-6779.
- (37) Kashiwagi, D.; Sim, S.; Niwa, T.; Taguchi, H.; Aida, T. Protein Nanotube Selectively Cleavable with DNA: Supramolecular Polymerization of “DNA-Appended Molecular Chaperones”. *J. Am. Chem. Soc.* **2018**, *140* (1), 26-29.
- (38) Mou, Y.; Huang, P. S.; Hsu, F. C.; Huang, S. J.; Mayo, S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (34), 10714-10719.
- (39) Mou, Y.; Yu, J.-Y.; Wannier, T. M.; Guo, C.-L.; Mayo, S. L. Computational design of co-assembling protein–DNA nanowires. *Nature* **2015**, *525* (7568), 230-233.
- (40) Jones, S.; Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **1996**, *93* (1), 13-20.
- (41) Kortemme, T.; Baker, D. Computational design of protein-protein interactions. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 91-97.
- (42) Kostianen, M. A.; Hiekkataipale, P.; Laiho, A.; Lemieux, V.; Seitsonen, J.; Ruokolainen, J.; Ceci, P. Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nat. Nanotech.* **2013**, *8* (1), 52-56.
- (43) Liljeström, V.; Mikkilä, J.; Kostianen, M. A. Self-assembly and modular functionalization of three-dimensional crystals from oppositely charged proteins. *Nat. Comm.* **2014**, *5*, doi:10.1038/ncomms5445.
- (44) Künzle, M.; Eckert, T.; Beck, T. Binary Protein Crystals for the Assembly of Inorganic Nanoparticle Superlattices. *J. Am. Chem. Soc.* **2016**, *138* (39), 12731-12734.

- (45) Simon, A. J.; Zhou, Y.; Ramasubramani, V.; Glaser, J.; Pothukuchy, A.; Gollihar, J.; Gerberich, J. C.; Leggere, J. C.; Morrow, B. R.; Jung, C. et al. Supercharging enables organized assembly of synthetic biomolecules. *Nat. Chem.* **2019**, *11* (3), 204-212.
- (46) Bolon, D. N.; Grant, R. A.; Baker, T. A.; Sauer, R. T. Specificity versus stability in computational protein design. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (36), 12724-12729.
- (47) Vizcarra, C. L.; Mayo, S. L. Electrostatics in computational protein design. *Curr. Opin. Chem. Biol.* **2005**, *9* (6), 622-626.
- (48) Mandell, D. J.; Kortemme, T. Computer-aided design of functional protein interactions. *Nature Chem. Biol.* **2009**, *5* (11), 797-807.
- (49) Norn, C. H.; André, I. Computational design of protein self-assembly. *Curr. Opin. Struct. Biol.* **2016**, *39* (Supplement C), 39-45.
- (50) Gainza-Cirauqui, P.; Correia, B. E. Computational protein design—the next generation tool to expand synthetic biology applications. *Curr. Opin. Biotech.* **2018**, *52*, 145-152.
- (51) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. High-Resolution Protein Design with Backbone Freedom. *Science* **1998**, *282* (5393), 1462-1467.
- (52) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331* (1), 281-299.
- (53) Liu, Y.; Kuhlman, B. RosettaDesign server for protein design. *Nucl. Acids. Res.* **2006**, *34*, W235-238.
- (54) Parmeggiani, F.; Huang, P. S.; Vorobiev, S.; Xiao, R.; Park, K.; Caprari, S.; Su, M.; Seetharaman, J.; Mao, L.; Janjua, H. et al. A general computational approach for repeat protein design. *J. Mol. Biol.* **2015**, *427* (2), 563-575.
- (55) 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.
- (56) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R. et al. Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* **2016**, *535* (7610), 136-139.
- (57) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **2016**, *353* (6297), 389-394.
- (58) 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. et al. Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **2017**, *552* (7685), 415-420.
- (59) Boyken, S. E.; Chen, Z.; Groves, B.; Langan, R. A.; Oberdorfer, G.; Ford, A.; Gilmore, J. M.; Xu, C.; DiMaio, F.; Pereira, J. H. et al. De novo design of protein homo-oligomers with

- modular hydrogen-bond network-mediated specificity. *Science* **2016**, 352 (6286), 680-687.
- (60) 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. et al. 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.
- (61) 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. et al. Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* **2013**, 340 (6132), 595-599.
- (62) Sontz, P. A.; Song, W. J.; Tezcan, F. A. Interfacial metal coordination in engineered protein and peptide assemblies. *Curr. Opin. Chem. Biol.* **2014**, 19, 42-49.
- (63) Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A. Metals in Protein-Protein Interfaces. *Ann. Rev. Biophys.* **2014**, 43, 409-431
- (64) Tainer, J. A.; Roberts, V. A.; Getzoff, E. D. Protein metal-binding sites. *Curr. Opin. Biotech.* **1992**, 3 (4), 378-387.
- (65) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **2010**, 43 (5), 661-672.
- (66) Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A. Controlling Protein-Protein Interactions through Metal Coordination: Assembly of a 16-Helix Bundle Protein. *J. Am. Chem. Soc.* **2007**, 129 (44), 13374-13375.
- (67) Salgado, E. N.; Lewis, R. A.; Mossin, S.; Rheingold, A. L.; Tezcan, F. A. Control of Protein Oligomerization Symmetry by Metal Coordination: C₂ and C₃ Symmetrical Assemblies through Cu^{II} and Ni^{II} Coordination. *Inorg. Chem.* **2009**, 48 (7), 2726-2728.
- (68) Salgado, E. N.; Lewis, R. A.; Faraone-Mennella, J.; Tezcan, F. A. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* **2008**, 130 (19), 6082-6084.
- (69) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. In *Methods in Enzymology*; Pecoraro, V. L., Ed.; Academic Press, 2016; Vol. 580.
- (70) Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A. Metal-Templated Design of Protein Interfaces. *Proc. Natl. Acad. Sci. USA* **2010**, 107, 1827-1832.
- (71) Brodin, J. D.; Ambroggio, X. I.; Tang, C. Y.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **2012**, 4 (5), 375-382.
- (72) Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A. Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **2015**, 137 (33), 10468-10471.

- (73) Song, W. J.; Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014**, *346* (6216), 1525-1528.
- (74) Song, W. J.; Yu, J.; Tezcan, F. A. Importance of Scaffold Flexibility/Rigidity in the Design and Directed Evolution of Artificial Metallo- β -lactamases. *J. Am. Chem. Soc.* **2017**, *139* (46), 16772-16779.
- (75) Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A. De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J. Am. Chem. Soc.* **2016**, *138* (40), 13163-13166.
- (76) Churchfield, L. A.; Alberstein, R. G.; Williamson, L. M.; Tezcan, F. A. Determining the Structural and Energetic Basis of Allostery in a De Novo Designed Metalloprotein Assembly. *J. Am. Chem. Soc.* **2018**, *140* (31), 10043-10053.
- (77) 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*, 369.
- (78) Sontz, P. A.; Bailey, J. B.; Ahn, S.; Tezcan, F. A. A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals. *J. Am. Chem. Soc.* **2015**, *137* (36), 11598-11601.
- (79) Bailey, J. B.; Zhang, L.; Chiong, J. A.; Ahn, S.; Tezcan, F. A. Synthetic Modularity of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2017**, *139* (24), 8160-8166.
- (80) Cristie-David, A. S.; Marsh, E. N. G. Metal-dependent assembly of a protein nano-cage. *Protein Sci.* **2019**, *28* (9), 1620-1629.
- (81) 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. et al. In *Nature*, 2019.
- (82) 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.
- (83) Tidor, B.; Karplus, M. The contribution of cross-links to protein stability: A normal mode analysis of the configurational entropy of the native state. *Proteins* **1993**, *15* (1), 71-79.
- (84) Mamathambika, B. S.; Bardwell, J. C. Disulfide-Linked Protein Folding Pathways. *Ann. Rev. Cell Dev. Bio.* **2008**, *24* (1), 211-235.
- (85) Bruce Fraser, R. D.; Parry, D. A. D. The role of disulfide bond formation in the structural transition observed in the intermediate filaments of developing hair. *J. Struct. Biol.* **2012**, *180* (1), 117-124.
- (86) Ballister, E. R.; Lai, A. H.; Zuckermann, R. N.; Cheng, Y.; Mougous, J. D. In vitro self-assembly from a simple protein of tailorable nanotubes building block. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (10), 3733-3738.

- (87) Banatao, D. R.; Cascio, D.; Crowley, C. S.; Fleissner, M. R.; Tienson, H. L.; Yeates, T. O. An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (44), 16230-16235.
- (88) Baneyx, F.; Matthaei, J. F. Self-assembled two-dimensional protein arrays in bionanotechnology: from S-layers to designed lattices. *Curr. Opin. Biotech.* **2014**, *28*, 39-45.
- (89) Fagan, R. P.; Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* **2014**, *12* (3), 211-222.
- (90) Baranova, E.; Fronzes, R.; Garcia-Pino, A.; Van Gerven, N.; Papapostolou, D.; Pehau-Arnaudet, G.; Pardon, E.; Steyaert, J.; Howorka, S.; Remaut, H. SbsB structure and lattice reconstruction unveil Ca²⁺ triggered S-layer assembly. *Nature* **2012**, *487* (7405), 119-122.
- (91) Bharat, T. A. M.; Kureisaite-Ciziene, D.; Hardy, G. G.; Yu, E. W.; Devant, J. M.; Hagen, W. J. H.; Brun, Y. V.; Briggs, J. A. G.; Lowe, J. Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. *Nat. Microbiol.* **2017**, *2*, 17059.
- (92) Sára, M.; Sleytr, U. B. Molecular sieving through S layers of *Bacillus stearothermophilus* strains. *J. Bacteriol.* **1987**, *169* (9), 4092-4098.
- (93) Sára, M.; Manigley, C.; Wolf, G.; Sleytr, U. B. Isoporous ultrafiltration membranes from bacterial cell envelope layers. *J. Membr. Sci.* **1988**, *36*, 179-186.
- (94) Farr, C.; Nomellini, J. F.; Ailon, E.; Shanina, I.; Sangsari, S.; Cavacini, L. A.; Smit, J.; Horwitz, M. S. Development of an HIV-1 Microbicide Based on *Caulobacter crescentus*: Blocking Infection by High-Density Display of Virus Entry Inhibitors. *PLoS One* **2013**, *8* (6), e65965.
- (95) Umelo-Njaka, E.; Nomellini, J. F.; Bingle, W. H.; Glasier, L. G.; Irvin, R. T.; Smit, J. Expression and testing of *Pseudomonas aeruginosa* vaccine candidate proteins prepared with the *Caulobacter crescentus* S-layer protein expression system. *Vaccine* **2001**, *19* (11-12), 1406-1415.
- (96) Moll, D.; Huber, C.; Schlegel, B.; Pum, D.; Sleytr, U. B.; Sara, M. S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (23), 14646-14651.
- (97) Huber, C.; Liu, J.; Egelseer, E. M.; Moll, D.; Knoll, W.; Sleytr, U. B.; Sara, M. Heterotetramers formed by an S-layer-streptavidin fusion protein and core-streptavidin as a nanoarrayed template for biochip development. *Small* **2006**, *2* (1), 142-150.
- (98) Kainz, B.; Steiner, K.; Moller, M.; Pum, D.; Schaffer, C.; Sleytr, U. B.; Toca-Herrera, J. L. Absorption, steady-state fluorescence, fluorescence lifetime, and 2D self-assembly properties of engineered fluorescent S-layer fusion proteins of *Geobacillus stearothermophilus* NRS 2004/3a. *Biomacromolecules* **2010**, *11* (1), 207-214.

- (99) Kainz, B.; Steiner, K.; Sleytr, U. B.; Pum, D.; Toca-Herrera, J. L. Fluorescence energy transfer in the bi-fluorescent S-layer tandem fusion protein ECFP-SgsE-YFP. *J. Struct. Biol.* **2010**, *172* (3), 276-283.
- (100) Charrier, M.; Li, D.; Mann, V. R.; Yun, L.; Jani, S.; Rad, B.; Cohen, B. E.; Ashby, P. D.; Ryan, K. R.; Ajo-Franklin, C. M. Engineering the S-Layer of *Caulobacter crescentus* as a Foundation for Stable, High-Density, 2D Living Materials. *ACS Synth. Biol.* **2019**, *8* (1), 181-190.
- (101) Brodin, J. D.; Carr, J. R.; Sontz, P. A.; Tezcan, F. A. Exceptionally stable, redox-active supramolecular protein assemblies with emergent properties. *Proc. Natl. Acad. Sci. USA* **2014**, *111* (8), 2897-2902.
- (102) Du, M.; Zhou, K.; Wang, X.; Zhang, J.; Zhang, Y.; Dong, J.; Wu, L.; Qiao, Z.; Chen, G.; Wang, Q. Precise Fabrication of De Novo Nanoparticle Lattices on Dynamic 2D Protein Crystalline Lattices. *Nano Lett.* **2020**, *20* (2), 1154-1160.
- (103) Thomas, A.; Matthaei, J. F.; Baneyx, F. A Self-Assembling Two-Dimensional Protein Array is a Versatile Platform for the Assembly of Multicomponent Nanostructures. *Biotechnol. J.* **2018**, *13* (12), 1800141.
- (104) Terasaka, N.; Azuma, Y.; Hilvert, D. Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (21), 5432-5437.

Chapter 2: Enzyme-directed Modification of Two-Dimensional Crystalline Protein Lattices

2.1 Abstract

The design and construction of crystalline protein arrays to selectively assemble ordered nanoscale materials has potential applications in sensing, catalysis and medicine. Whereas a variety of design strategies have been implemented for the bottom-up construction of novel protein assemblies, the generation of functional materials is generally limited to the capabilities of the attendant protein building blocks. Enzyme-directed post-translational modifications are responsible for the functional diversity of the proteome and thus, can be harnessed to manipulate protein-based biomaterials. In this study, we describe the use of phosphopantetheinyl transferases (PPTases), a class of enzymes that covalently modify proteins using CoA substrates, to tailor the surface of 2D protein arrays. A short peptide surrogate for Sfp PPTase, ybbR, is covalently tethered to 2D crystalline platform to facilitate enzymatic functionalization. Our results show that substrate promiscuity of PPTases allows for the site-specific modification of two different self-assembling protein crystals with both small molecule fluorophores and proteins. This work highlights the potential for using enzyme-mediated modification of large protein surfaces towards the generation of sophisticated protein platforms reminiscent of the complex landscape of the extracellular matrix.

2.2 Introduction

Proteins are Nature's premier building blocks for the construction of supramolecular assemblies to fulfill a myriad of cellular processes necessary for sustained life.^{1,2} Inspired by the diversity of native protein structures, the hierarchical construction of intricate biological machines to perform complex chemical processes remains an outstanding goal of bio-nanotechnology.³⁻⁶ Whereas the self-assembly of multicomponent protein architectures is based on a complex

interplay of non-covalent protein-protein interactions in the cellular environment, the structural and chemical diversity of the proteome is largely enhanced through enzyme-mediated post-translational modifications.⁷ Such modifications play vital roles in the biosynthesis of complex biomolecules (*e.g.* fatty acids, peptides, polyketides)⁸ or in large-scale cellular functions such as cell-cell signaling and recognition (*e.g.* via glycosylation or membrane anchoring through the attachment of lipid tails)⁹. Bacterial and archaeal S-layers, prototypical two-dimensional self-assembling protein lattices, perform critical roles in cell protection, virulence and surface recognition through post-translational modifications.¹⁰ Developing protein-based biomaterials amenable to post-translational modifications presents potential applications in diagnostic sensors, vaccines, drug-delivery, or biomineralization matrices.^{11,12}

In particular, there is broad interest in the use of two-dimensional (2D) protein arrays as a biotechnological platform due to their high-density display of polypeptides with nanoscale tunability and reconfigurability.¹³ Engineering the surface of crystalline protein materials provides an opportunity to selectively organize target molecules of interest in a site-dependent manner. Over the years, various chemical biology tools have been employed for the immobilization of biomolecules on solid substrates (*e.g.* adsorption into a gel matrix, biotinylation, azide-alkyne cycloaddition).^{14,15} In particular, enzymatic processes naturally afford selective labeling of biological architectures while maintaining efficiency and specificity, making them particularly attractive for developing functional biomaterials. Proteins have been employed for directing protein-protein ligation or surface immobilization in aqueous environments (*e.g.* Sortase¹⁶, split-intein¹⁷, SpyTag-SpyCatcher¹⁸); for instance, the transpeptidation reaction of *Staphylococcus aureus*, Sortase A has been used in numerous instances to demonstrate the modification of polymer or agarose beads, gold surfaces, molecular films and even tailoring the surfaces of cells.¹⁹⁻²¹

Sleytr and coworkers have shown that S-layer lattices can be reconstituted *in vitro* for applications in membrane filtration, drug delivery, and spatial organization of immunogenic biomolecules or inorganic nanoparticles.¹² Two recent reports describe the genetic incorporation of peptide tags onto crystalline S-layer substrates using the split-protein system SpyTag-SpyCatcher¹⁸, which forms an irreversible isopeptide bond between the SpyCatcher protein and SpyTag peptide, to generate hierarchical 3D materials and high-density displays on living cells.^{22,23} However, given the dearth of atomic resolution structures of S-layer proteins, it remains challenging to engineer 2D S-layer lattices into functional materials. There have been considerable protein design efforts for the construction of artificial 2D lattices using a variety of strategies, ranging from metal coordination²⁴, reversible covalent bonds²⁵, noncovalent hydrophobic contacts²⁶ and gene fusions^{27,28}. A handful of recent reports demonstrate the use of artificial 2D protein arrays as a molecular display platform.^{29,30}

2.3 Results and Discussion

2.3.1 Design and characterization of chemically modified RIDC3+ybbR arrays

In this work, we set out to develop a crystalline framework whose surfaces can be tunably modified to display biomolecules. In designing our system, we wanted to ensure that (i) the incorporation of a functional peptide handle would not disrupt the underlying lattice packing while allowing for (ii) site-specific modification with (iii) facile genetic encodability and (iv) broad substrate specificity. We chose an engineered variant of cytochrome *cb₅₆₂*, RIDC3, which can readily assemble into 1-, 2- and 3D protein lattices via Zn²⁺ coordination as our crystalline protein substrate.²⁴ Self-assembled RIDC3 arrays have been previously shown to tolerate a broad pH, temperature and solvent range, providing a robust platform for the surface display of biomolecules.³¹ We developed two strategies for the incorporation of a functional peptide handle

onto RIDC3 crystals (**Figure 2.1a, b**): a chemical modification of pre-formed crystals with the peptide and the genetic incorporation of the peptide onto RIDC3. In order to tunably and specifically modify array surfaces, we use the post-translational modifying phosphopantetheinyl transferase (PPTase) as our enzyme of choice. PPTases play a crucial role in various biosynthetic pathways through the covalent modification of acyl carrier proteins (ACPs) at a conserved serine residue via the transfer of phosphopantetheine (Ppant) from coenzyme A (CoA).⁸ Walsh and coworkers used phage display to discover a 11 amino acid peptide (ybbR: DSLEFIASKLA) that acts as a minimal recognition sequence for PPTases and can be used as a short peptide tag for site-specific protein labeling at the first serine residue (**Figure 2.1c, d**).^{32,33} While CoA is the biological

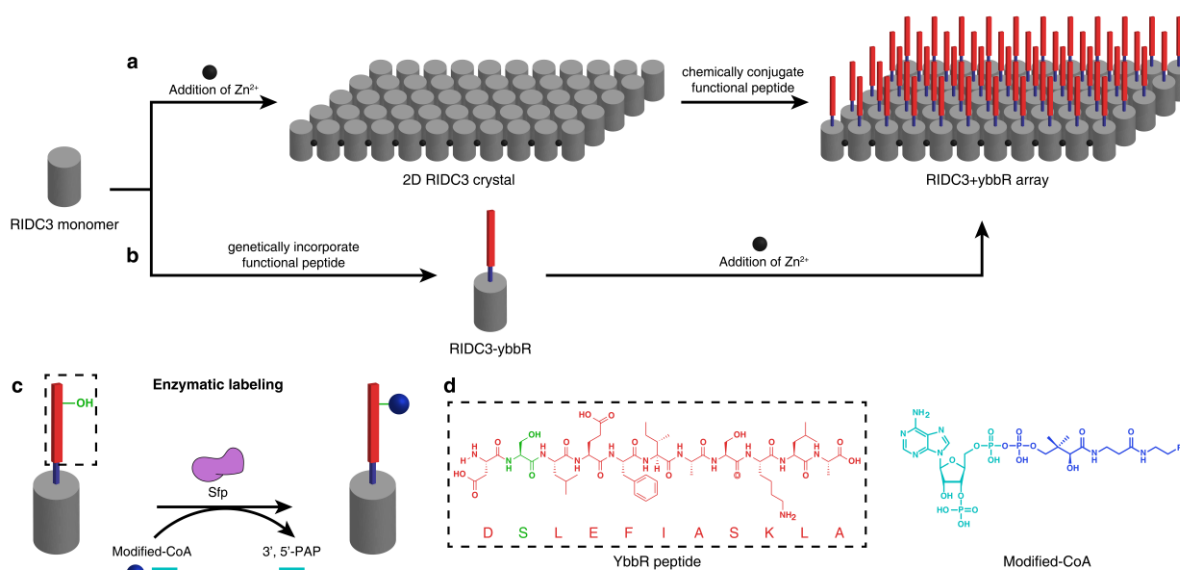


Figure 2.1 Schematic of ybbR addition onto RIDC3 for enzymatic labeling. The chemical conjugation (**a**) or genetic incorporation (**b**) of a functional peptide is shown. (**a**) RIDC3 self-assembles into crystalline 2D lattices upon the addition of Zn²⁺ and be further modified via chemical conjugation of a functional peptide (ybbR) to yield a RIDC3+ybbR array. (**b**) RIDC3 can be genetically modified to express as the RIDC3-ybbR construct, which can self-assemble to form peptide-bearing 2D crystals. (**c**) The ybbR peptide contains a serine residue that can be site-specifically modified using the PPTase Sfp with modified-CoA substrates. (**d**) The sequence for ybbR and a modified-CoA are shown color coded to match the cartoon representation in (c).

substrate for the PPTase used in our study, Surfactin phosphopantetheinyl transferase (Sfp) from *Bacillus subtilis*, previous work has shown Sfp can recognize a broad range of substrates and can enzymatically transfer different sets of biomolecules tethered to CoA in a site-specific manner to a serine residue of ACP or ybbR (**Figure 2.1c, d**).³⁴⁻³⁷ The promiscuity of Sfp for a range of modified-CoA substrates provides a model system to explore the enzymatic surface functionalization of our crystalline protein materials.

We first chemically modified the surface of RIDC3 arrays with a synthesized ybbR peptide in order to ensure that crystallinity was retained upon the addition of peptide and chemoenzymatic labeling of ybbR remained possible on the surface. A close examination of the RIDC3 crystal packing (PDB ID: 3TOM) revealed up to 7 surface-exposed lysine residues, which we targeted for modification. Suspensions of pre-formed RIDC3 crystals were treated with a 10-fold excess of a dibenzocyclooctyne-N-hydroxysuccinimidyl ester (DBCO-NHS) to modify surface-exposed lysine residues (**Figure 2.2a**).³⁸ The strained cyclooctyne, DBCO, was used to avoid using Cu(I) with Zn-coordinating RIDC3 lattices. Following treatment with DBCO, 10 equiv. of a synthesized ybbR peptide containing an azide-terminated non-natural amino acid, DSLEFIASKLA-G-K(N₃) (**Figure 2.3**), was added to the RIDC3-DBCO crystal suspension and allowed to react overnight. Negative-stain transmission electron microscopy (ns-TEM) snapshots taken during chemical modification confirmed that RIDC3 crystals remained intact during labeling (**Figure 2.4**). In order to quantify ybbR conjugation, suspensions of RIDC3+ybbR crystals were washed and dissolved with the addition of the metal chelator ethylenediaminetetraacetic acid (EDTA) prior to characterization with electrospray ionization mass spectrometry (ESI-MS) and UV-Vis spectroscopy to monitor chemical labeling (**Figure 2.2b, c**). ESI-MS measurements revealed up to 3 additions of DBCO and ybbR per RIDC3 monomer and were corroborated by UV-Vis

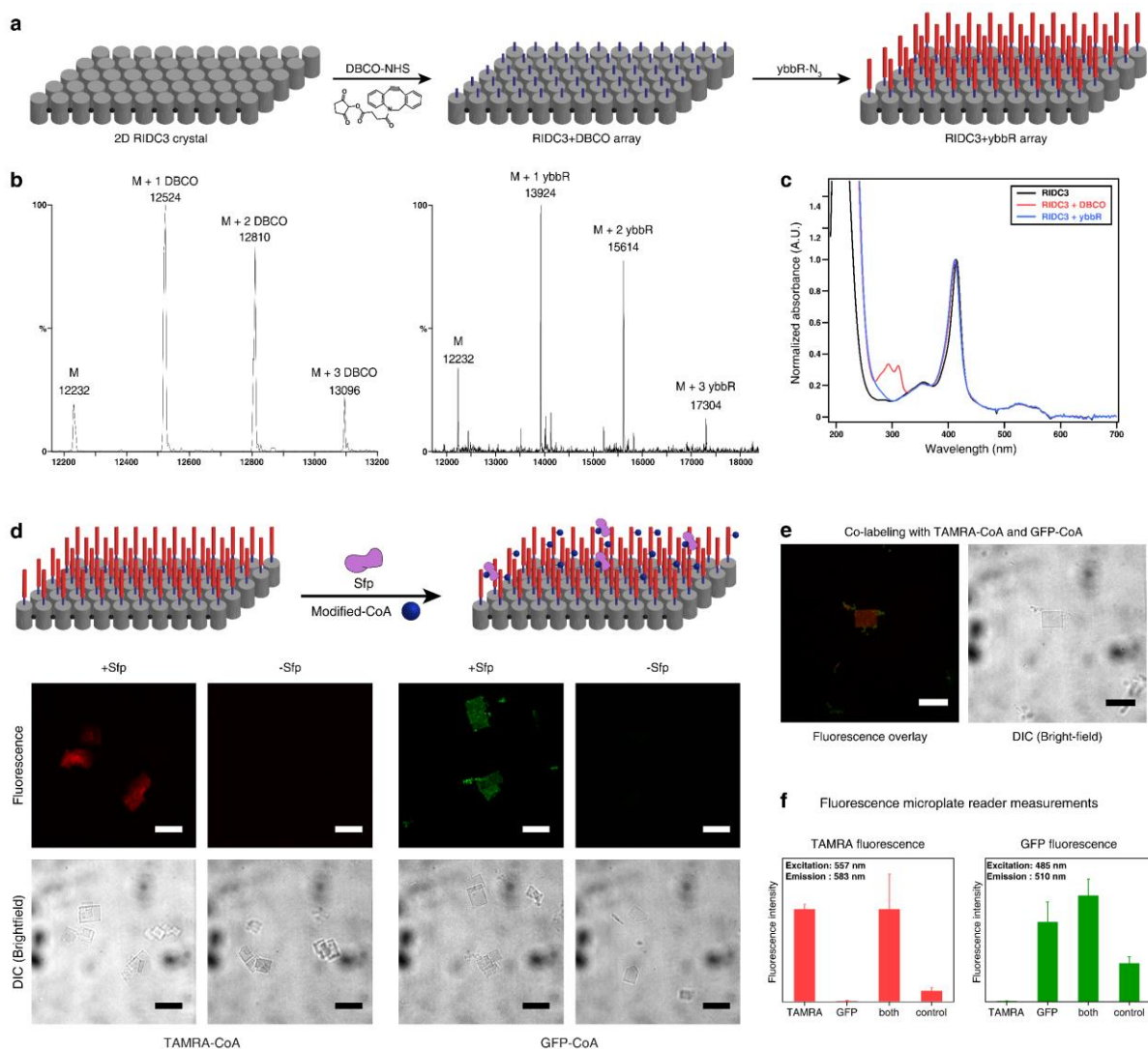


Figure 2.2 Characterization and enzymatic labeling of RIDC3+ybbR arrays. **(a)** Cartoon representations showing the generation of RIDC3+ybbR arrays from pre-formed RIDC3 crystals. Mass spectra **(b)** and UV-Vis profiles **(c)** are shown to confirm the addition of DBCO and ybbR. **(d)** Cartoon representation of enzymatic labeling and confocal microscopy images showing enzymatic modification of RIDC3+ybbR arrays with TAMRA-CoA and GFP-CoA. Samples in the absence of Sfp PPTase are not fluorescent. **(e)** Confocal microscopy images of RIDC3+ybbR arrays co-labeled with TAMRA and GFP. The fluorescence overlay of the red and green fluorescence confirms the addition of both labels. **(f)** Fluorescence plots of RIDC3+ybbR crystals after incubation with modified-CoA and Sfp. Scale bars are 10 μm .

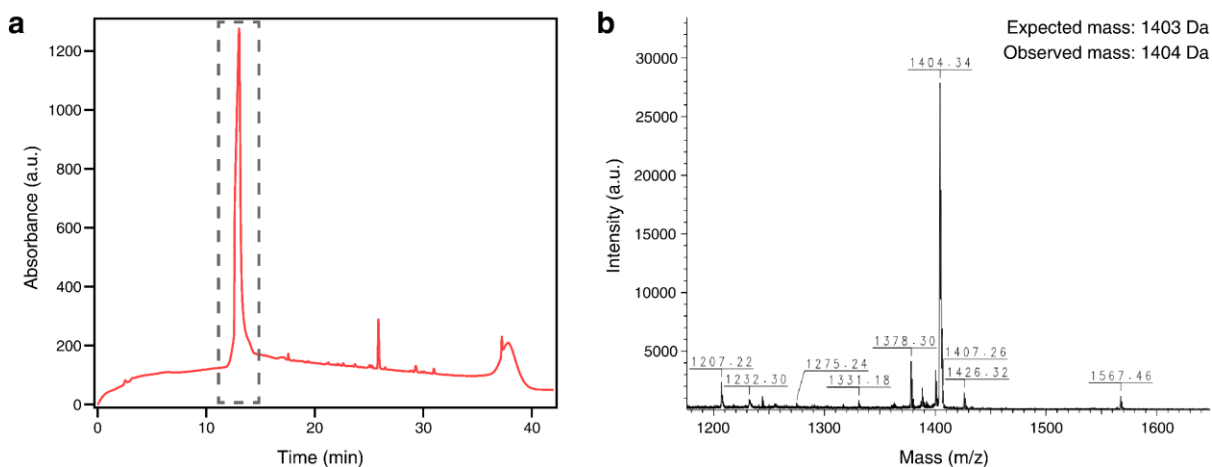


Figure 2.3 Characterization of synthesized ybbR-N₃ peptide. **(a)** HPLC chromatogram showing elution profile of ybbR-N₃ purification. **(b)** MALDI spectra of boxed region in **(a)** shows the expected mass for pure ybbR-N₃.

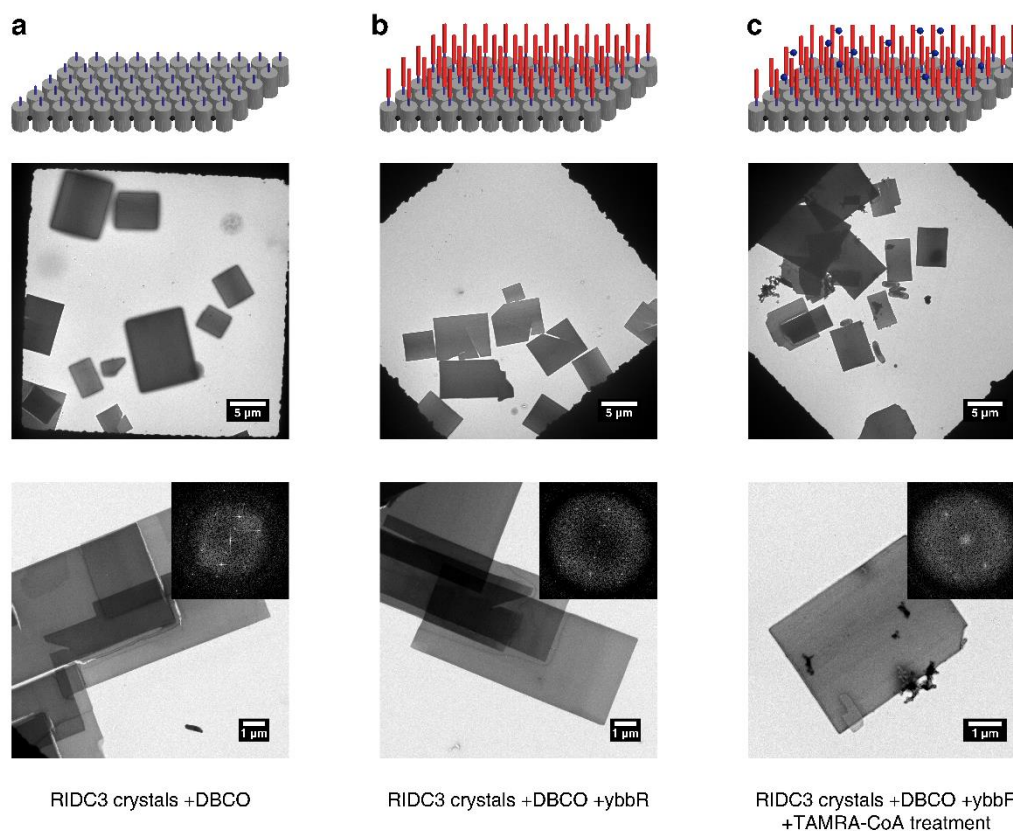
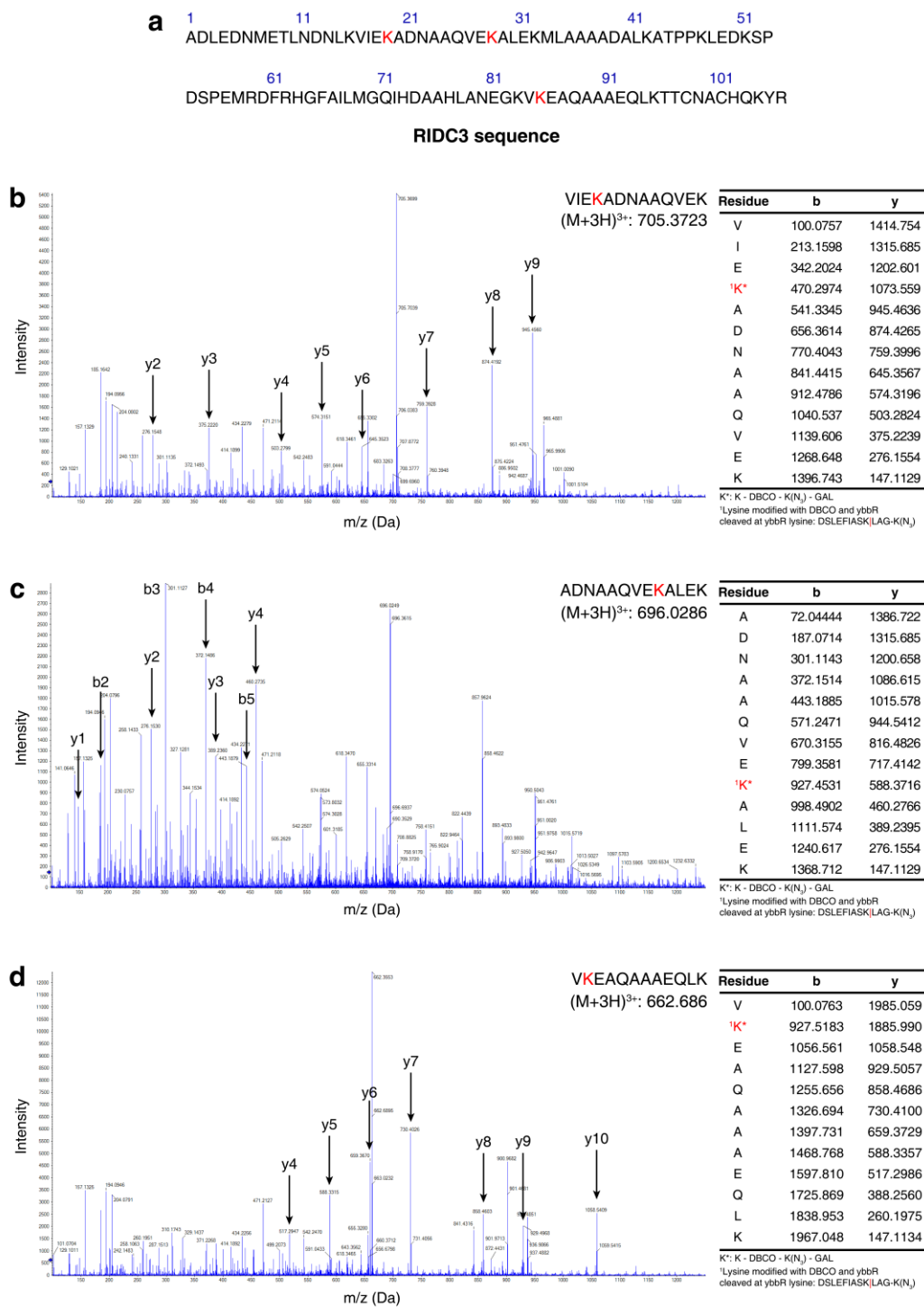


Figure 2.4 TEM characterization of modified RIDC3 crystals. Cartoon representation and TEM micrographs for **(a)** RIDC3 crystals modified with DBCO, **(b)** RIDC3+ybbR chemically modified crystals, and **(c)** RIDC3+ybbR crystals enzymatically modified using Sfp and TAMRA-CoA. Fast-fourier transform (FFT) inserts show that lattice crystallinity is retained throughout the process.

spectroscopy measurements monitoring the addition and consumption of DBCO ($\epsilon_{309} = 12000 \text{ M}^{-1} \text{ cm}^{-1}$)³⁹, with approximately 2.4 ybbR peptides added per RIDC3 monomer. MS/MS analysis of trypsin digested RIDC3-ybbR samples identified three sites of modification on RIDC3: Lys19, Lys28, Lys85 (**Figure 2.5**). While there are more lysines present on RIDC3 suitable for modification (**Figure 2.5a**), the length of the digested peptide may be too large to confidently identify through LC-MS/MS analysis. Nevertheless, spectroscopic and mass spectrometry data confirmed the covalent attachment of ybbR onto RIDC3 crystals at surface-exposed lysines.

Enzymatic labeling of RIDC3+ybbR lattices was performed using Sfp PPTase and fluorescent CoA analogs, which provided a facile visual handle for identifying labeled crystals. As a test case, a small molecule CoA analog (TAMRA-CoA) was added with Sfp and MgCl_2 to RIDC3-ybbR lattices and gently shaken overnight at room temperature. Control samples were similarly prepared with the absence of Sfp and all solutions were thoroughly washed with buffer to remove any unbound dye. Confocal microscopy measurements showed RIDC3+ybbR crystals incubated with Sfp and CoA were fluorescent whereas a negative control sample missing Sfp contained no fluorescence in the field-of-view (**Figure 2.2d**). Importantly, differential interference contrast (DIC) images of RIDC3 crystals show a perfect overlap between areas of fluorescence and the location of RIDC3+ybbR crystals (**Figure 2.2d**). Corresponding ns-TEM micrographs of the labeling samples confirmed that RIDC3+ybbR lattices remained intact after enzymatic treatment and retained crystallinity (**Figure 2.6**). Furthermore, the promiscuity of PPTase to CoA analogs permitted the use of a larger biomolecule, GFP-CoA, for the enzymatic transfer of a GFP-Ppant group onto ybbR. A GFP-CoA analog was generated by first incubating sfGFP with the bifunctional linker, sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate (sSMCC), to label lysine residues with the NHS-ester moiety followed by the addition of CoA-SH



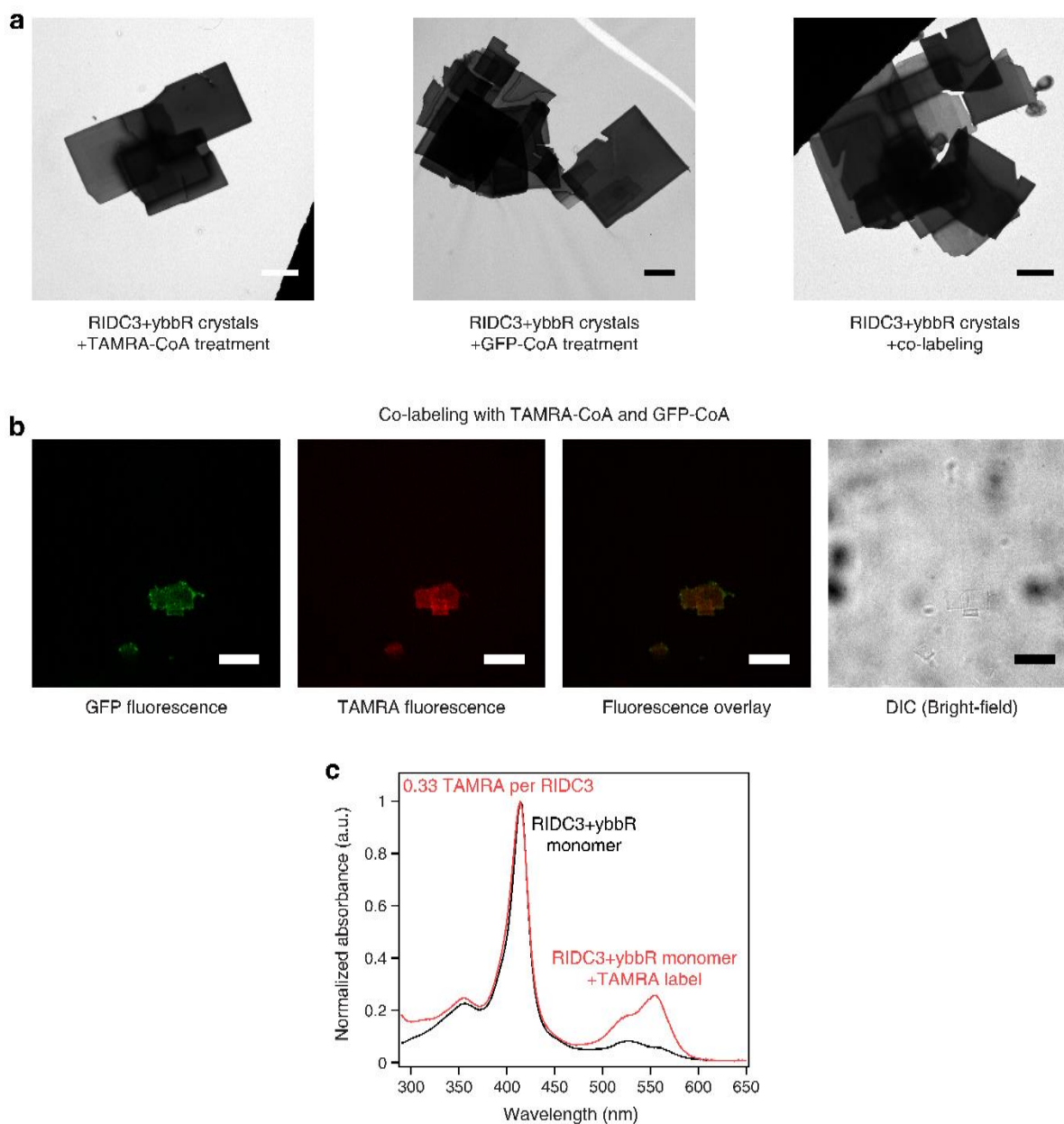


Figure 2.6 TEM and confocal microscopy characterization of enzymatically labeled RIDC3+ybbR arrays. **(a)** TEM analysis of crystals after enzymatic labeling to confirm that crystals remain intact. **(b)** Confocal microscopy images of RIDC3+ybbR arrays co-labeled with TAMRA-CoA and GFP-CoA. Individual fluorescence channels are shown followed by the overlay to confirm colocalization of both fluorescent molecules on the crystal. **(c)** UV-Vis characterization of RIDC3+ybbR monomer before and after enzymatic labeling with Sfp and TAMRA-CoA. Scale bars: **(a)** 1 μm , **(b)** 10 μm

for thiol-maleimide coupling (**Figure 2.7a**). CoA addition was quantified using UV-Vis measurements to yield ~1.4 CoA per sfGFP (**Figure 2.7b**). As with TAMRA-CoA modification, incubation of RIDC3+ybbR crystals with Sfp and GFP-CoA resulted in brightly fluorescent crystals (**Figure 2.2d**) and the absence of Sfp resulted in minimal background fluorescence. Based on the successful chemoenzymatic transfer of a small molecule dye (TAMRA-Ppant) or a protein (sfGFP-Ppant) onto RIDC3+ybbR crystals, we posited that a solution containing TAMRA-CoA and GFP-CoA could be used to colocalize both labels on the arrays. Suspensions of RIDC3+ybbR crystals incubated with both CoA substrates and Sfp displayed distinct fluorescent signals at TAMRA and GFP wavelengths, and an overlay of both images showed an unambiguous overlap of both their signals on the crystal surface (**Figure 2.2e**).

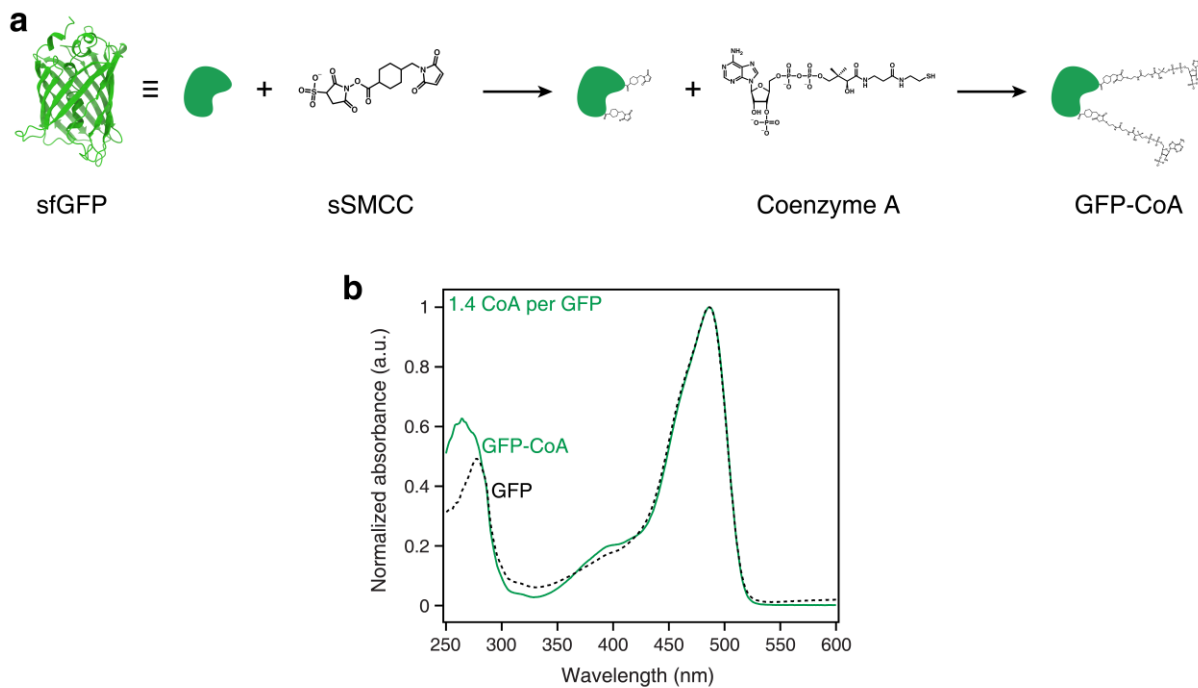


Figure 2.7 Generation and characterization of GFP-CoA. **(a)** Scheme for generation of GFP-CoA **(b)** UV-Vis profiles for GFP and GFP-CoA are shown normalized to the absorbance of GFP. The increase in absorbance at 260 nm is indicative of CoA conjugated onto GFP.

While identifying fluorescently labeled crystals was straightforward using confocal microscopy, quantification of enzymatic labeling of RIDC3+ybbR crystals proved was challenging. Crystals dissolved upon the addition of EDTA yielded soluble protein for UV-Vis spectroscopy but the signals for TAMRA or GFP were too low relative to RIDC3 to confidently position above background absorbance. This is due in part to the intense absorbance of the Soret and Q bands (415 and 527 nm, respectively) of the c-type heme covalently bound to the RIDC3 monomer, which partially overlaps with the absorption maximum of TAMRA (554 nm) and GFP (485 nm). Instead, we directly probed crystal suspensions using a microplate reader to measure fluorescence intensities for labeled vs. unlabeled samples (**Figure 2.8**). A standard curve of [fluorophore] vs. fluorescence intensity was prepared using free TAMRA-CoA or GFP-CoA to calculate concentration values for RIDC3+ybbR samples based on their fluorescence intensities. Crystals were dissolved after fluorescence measurements to obtain protein concentrations via UV-Vis spectroscopy. At a first glance, these measurements suggest less than 0.5% of RIDC3 is enzymatically modified by Sfp.

Since enzymatic reactivity at the surface may be lower than that of solution labeling, we obtained RIDC3+ybbR monomers (from dissolving chemically modified RIDC3+ybbR crystals with EDTA). Functionalization of these monomers revealed approximately 0.33 TAMRA per RIDC3, significantly higher than labeling on the crystal surface (**Figure 2.6c**). Previous studies on RIDC3 lattices have shown that solution-assembled crystals are multilayered³¹ perhaps limiting enzyme access to “interior” RIDC3 proteins with 3D stacked protein layers. Indeed, dry-state AFM characterization of RIDC3+ybbR crystals confirmed that they too consisted of ~4 nm thick 2D layers that stacked to an average height of 133±40 nm. Considering that only ybbR peptides decorating the exterior of RIDC3+ybbR crystals are likely available sites of modification with Sfp,

we used a combination of spectroscopic data (to quantify ybbR conjugation onto RIDC3) and AFM measurements (to gather crystal dimensions) to determine the enzyme-accessible fraction of the average RIDC3+ybbR crystal. From 46 crystals, we determined average dimensions of $a = 3.2 \pm 0.84 \mu\text{m}$, $b = 8.1 \pm 2.4 \mu\text{m}$, $c = 0.133 \pm 0.04 \mu\text{m}$ (**Figure 2.8**). We next used the RIDC3 crystal structure to determine the protein “step” in each crystal dimension: 1 protein per 3.79 nm in a , 1 protein per 3.46 nm in b , and 1 protein per 2.3 nm in c (PDB ID: 3TOM). Based on these measurements, we calculated the surface coverage of RIDC3 proteins to be ~3.5% of total protein within the crystal. After correlating fluorescence intensities to absorbance values (**Figure 2.8**), we computed ~3% and ~10% enzymatic labeling of RIDC3+ybbR crystal surfaces for GFP-CoA and TAMRA-CoA respectively. Overall, this confirmed that labeling on the surface was less effective than in solution and the formation of multilayered crystals largely prevented the bulk of RIDC3 proteins from easily accessing to Sfp and CoA.

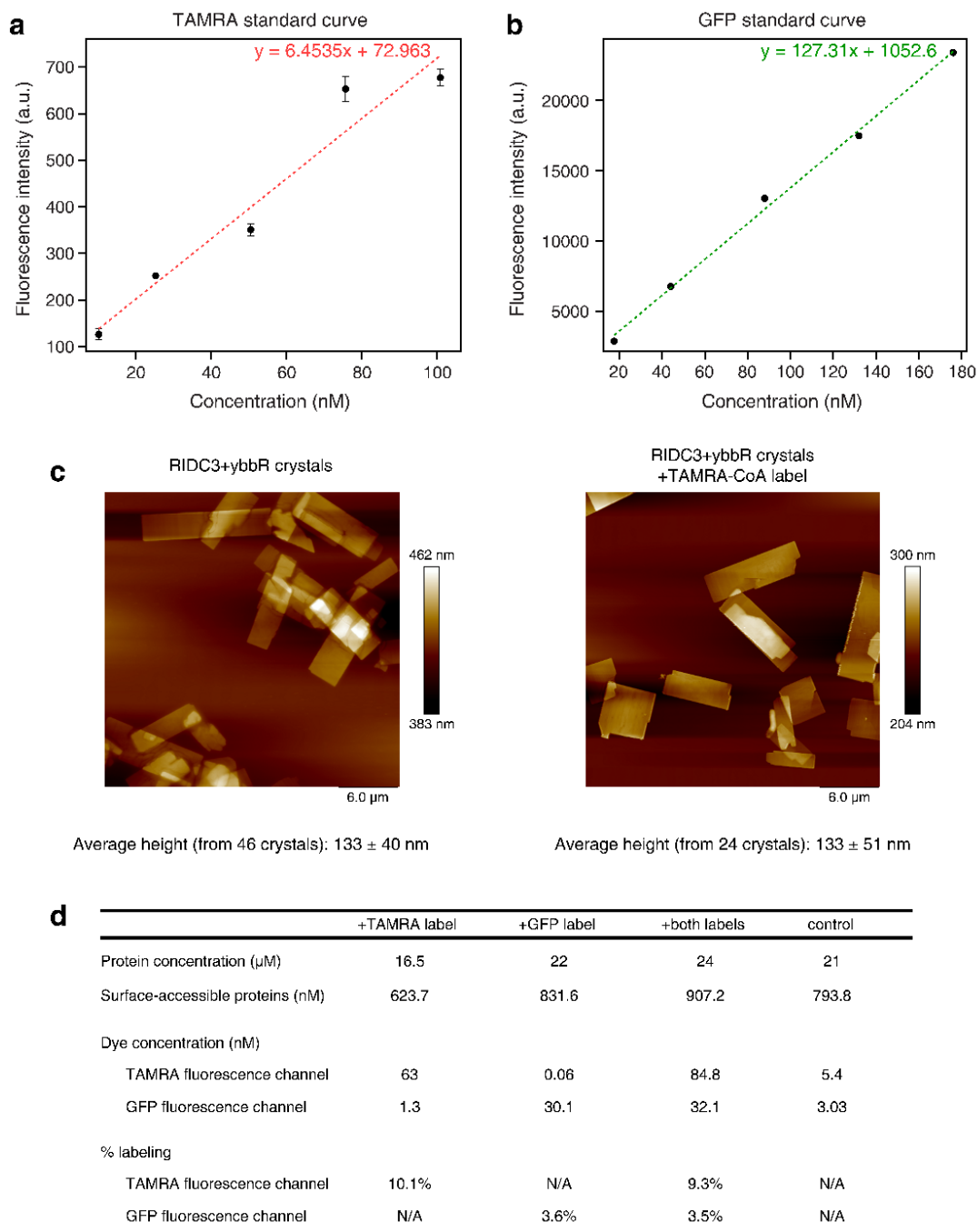


Figure 2.8 Quantification of fluorescent labeling of RIDC3+ybbR crystals. **(a), (b)** Standard curves plotting dye concentration vs. fluorescence. **(c)** Representative AFM images of RIDC3+ybbR crystals before and after enzymatic labeling with TAMRA-CoA. The average height of measured crystals are listed below. **(d)** Table showing protein and dye concentrations calculated based on UV-Vis (for protein) or using the standard curves in **(a), (b)** for the dye. Percentage of surface labeling is calculated based off the concentration of surface-accessible proteins available for modification.

2.3.2 Generation of RIDC3+CoA lattices

In addition to chemically conjugating ybbR onto RIDC3 arrays, we wondered whether a surface displaying CoA molecules and enzymatically linked to a soluble ybbR-bearing molecule would alter labeling efficiency or fluorophore distribution on the crystals. Thus, RIDC3+CoA crystals were prepared using a similar conjugation strategy to that used for generating GFP-CoA (**Figure 2.9a**). RIDC3 crystals were first treated with sSMCC to covalently attach to surface exposed lysines followed by the addition of CoA to react with the maleimide portion of SMCC. CoA addition was quantified via UV-Vis spectroscopy ($\lambda_{\text{max}} = 259 \text{ nm}$) and determined to be ~ 0.8 per RIDC3 (**Figure 2.9b**). Though a 25-fold excess of CoA was added after SMCC coupling, lower efficacy of thiol-maleimide chemistry at the crystal surface (especially compared to the strained cyclooctyne click chemistry used for ybbR conjugation) and the potential for CoA self-dimerization were likely the primary factors that resulted in a lower CoA:RIDC3 ratio. Nevertheless, incubation of RIDC3+CoA crystals with Sfp and a genetically encoded GFP-ybbR resulted in fluorescent green rectangles under a confocal microscope (**Figure 2.9c**). Control experiments in the absence of PPTase or GFP-ybbR showed no fluorescence, indicating that there is little to no non-specific association of GFP to the RIDC3+CoA crystals. Interestingly, GFP-ybbR-labeled RIDC3+CoA crystals consist of punctate patches of GFP forming an outline along the edges of the crystal unlike RIDC3+ybbR samples, which showed uniform fluorescence intensity across the crystal surface. This could arise from a greater density of CoA molecules on the edges of the crystal that enables easier access of Sfp and GFP-ybbR to the edges relative to the interior. Quantification of GFP-ybbR labeling onto RIDC3+CoA labels revealed $\sim 9.7\%$ of surface exposed proteins were labeled, on par with that of RIDC3+ybbR crystals. In both RIDC3+ybbR and RIDC3+CoA preparations, enzymatic labeling only covered a fraction of the surface sites

containing ybbR or CoA respectively. This could be due to a lower efficacy of enzymatic transfer on crystal suspensions relative to free protein (as observed in enzymatic transfer onto RIDC3+ybbR monomers vs. RIDC3+ybbR crystals) and a non-periodic display of the functional moiety leading to clusters of labeled regions on the surface of the lattices. Thus, a RIDC3-ybbR construct was created with the ybbR peptide genetically appended to the c-terminus of RIDC3 to allow for (1) a uniform distribution of peptide upon forming crystals and (2) avoid post-assembly conjugation of molecules to generate functional crystals (**Figure 2.1a**).

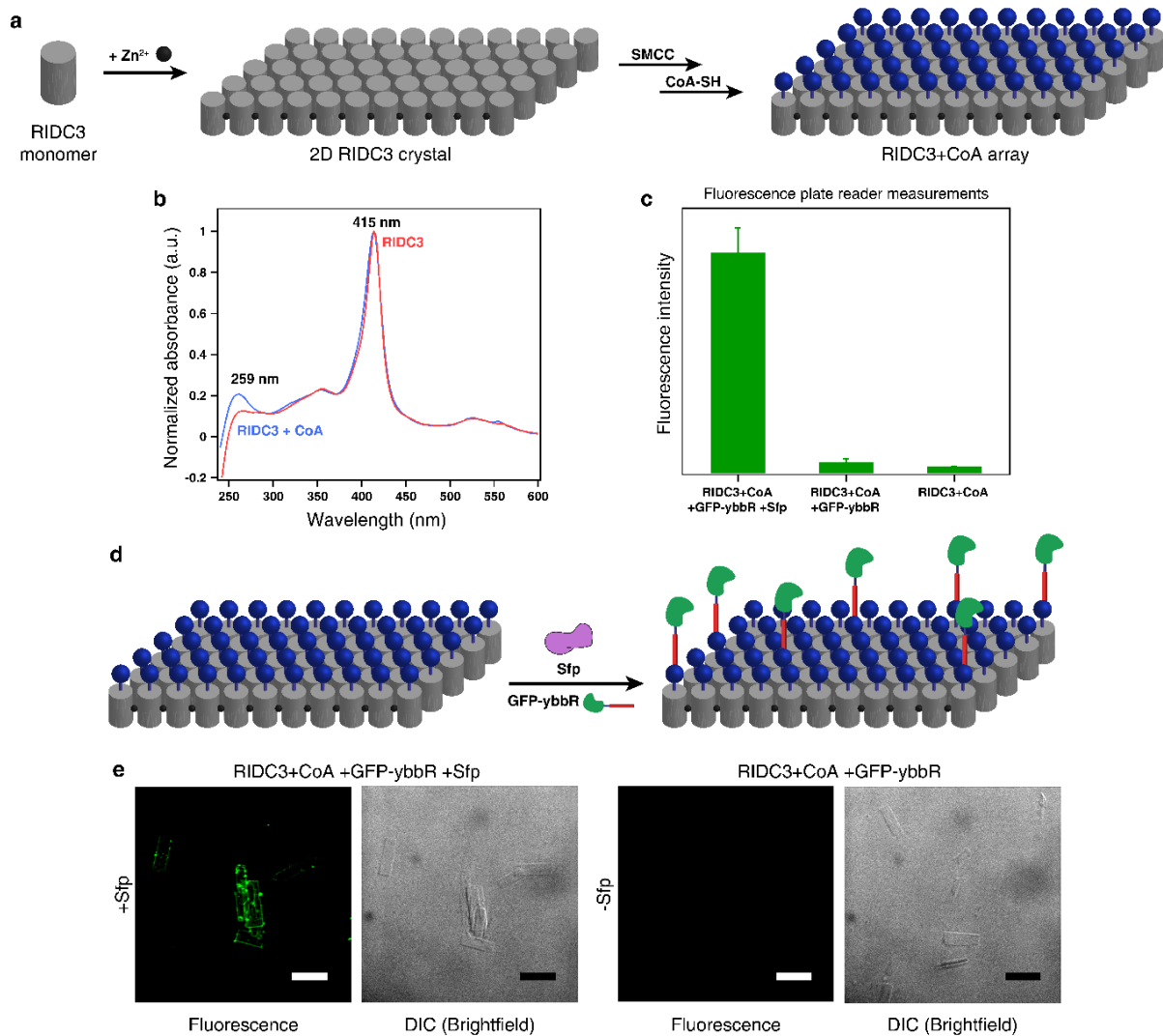


Figure 2.9 Characterization and enzymatic labeling of RIDC3+CoA arrays. **(a)** Cartoon schematic for the generation of RIDC3+CoA arrays that can be enzymatically modified with GFP-ybbR. **(b)** UV-Vis spectra of RIDC3 and RIDC3+CoA normalized to the RIDC3 absorption maximum at 415 nm. **(c)** Fluorescence intensities for enzymatic labeling samples with and without Sfp PPTase or GFP-ybbR. **(d)** Cartoon schematic showing the proposed labeling of arrays with GFP-ybbR. **(e)** Confocal microscopy images of RIDC3+CoA samples incubated with GFP-ybbR with and without Sfp PPTase. Scale bars are 10 μ m.

2.3.3 Design and characterization of genetically modified RIDC3-ybbR and RhuA-ybbR

The C-terminus of RIDC3 lie orthogonal to the Zn-protein interface and face the exterior of the crystal surface, making it an ideal location for the genetic insertion of ybbR. RIDC3 was genetically modified with a Gly-Gly spacer followed by the ybbR peptide (RIDC3 – GG – DSLEFIASKLA) to generate the RIDC3-ybbR plasmid, which was expressed and purified in a similar fashion to RIDC3. In the final stage of the purification, we noticed two peaks in the FPLC chromatogram (**Figure 2.10a**) corresponding to two different protein species, as confirmed by mass spectrometry analysis (**Figure 2.11a**). A small fraction contained the intact RIDC3-ybbR protein, but a majority constituted of a truncated protein ~400 Da smaller than RIDC3-ybbR. This cleaved protein product was identified as RIDC3-ybbR missing the final four amino acids (RIDC3 – GG – DSLEFIA). Using Native PAGE as a facile handle for assessing the abundance of uncleaved and cleaved protein (**Figure 2.10b**), various purification conditions were tested (temperature, addition of protease inhibitors, pH); however, in all cases, very little intact RIDC3-ybbR was isolated. The addition of the peptide to RIDC3 may have rendered it unstable, leading to the autolytic conversion into the truncated RIDC3-ybbR product. Nevertheless, array formation and enzymatic labeling efficacy were evaluated using limited quantities of pristine RIDC3-ybbR.

We first tested whether the addition of the C-terminal peptide deterred the formation of Zn-mediated RIDC3-ybbR lattices. Crystal formation was assessed after screening a variety of conditions (pH, Zn:protein molar ratios, and protein concentration). Ordered RIDC3-ybbR assemblies were found in 100 μ M protein preparations at pH 7.5 and $[\text{Zn}^{2+}]:[\text{RIDC3-ybbR}]$ ratios of 5 (**Figure 2.11b**). Interestingly, protein nanotubes that form at pH > 8 or $[\text{Zn}^{2+}]:[\text{protein}]$ ratios > 50 with RIDC3 were observed at pH 5.5 and $[\text{Zn}^{2+}]:[\text{RIDC3-ybbR}]$ ratios of 5 (**Figure 2.10c**). The difference in assembly conditions for similar assembly morphologies between RIDC3 and RIDC3-

ybbR is likely due to the addition of the peptide, which may disfavor stacking of 2D layers at low pH and thus encourage the formation of helical nanotubes. Encouraged by RIDC3-ybbR crystal formation, enzymatic labeling was tested using RIDC3-ybbR monomers incubated with Sfp PPTase and TAMRA-CoA. SDS PAGE and mass spectrometry confirmed that RIDC3-ybbR monomers were amenable to chemoenzymatic labeling with TAMRA-CoA (**Figure 2.11c, d**). Truncated RIDC3-ybbR protein, the major isolated fraction of RIDC3-ybbR purifications, was also tested for enzymatic labeling and found to be unreactive. Previous studies on ybbR have shown that truncated peptides cannot form the alpha-helical motif hypothesized to be necessary for PPTase recognition.³² Due to difficulties in preparing appreciable quantities of intact RIDC3-ybbR, we sought out an alternative self-assembly system that would sustain the stable incorporation of a genetic tag and assemble into crystalline 2D lattices. Additionally, using a secondary platform also allowed us to assess the generalizability of using ybbR as a substrate for enzymatic modification of protein lattices.

In a recent report from our laboratory, the C_4 -symmetric L-rhamnulose-1-phosphate aldolase (RhuA) was modified with cysteine residues at its corners to facilitate the self-assembly of unsupported 2D lattices in solution under controlled oxidation conditions via disulfide bond formation. The C_98 RhuA proteins tessellate to form an alternating arrangement of tetramers in the crystalline lattice. Furthermore, rotations about the individual axes of symmetry at the flexible disulfide linkages create a coherently dynamic, auxetic lattice (*i.e.* a longitudinal expansion when stretched in the transverse direction) with a theoretical maximum Poisson's ratio of -1. The larger size and inherent symmetry of C_98 RhuA could provide a more stable scaffold for the genetic incorporation of ybbR while retaining its self-assembly properties.

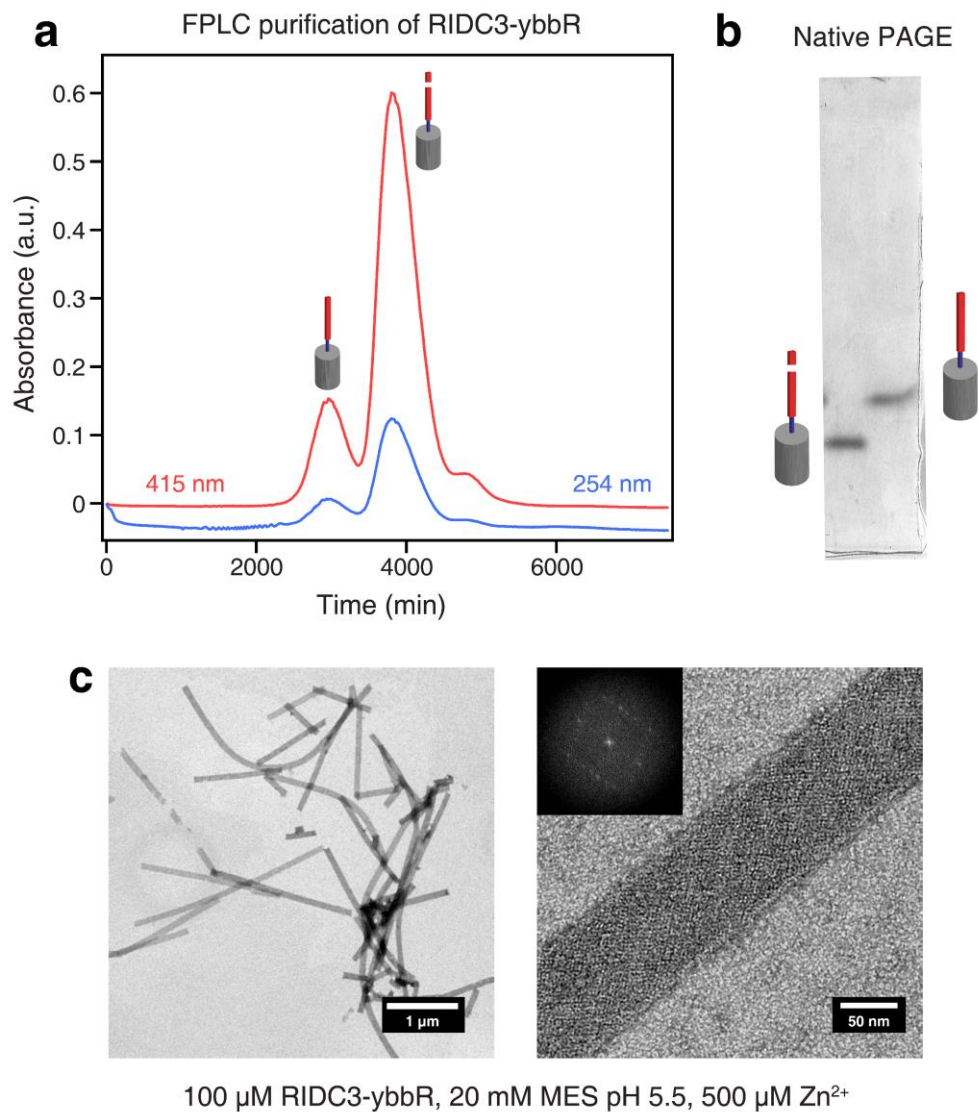


Figure 2.10 Characterization of genetically modified RIDC3-ybbR. **(a)** FPLC chromatogram showing separation of intact and cleaved RIDC3-ybbR protein during purification. **(b)** Native PAGE can be used to identify cleaved vs. intact proteins. The difference in molecular weight is too small to discern by SDS PAGE. **(c)** TEM micrographs of RIDC3-ybbR nanotubes formed at pH 5.5. FFT inset shows the same diffraction pattern as RIDC3 nanotubes.

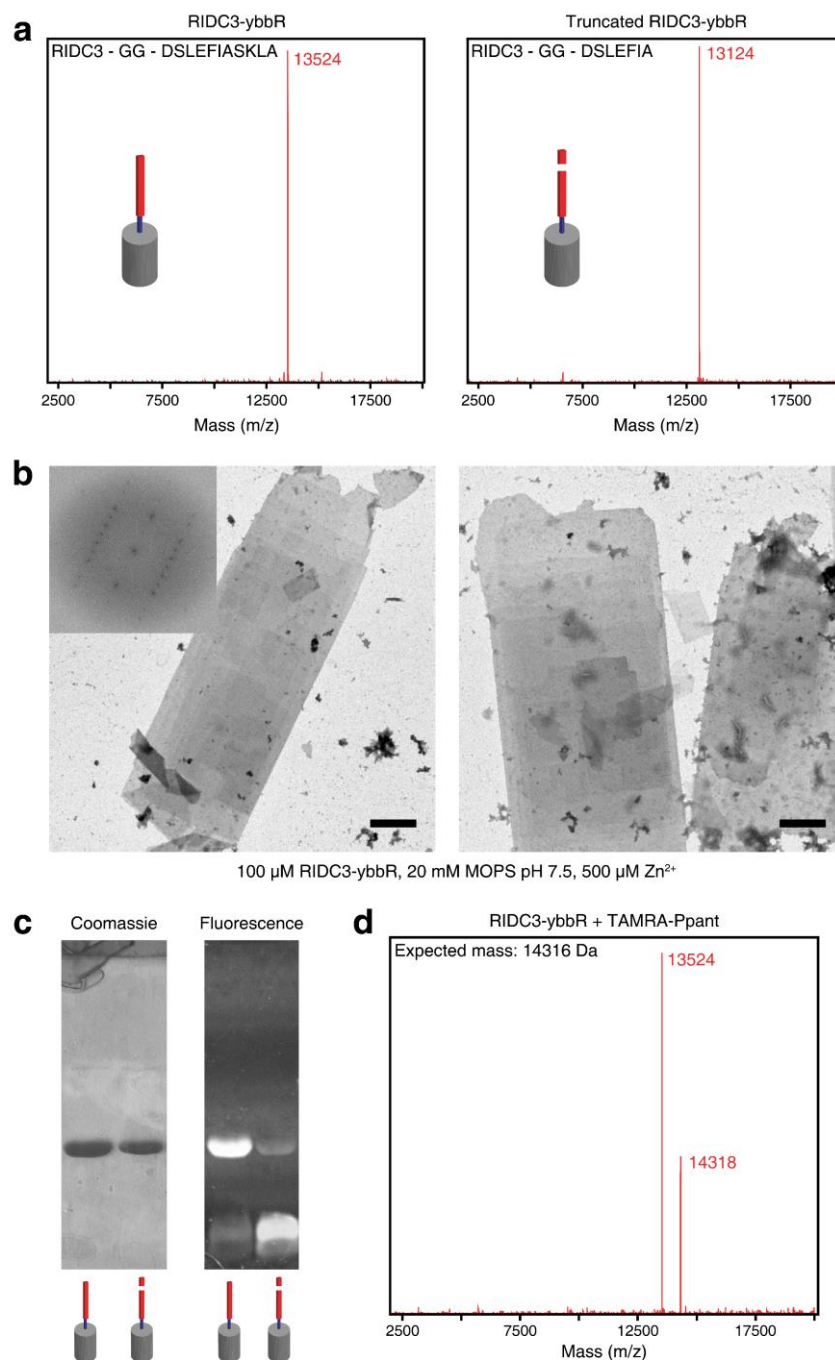


Figure 2.11 Self-assembly and enzymatic labeling of genetically incorporated RIDC3-ybbR. **(a)** Mass spectra of intact and truncated RIDC3-ybbR observed during purification. **(b)** TEM micrographs of 2D RIDC3-ybbR arrays formed at pH 7.5. FFT inset shows the same diffraction pattern as RIDC3 crystals. **(c)** SDS PAGE gels showing enzymatic labeling of RIDC3-ybbR and truncated RIDC3-ybbR protein. Fluorescence images of the gel confirm that the truncated protein cannot be enzymatically modified. **(d)** Mass spectrum showing peaks for RIDC3-ybbR and RIDC3-ybbR modified using TAMRA-CoA.

As with RIDC3, *ybbR* was genetically appended to the c-terminus of C98 RhuA with a short spacer (C98 RhuA – SGSG – DSLEFIASKLA) so that the peptide is displayed at the surface of 2D lattices minimally interfering with disulfide formation (**Figure 2.12a**). Importantly, no truncation of *ybbR* was observed during the purification and isolation in this construct; RhuA may be less susceptible to truncations due to increased protein stability from its innate 4-fold symmetry and larger size relative to *cb562*. C98 RhuA-*ybbR* proteins self-assemble into a homogenous suspension of ordered crystals with identical crystallinity and lattice coherence to C98 RhuA lattices in similar solution conditions (**Figure 2.12b**). Furthermore, C98 RhuA-*ybbR* tetramers were successfully modified after incubation with Sfp PPTase and TAMRA-CoA, containing ~0.2 TAMRA per polypeptide (*i.e.* 0.8 TAMRA per C98 RhuA-*ybbR* tetramer) after enzymatic labeling (**Figure 2.12c**). The presence of 4 *ybbR* peptides on each C-terminal face of the RhuA tetramer may have a crowding effect that reduces enzymatic labeling efficacy, potentially leading to a lower yield of functionalized C98 RhuA-*ybbR* proteins. Pre-formed C98 RhuA-*ybbR* crystals were observed via confocal microscopy and ns-TEM after incubation with Sfp PPTase and modified CoA molecules. Brightly fluorescent crystals were found with both TAMRA-CoA and GFP-CoA-incubated solutions whereas C98 RhuA-*ybbR* lattices in the corresponding DIC images were much harder to discern likely due to the increased porosity and smaller size of the crystals relative to RIDC3 (**Figure 2.12d**). TEM analysis of the labeled samples confirmed that the arrays remained intact after enzymatic modification (**Figure 2.12d**). Confocal microscopy and ns-TEM analysis of control samples devoid of Sfp PPTase were minimally fluorescent and remained crystalline (**Figure 2.13**). In contrast to genetically modified RIDC3-*ybbR*, the C98 RhuA-*ybbR* construct was found to be intact and stable and could be enzymatically modified without loss in crystallinity.

Selective removal of functional labels after post-translational modification of ybbR has been previously demonstrated *in vitro* on acyl carrier proteins and short peptides using an acyl carrier protein hydrolase (AcpH). Though these experiments were performed on soluble proteins, we wondered whether suspensions of crystals functionalized with TAMRA or GFP were amenable to reversible labeling using Sfp and AcpH. Isolated GFP or TAMRA-labeled RIDC3+ybbR crystals were incubated with recombinantly expressed AcpH from *Pseudomonas fluorescens*, gently shaken overnight at ambient temperatures, and exchanged into fresh buffer (to remove any free dye from hydrolysis) prior to analysis by confocal microscopy. RIDC3+ybbR crystals were imaged and fluorescence intensity was normalized with crystals prior to AcpH treatment. However, we found that the crystals remained fluorescent after incubation with AcpH, indicating that the labels were not hydrolyzed (**Figure 2.14**). In order to confirm that AcpH was functioning properly, we generated an enzymatically labeled TAMRA-ACP and incubated with AcpH. SDS PAGE gels of ACP samples before and after incubation with AcpH indicated that a substantial amount of dye (ca. 60%) was hydrolyzed (**Figure 2.14**). We next prepared enzymatically labeled suspensions of RIDC3+ybbR crystals, dissolved the crystals and incubated with *Pf* AcpH. These reactions were performed using soluble protein instead of crystal suspensions to assess AcpH activity on ybbR-containing RIDC3 proteins. However, AcpH treatment of RIDC3+ybbR did not result in any appreciable reduction in fluorescence (**Figure 2.14**). This suggests that AcpH affinity for our ybbR-protein fusions is insufficient for hydrolytic activity, thus precluding reversible labeling of our arrays.

2.4 Conclusions

In this study, we investigated the incorporation of a functional peptide handle, serving as a target for enzymatic modification, onto 2D RIDC3 and RhuA crystalline lattices. Chemical conjugation of ybbR onto RIDC3 arrays provided a nonuniform surface coating of functional peptides that allowed for the enzymatic transfer of fluorescent phosphopantetheine analogs onto ybbR. A genetically modified RIDC3-ybbR was implemented for the periodic patterning of ybbR on RIDC3 proteins, but the inherent instability of the appended peptide on this construct precluded extensive analysis of the RIDC3-ybbR platform. However, incorporation of ybbR onto the larger RhuA scaffold provided an alternative path to generating genetically encoded lattices. In both instances, genetic modification of the proteins did not perturb their self-assembly behavior and showed that the rational design of crystalline materials with tunable functionality is possible through a minimal peptide insertion. Enzyme-mediated post-translational modifications are ubiquitous in nature and are, in part, responsible for the complexity of the proteome. The bottom-up assembly of hierarchical materials is of active interest and the tailoring the surfaces of protein biomaterials using enzymes provides a promising approach for generating functional and hybrid materials for use in sensing, catalysis, or lab-on-a-chip design frameworks.

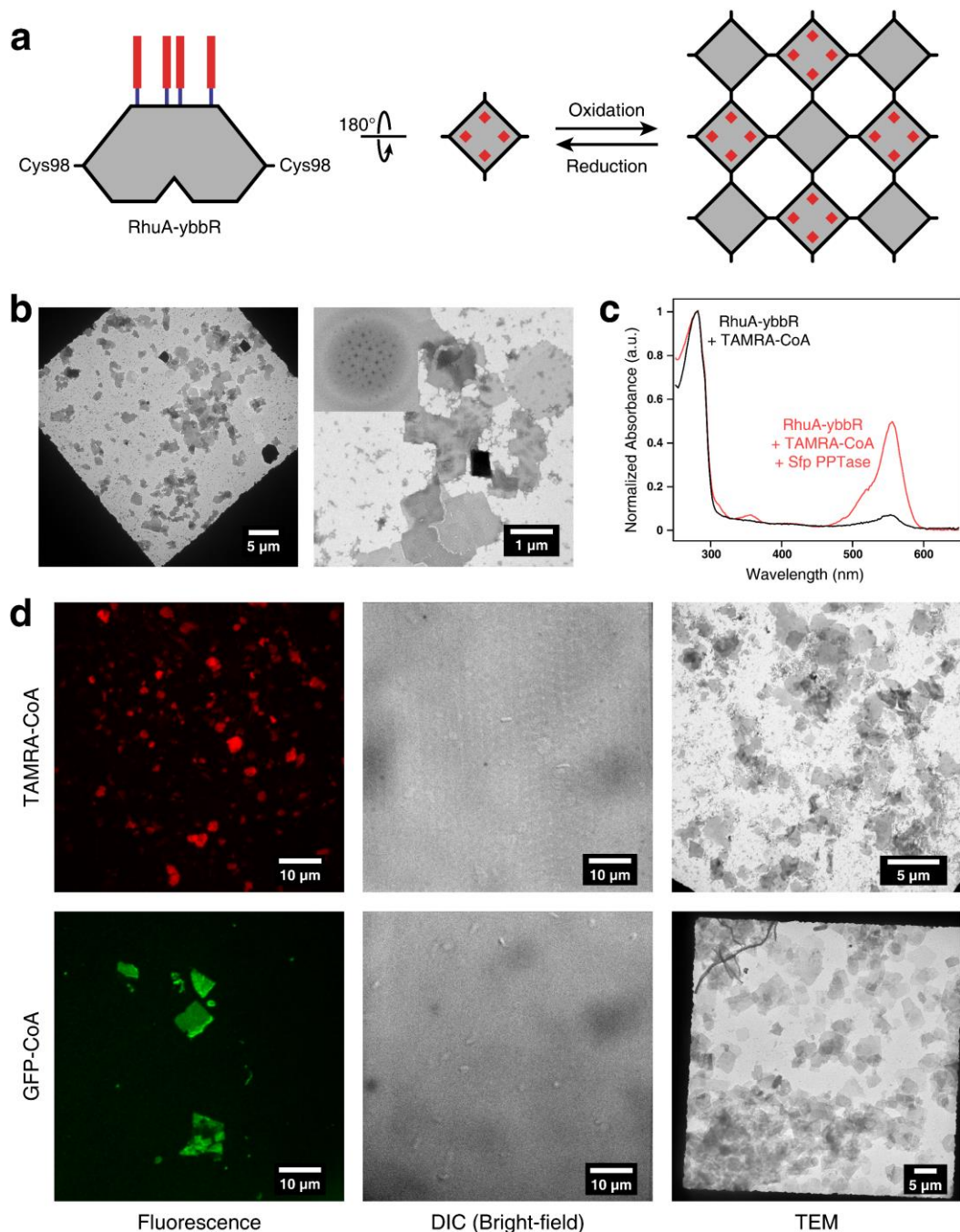


Figure 2.12 Characterization and enzymatic modification of C^{98} RhuA-ybbR. **(a)** Cartoon schematic of genetic incorporation of ybbR to the C-terminus of RhuA. **(b)** TEM micrographs of 2D C^{98} RhuA-ybbR crystals. FFT inset shows the same pattern as C^{98} RhuA crystals. **(c)** UV-Vis profiles of RhuA-ybbR before and after enzymatic labeling using TAMRA-CoA. Samples are normalized to the RhuA-ybbR absorbance maximum at 280 nm. **(d)** Confocal microscopy and TEM characterization of RhuA-ybbR incubated with Sfp PPTase and either TAMRA-CoA or GFP-CoA.

RhuA-ybbR controls: no Sfp added

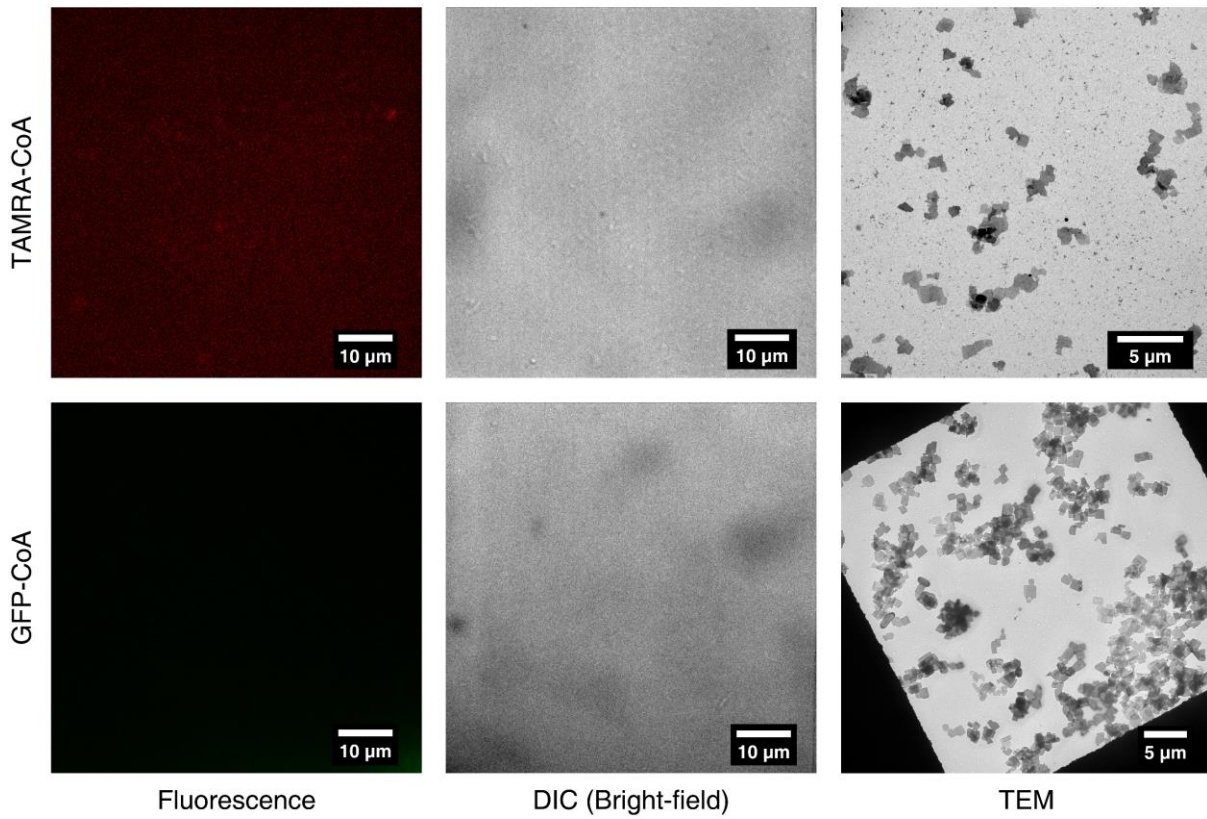


Figure 2.13 Confocal microscopy and TEM characterization of RhuA-ybbR control samples. In the absence of Sfp, no fluorescence is observed. RhuA-ybbR crystals remain intact in the labeling conditions.

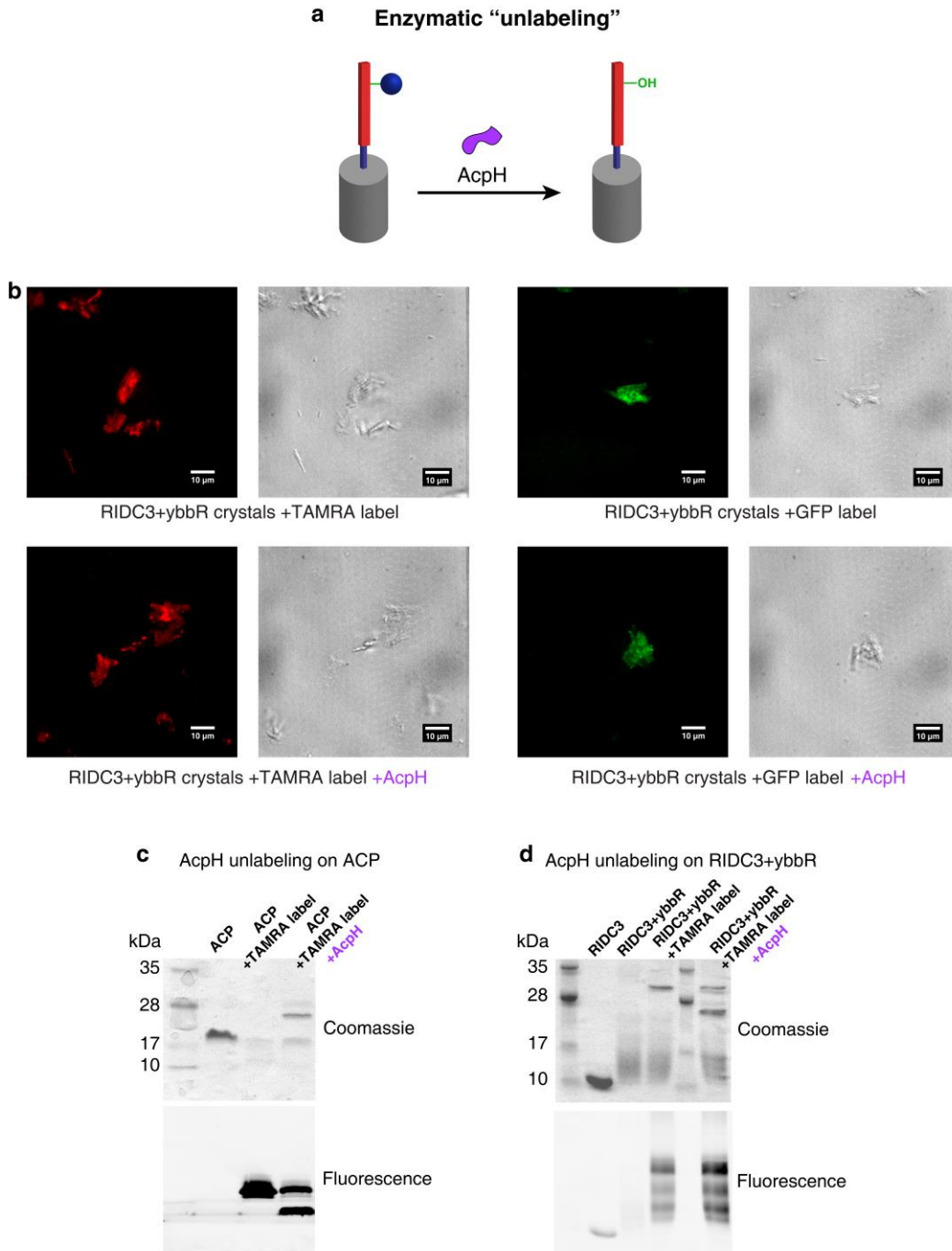


Figure 2.14 Attempts at reversible labeling of RIDC3+ybbR using AcpH. **(a)** Cartoon schematic of enzymatic “unlabeling” of CoA-modified RIDC3+ybbR proteins with AcpH. **(b)** Confocal microscopy images before and after incubation of labeled RIDC3+ybbR crystals with AcpH. Fluorescence intensity is not noticeably different after AcpH treatment. **(c)** SDS PAGE gel of AcpH treatment of acyl carrier protein (ACP) shows unlabeling of ACP-TAMRA sample after AcpH incubation. **(d)** SDS PAGE gel of RIDC3+ybbR sample labeled with TAMRA-CoA does not show decreased fluorescence after AcpH treatment.

2.5 Materials and Methods

2.5.1 Protein expression and purification

The ybbR peptide was cloned into a pET20b(+) plasmid housing the RIDC3 vector using Quikchange mutagenesis protocols. RIDC3 and RIDC3-ybbR was expressed and purified.⁴⁰ Briefly, cells harboring the RIDC3 or RIDC3-ybbR plasmid were grown in a lysogeny broth (LB) medium containing 100 mg/L ampicillin and 34 mg/L chloramphenicol in 2.8-L flasks and shaken at 200 rpm for 16-20 h at 37 °C. Cells were harvested by centrifugation at $5,000 \times g$ for 10 min, resuspended in a buffered solution containing 5 mM sodium acetate (pH 5.0) and 100 mg lysozyme and lysed by sonication on ice. The crude lysate was subject to pH titration with 40% sodium hydroxide to pH 10 immediately followed by the addition of 50% (v/v) hydrochloric acid to pH 4.5 and centrifuged at $10,000 \times g$ for 30 min. The cleared lysate was applied to a CM Sepharose gravity column and eluted using a stepwise gradient (0 – 500 mM NaCl). Collected fractions were concentrated using an Amicon Stirred Cell (Millipore) and dialyzed against a buffered solution containing 10 mM sodium phosphate (NaPi) (pH 8.0) overnight. The protein was collected and purified using a DuoFlow workstation station equipped with a Bio-Scale Mini Macro-prep High Q-cartridge column (BioRad) and eluted using a linear gradient (0 – 500 mM NaCl). Protein purity was assessed by ultraviolet-visible (UV-Vis) spectroscopy and samples with a RZ ratio (A_{415}/A_{280}) > 6 were pooled, concentrated and flash frozen for storage at -80 °C.

A pJ414 plasmid housing the RhuA-ybbR gene was purchased from DNA 2.0. RhuA-ybbR was expressed and purified as previously described.²⁵ Cells were grown in 30-mL LB cultures containing 100 mg/L ampicillin and shaken overnight at 37 °C. From these starter cultures, 5 mL was inoculated into 1-L LB media and shaken at 200 rpm at 37 °C to an optical density (OD) of 0.8 – 1. Protein expression was inducted with 1 mM isopropyl β -D-1- thiogalactopyranoside

(IPTG) for 12 h and cells were harvested by centrifugation at $5,000 \times g$ for 10 min. Cells were resuspended in a buffered solution containing 10 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris) (pH 7.5), 10 mM β -mercaptoethanol (β ME) and 1 mM $ZnCl_2$ supplemented with 100 mg lysozyme and 50 mg PMSF and lysed by sonication on ice. Cell lysate was centrifuged at $10,000 \times g$ for 30 min, the cleared lysate was subject to the addition of Polymin-P (Acros) to a final concentration of 0.15% (w/v), and stirred for 30 min at 4 °C prior to centrifugation at $10,000 \times g$ for 30 min. The supernatant was loaded onto a DEAE Sepharose gravity column and eluted using a stepwise gradient (0 – 500 mM NaCl). Fractions containing RhuA-ybbR were subject to the addition of ammonium sulfate to a final concentration of 1.7 M to precipitate the protein and stirred for 30 min prior to centrifugation at $10,000 \times g$ for 30 min. The precipitated protein was resuspended in a buffered solution containing 10 mM Tris (pH 7.5), 10 mM β ME, and 1 mM $ZnCl_2$ and dialyzed against the same solution three times. The protein was collected and purified using a DuoFlow workstation station equipped with a Bio-Scale Mini Macro-prep High Q-cartridge column (BioRad) and eluted using a linear gradient (0 – 500 mM NaCl). Protein purity was assessed by SDS PAGE and pure fractions were dialyzed against a buffered solution containing 10 mM Tris (pH 7.5), 10 mM β ME, and 1 mM $ZnCl_2$ and flash frozen for storage at -80 °C.

B. subtilis Sfp, *P. fluorescens* AcpH and superfolder GFP (sfGFP) were expressed and purified as previously described.^{34,41} Briefly, cells were inoculated into terrific broth containing 100 mg/L kanamycin and shaken at 200 rpm at 37 °C to an OD of 0.8. Protein expression was induced with 1 mM IPTG and shaken at 200 rpm overnight at 16 °C. Cells were harvested by centrifugation at $5,000 \times g$ for 10 min, resuspended in a buffered solution containing 50 mM Tris (pH 7.5) and 250 mM NaCl for Sfp and GFP or 50 mM Tris (pH 8.0) and 250 mM NaCl for AcpH

supplemented with 100 mg lysozyme, 5 $\mu\text{g/mL}$ DNase I, and 5 $\mu\text{g/mL}$ RNase and lysed by French pressure cell press (500 – 1,000 psi). The cell lysate was centrifuged at $12,000 \times g$ for 45 min and the cleared supernatant was subject to a Ni-NTA gravity column. The column was washed with a buffered solution containing 50 mM Tris (pH 7.5), 250 mM NaCl and 10 mM imidazole for Sfp and GFP or 50 mM Tris (pH 8.0), 250 mM NaCl and 10 mM imidazole for AcpH prior to protein elution at 300 mM imidazole. Pure fractions were desalted into a buffered solution containing 50 mM Tris (pH 7.5) using a PD-10 desalting column (GE Healthcare Life Sciences), concentrated using an Amicon spin filter (Millipore), exchanged into a buffered solution containing 50 mM Tris (pH 7.4), 150 mM NaCl and 20% glycerol for Sfp and GFP or 50 mM Tris (pH 8.0), 250 mM NaCl and 25% glycerol for AcpH and flash frozen at $-80\text{ }^{\circ}\text{C}$.

2.5.2 Preparation of ybbR-N₃ peptide

The desired sequence of the synthesized ybbR-N₃ peptide is: DSLEFIASKLAG-K(N₃). The Fmoc-Lys(N₃) amino acid was purchased separately from Anaspec. Solid phase peptide synthesis using Rink amide MBHA resins was used to generate the peptide. A solution of 20% 4-methylpiperidine in dimethylformamide (DMF) was used for Fmoc deprotection (2×5 min) and peptide coupling was performed using 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) and N,N-diisopropylethylamine (DIPEA). Deprotection and coupling cycles were repeated for the addition of every amino acid and the final peptide was cleaved from the resin using a solution of trifluoroacetic acid (TFA) and dichloromethane (DCM) at a 9:1 ratio. Peptides were precipitated using cold ether, pelleted by centrifugation at 5,000 rpm for 15 min and dried in vacuo. The peptide was purified by reverse-phase high-performance liquid chromatography (HPLC) using a Hitachi-Elite LaChrom L2130

pump with a binary gradient and a UV-Vis detector and eluted using a linear gradient of acetonitrile in water. Peptide purity was assessed by matrix-assisted laser desorption/ionization (MALDI).

2.5.3 Preparation of modified-CoA conjugates

TAMRA-CoA was synthesized and purified as previously described.²⁸ GFP-CoA was prepared by first exchanging sfGFP into a buffered solution containing 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS) (pH 7.5). Approximately 5 equiv. of sulfosuccinimidyl-4-[N-maleimidomethyl]cyclohexane-1-carboxylate (sulfo-SMCC) dissolved in a solution of 100 μ L DMF was added to a 1 mL solution containing 100 μ M GFP and allowed to react at 37 °C for 1.5 h. The sample was buffer exchanged into a fresh solution of 20 mM MOPS (pH 7.5) using a 10 D/G desalting column to remove any unreacted SMCC and 5 equiv. of coenzyme A, free acid dissolved in a buffered solution containing 20 mM MOPS (pH 7.5) was added and allowed to gently shake overnight at room temperature on an orbital shaker. The sample was buffer exchanged into a fresh solution of 20 mM MOPS (pH 7.5) using a 10 D/G desalting column and characterized by UV-Vis spectroscopy.

2.5.4 Chemical conjugation of self-assembled RIDC3 crystals

RIDC3 crystals were self-assembled by first exchanging RIDC3 into a buffered solution of 20 mM 2-(N-morpholine)ethanesulfonic acid (MES) (pH 5.5) using an Amicon spin filter. A concentrated stock solution of RIDC3 was supplemented with a solution of 20 mM MES (pH 5.5) and 5 equiv. $ZnCl_2$ dissolved in water to a final protein concentration of 100 μ M. After mixing with $ZnCl_2$, the solution turned cloudy within 5 min and crystals matured in 7-10 days. Mature crystals were buffer exchanged \times 5 into a fresh solution containing 20 mM MOPS (pH 7.5) by pelleting suspensions of crystals at 3,500 rpm in 30 s bursts for 2 min and carefully pipetting out the clear supernatant. Approximately 10 equiv. of a dibenzocyclooctyne-N-hydroxysuccinimidyl

ester (DBCO-NHS) dissolved in DMF (25 mg/mL) was added to a suspension of RIDC3 crystals and allowed to gently shake at room temperature for 2 h. Crystals were buffer exchanged $\times 5$ into a fresh solution containing 20 mM MOPS (pH 7.5) via centrifugation to remove any unreacted DBCO. A stock solution of ybbR-N₃ dissolved in DMF was added to the suspension of RIDC3 crystals to a final ratio of 10:1 peptide:crystals and gently shaken overnight at room temperature. RIDC3+ybbR crystals were again buffer exchanged $\times 5$ into a fresh solution containing 20 mM MOPS (pH 7.5) and analyzed by UV-Vis spectroscopy and electrospray ionization mass spectrometry (ESI-MS).

The preparation of RIDC3-CoA crystals was performed similarly to RIDC3+ybbR crystals, starting with buffer exchanging mature crystals into a solution containing 20 mM MOPS (pH 7.5) via centrifugation. Approximately 10 equiv. of a stock solution of sulfo-SMCC dissolved in DMF was added to the suspension of crystals and allowed to gently shake at room temperature for 1.5-2 h. Crystals were buffer exchanged $\times 5$ into a fresh solution containing 20 mM MOPS (pH 7.5) via centrifugation and 10 equiv. of coenzyme A, free acid dissolved in a buffered solution containing 20 mM MOPS (pH 7.5) was added and gently shaken overnight at room temperature. Crystals were buffer exchanged one final time and analyzed by UV-Vis spectroscopy.

2.5.5 Preparation of RIDC3-ybbR and RhuA-ybbR crystals

RIDC3-ybbR solutions were stored at 4 °C to limit degradation of the appended peptide. A stock solution of RIDC3-ybbR was combined with Zn²⁺ in different buffered solutions. In general, a 50- μ L sample contained 100 μ M protein at [Zn²⁺]:[protein] ratios ranging from 1:1 to 10:1. Buffered solutions containing 20 mM MES were used at pH 5.5 and pH 6.5, 20 mM MOPS at pH 7.5 and 20 mM CHES at pH 8.5. Crystalline materials were found in solutions at pH 5.5 and pH 7.5. RhuA-ybbR crystals were prepared by buffer exchanging protein into a fresh solution

containing 10 mM Tris, 10 mM β ME, and 1 mM $ZnCl_2$ using an Amicon spin filter to a final protein concentration of 125-150 μ M RhuA-ybbR tetramer. Samples were gently shaken at 4 °C for 1-2 weeks until crystals matured.

2.5.6 Gel digestion and MS-MS analysis of RIDC3+ybbR crystals

A suspension of RIDC3+ybbR crystals was dissolved using a solution containing 1 mM ethylenediaminetetraacetic acid (EDTA) and separated from any possible contaminants on a SDS-PAGE gel and stained with Coomassie Brilliant Blue. The band corresponding to RIDC3+ybbR was cut out into 1 mm \times 1 mm cubes and destained \times 3 with a solution containing 100 mM ammonium bicarbonate for 15 min followed by the addition of 100 μ L acetonitrile (ACN) for 15 min. The supernatant was removed, and gel pieces were dried in vacuo and chemically reduced with 200 μ L of a solution containing 100 mM ammonium bicarbonate and 10 mM dithiothreitol (DTT) at 55 °C for 30 min. The supernatant was removed, and gel pieces were incubated with 200 μ L of a solution containing 100 mM ammonium bicarbonate and 55 mM iodoacetamide and incubated for 20 min at room temperature in the absence of light. Gel samples were washed first with a fresh solution of 100 mM ammonium bicarbonate and finally acetonitrile prior to dehydrating gel pieces using a SpeedVac. Gel pieces were digested by first covering them in an ice-cold solution containing 50 mM ammonium bicarbonate and trypsin (0.1 μ g/ μ L) and incubated on ice for 30 min. After the gel was completely rehydrated, excess solution was removed and replaced with a fresh solution containing 50 mM ammonium bicarbonate overnight at 37 °C. Peptide extraction was performed by the addition of a 50 μ L solution containing 0.2% formic acid (v/v) and 5% ACN (v/v) in water and mixing at room temperature for 30 min. The supernatant was collected and the extraction procedure was repeated again, combining the 2nd supernatant with the

previous solution. Samples were analyzed by liquid chromatography (LC) with tandem mass spectrometry (MS/MS) using electrospray ionization.

Trypsin-digested peptide solutions were analyzed using ultra high pressure liquid chromatography coupled with LC-MS/MS using nanospray ionization. Ionization experiments were performed using a TripleTof 5600 hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reverse-phase UPLC equipped with a 20-cm 75-micron ID glass capillary packed with 2.5- μm C18 CSHTM beads (Waters corporation). Peptides were eluted using a linear gradient (5-80% ACN) at a flow rate of 250 $\mu\text{L}/\text{min}$ for 1 h. The UPLC solutions used were: Buffer A - 98% H₂O, 2% ACN, 0.1% formic acid, and 0.005% TFA and Buffer B - 100% ACN, 0.1% formic acid, and 0.005% TFA. MS/MS data were acquired in a data-dependent manner; the MS1 data was acquired for 250 ms at m/z of 400 to 1250 Da and the MS/MS data was acquired from m/z of 50 to 2,000 Da. The independent data acquisition (IDA) parameters were as follows: MS1-TOF acquisition time of 250 ms, followed by acquisition of 50 MS2 events of 48 ms for each event. The threshold to trigger a MS2 event was set to 150 counts for ion charge states of +2, +3 and +4. The ion exclusion time was set to 4 s. Finally, the collected data were analyzed using Protein Pilot 5.0 (ABSCIEX) for peptide identifications.

2.5.7 Enzymatic labeling procedure

Samples for enzymatic labeling were buffer exchanged via centrifugation (for crystal suspensions) or using Amicon spin filters (for soluble protein solutions) into a solution containing 20 mM MOPS (pH 7.5). Crystal suspensions were dissolved via the addition of a stock solution of EDTA for a final concentration of 1 mM and buffer exchanged $\times 5$ for solution experiments. In general, a 20- μL reaction consisted of 10 μM Sfp PPTase, 15 mM MgCl₂, 100-200 μM modified-CoA and 50 μM protein-ybbR conjugate in a buffered solution containing 20 mM MOPS (pH 7.5).

The solution was gently shaken for 16-24 h at room temperature. Completed reactions were buffer exchanged $\times 5$ via centrifugation to remove unbound dye and enzyme into a solution containing 20 mM MOPS (pH 7.5).

AcpH treatment of labeled solutions was performed similarly. In general, a 20- μ L reaction consisted of 10 μ M AcpH, 15 mM MgCl₂, 1 mM MnCl₂ and 50 μ M labeled protein in a buffered solution containing 20 mM Tris (pH 8.0). Samples were allowed to react for 16 h at 37 °C. Samples were buffer exchanged as previously described into a fresh solution of 20 mM MOPS (pH 7.5). PAGE gels were visualized on a Typhoon TRIO Variable Mode Imager (GE Healthcare Biosciences) at 50- μ m resolution using a 532 nm excitation wavelength and 580 nm emission filter and photomultiplier tube setting of 500.

2.5.8 Confocal microscopy

A 5- μ L suspension of crystals was pipetted onto a glass slide and covered with a cover slip, sealing the edges with clear nail polish to prevent sample drying. Samples were imaged with a 100x oil objective on a spinning-disk confocal Zeiss Axio Observer inverted microscope equipped with a pair of Roper Quantum 5125C cameras. Samples were excited at 488 nm for green fluorescence and 564 nm for red fluorescence. Differential interference contrast and fluorescence images were captured at 1-s and 100-ms exposures, respectively. Images were collected in Slidebook 6 (Intelligent Imaging Innovations) and analyzed using Fiji (<http://fiji.sc/Fiji>).

2.5.9 Negative-stain transmission electron microscopy (ns-TEM)

A 3- μ L suspension of crystals was pipetted onto formvar/carbon-coated Cu grids (Ted Pella, Inc.) that had been glow discharged for 45-60 s. Samples were incubated for 5 min, washed with 50 μ L of MilliQ water and blotted with filter paper. A 3- μ L drop of 1% uranyl acetate in water was pipetted onto the grid and incubated for 30-60 s and blotted dry with filter paper. Grids

were imaged using a FEI Sphera transmission electron microscope operating at 200 keV, equipped with a LaB6 filament and a Gatan 4K charged-coupled device (CCD). Micrographs were collected using objective-lens underfocus settings ranging from 250 nm to 2 μ m and analyzed using Fiji (<http://fiji.sc/Fiji>).

2.5.10 Atomic force microscopy (AFM)

A 10- μ L suspension of crystals was deposited onto freshly cleaved mica (Ted Pella, Inc.) and incubated for 10 min. The mica disc was gently dried using a stream of nitrogen with care not to push the drop over the edge of the disc. AFM measurement were performed on a Bruker Dimension Icon ScanAsyst atomic force microscope using a ScanAsyst-Air tip (Bruker) operating in tapping mode. Images were analyzed using NanoScope Analysis (v.1.5, Bruker).

2.5.11 Fluorescence microplate reader measurements

Fluorescence measurements were performed using a 96-well plate (Falcon) containing 50- μ L solutions of each sample. Excitation/emission wavelengths of 485/510 nm and 557/583 nm were used for green and red fluorescence respectively with a 2 nm slit width, a 0.2 s integration time and a gain of 100.

2.6 Acknowledgments

Rohit Subramanian was supported by the National Institute of Health Chemical Biology Interfaces Training Grant UC San Diego (T32GM112584-01). ns-TEM data were collected at the UCSD EM facilities supported by funding to T. S. B. from the NIH (R01-GM033050) and the Agouron Foundation.

Chapter 2 is reproduced, in part, from a manuscript currently being prepared for submission: Subramanian, R. H.; Suzuki, Y.; Tallorin, L.; Sahu, S.; Tezcan, F.A. Enzyme-directed Modification of Two-Dimensional Crystalline Protein Lattices.

The dissertation author is primary author on all reprinted materials.

2.7 References

- (1) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551-575.
- (2) Goodsell, D. S.; Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Struct.* **2000**, *29*, 105-153.
- (3) Churchfield, L. A.; Tezcan, F. A. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **2019**, *52* (2), 345-355.
- (4) Song, W. J.; Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014**, *346* (6216), 1525-1528.
- (5) Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **2017**, *46* (1), 23-42.
- (6) 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. et al. Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **2017**, *552* (7685), 415-420.
- (7) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto Jr., G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem., Int. Ed.* **2005**, *44* (45), 7342-7372.
- (8) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep* **2014**, *31* (1), 61-108.
- (9) Paulick, M. G.; Bertozzi, C. R. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry* **2008**, *47* (27), 6991-7000.
- (10) Fagan, R. P.; Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* **2014**, *12* (3), 211-222.
- (11) Khoury, G. A.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.* **2014**, *32* (2), 99-109.
- (12) Sleytr, U. B.; Schuster, B.; Egelseer, E. M.; Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **2014**, *38* (5), 823-864.

- (13) Baneyx, F.; Mattheai, J. F. Self-assembled two-dimensional protein arrays in bionanotechnology: from S-layers to designed lattices. *Curr. Opin. Biotech.* **2014**, *28*, 39-45.
- (14) Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS. Chem. Biol.* **2006**, *1* (10), 644-648.
- (15) Jonkheijm, P.; Weinrich, D.; Schroder, H.; Niemeyer, C. M.; Waldmann, H. Chemical strategies for generating protein biochips. *Angew. Chem., Int. Ed.* **2008**, *47* (50), 9618-9647.
- (16) Marraffini, L. A.; Dedent, A. C.; Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **2006**, *70* (1), 192-221.
- (17) Kwon, Y.; Coleman, M. A.; Camarero, J. A. Selective immobilization of proteins onto solid supports through split-intein-mediated protein trans-splicing. *Angew. Chem., Int. Ed.* **2006**, *45* (11), 1726-1729.
- (18) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (12), 690-697.
- (19) Pishesha, N.; Ingram, J. R.; Ploegh, H. L. Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol.* **2018**, *34*, 163-188.
- (20) Schmohl, L.; Schwarzer, D. Sortase-mediated ligations for the site-specific modification of proteins. *Curr. Opin. Chem. Biol.* **2014**, *22*, 122-128.
- (21) Shadish, J. A.; Benuska, G. M.; DeForest, C. A. Bioactive site-specifically modified proteins for 4D patterning of gel biomaterials. *Nat. Mater.* **2019**, *18* (9), 1005-1014.
- (22) Charrier, M.; Li, D.; Mann, V. R.; Yun, L.; Jani, S.; Rad, B.; Cohen, B. E.; Ashby, P. D.; Ryan, K. R.; Ajo-Franklin, C. M. Engineering the S-Layer of *Caulobacter crescentus* as a Foundation for Stable, High-Density, 2D Living Materials. *ACS Synth. Biol.* **2019**, *8* (1), 181-190.
- (23) Manea, F.; Garda, V. G.; Rad, B.; Ajo-Franklin, C. M. Programmable assembly of 2D crystalline protein arrays into covalently stacked 3D bionanomaterials. *Biotechnol. Bioeng.* **2019**.
- (24) Brodin, J. D.; Ambroggio, X. I.; Tang, C. Y.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **2012**, *4* (5), 375-382.
- (25) 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*, 369.

- (26) 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.
- (27) Sinclair, J. C.; Davies, K. M.; Venien-Bryan, C.; Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **2011**, *6* (9), 558-562.
- (28) Mattheai, J. F.; DiMaio, F.; Richards, J. J.; Pozzo, L. D.; Baker, D.; Baneyx, F. Designing Two-Dimensional Protein Arrays through Fusion of Multimers and Interface Mutations. *Nano. Lett.* **2015**, *15* (8), 5235-5239.
- (29) Thomas, A.; Mattheai, J. F.; Baneyx, F. A Self-Assembling Two-Dimensional Protein Array is a Versatile Platform for the Assembly of Multicomponent Nanostructures. *Biotechnol. J.* **2018**, *13* (12), 1800141.
- (30) Du, M.; Zhou, K.; Wang, X.; Zhang, J.; Zhang, Y.; Dong, J.; Wu, L.; Qiao, Z.; Chen, G.; Wang, Q. Precise Fabrication of De Novo Nanoparticle Lattices on Dynamic 2D Protein Crystalline Lattices. *Nano Lett.* **2020**, *20* (2), 1154-1160.
- (31) Brodin, J. D.; Carr, J. R.; Sontz, P. A.; Tezcan, F. A. Exceptionally stable, redox-active supramolecular protein assemblies with emergent properties. *Proc. Natl. Acad. Sci. USA* **2014**, *111* (8), 2897-2902.
- (32) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (44), 15815-15820.
- (33) Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **2006**, *1* (1), 280-285.
- (34) Kosa, N. M.; Pham, K. M.; Burkart, M. D. Chemoenzymatic exchange of phosphopantetheine on protein and peptide. *Chem. Sci.* **2014**, *5* (3), 1179-1186.
- (35) Tallorin, L.; Wang, J.; Kim, W. E.; Sahu, S.; Kosa, N. M.; Yang, P.; Thompson, M.; Gilson, M. K.; Frazier, P. I.; Burkart, M. D. et al. Discovering de novo peptide substrates for enzymes using machine learning. *Nat. Commun.* **2018**, *9* (1), 5253.
- (36) Grunewald, J.; Klock, H. E.; Cellitti, S. E.; Bursulaya, B.; McMullan, D.; Jones, D. H.; Chiu, H. P.; Wang, X.; Patterson, P.; Zhou, H. et al. Efficient Preparation of Site-Specific Antibody-Drug Conjugates Using Phosphopantetheinyl Transferases. *Bioconjug. Chem.* **2015**, *26* (12), 2554-2562.
- (37) Wong, L. S.; Okrasa, K.; Micklefield, J. Site-selective immobilisation of functional enzymes on to polystyrene nanoparticles. *Org. Biomol. Chem.* **2010**, *8* (4), 782-787.
- (38) Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. Rapid Cu-Free Click Chemistry with Readily Synthesized Biarylazacyclooctynones. *J. Am. Chem. Soc.* **2010**, *132* (11), 3688-3690.

- (39) Kotagiri, N.; Li, Z.; Xu, X.; Mondal, S.; Nehorai, A.; Achilefu, S. Antibody quantum dot conjugates developed via copper-free click chemistry for rapid analysis of biological samples using a microfluidic microsphere array system. *Bioconjug. Chem.* **2014**, *25* (7), 1272-1281.
- (40) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. In *Methods in Enzymology*; Pecoraro, V. L., Ed.; Academic Press, **2016**; Vol. 580.
- (41) Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24* (1), 79-88.

Chapter 3: Self-assembly of a Designed Nucleoprotein Architecture Through Multimodal Interactions

3.1 Abstract

The co-self-assembly of proteins and nucleic acids (NAs) produces complex biomolecular machines (*e.g.*, ribosomes and telomerases) that represent some of the most daunting targets for biomolecular design. Despite significant advances in protein and DNA or RNA nanotechnology, the construction of artificial nucleoprotein complexes has largely been limited to cases that rely on the NA-mediated spatial organization of protein units, rather than a cooperative interplay between protein- and NA-mediated interactions that typify natural nucleoprotein assemblies. We report here a structurally well-defined synthetic nucleoprotein assembly that forms through the synergy of three types of intermolecular interactions: Watson-Crick base pairing, NA-protein interactions, protein-metal coordination. The fine thermodynamic balance between these interactions enables the formation of a crystalline architecture under highly specific conditions.

3.2 Introduction

Modularity is a key driver of biological complexity, whereby nature employs a limited set of building blocks to create strikingly diverse machines and materials. Hence, a prominent goal of synthetic biology and nanotechnology is to use biological or biomimetic building blocks (*e.g.*, DNA or RNA, proteins, peptides, foldamers) in new structural contexts to construct self-assembled systems with properties that complement or surpass those produced by evolution. The emergence of the field of bio-nanotechnology has largely been fueled by the high fidelity of Watson-Crick base-pairing interactions, which have enabled the fabrication of DNA and RNA assemblies with complex structures that extend far beyond the natural scope of nucleic acids (NAs)¹⁻⁷. However,

the inherent functional properties of these assemblies are restricted by the limited chemical scope of NAs. In parallel, there have been considerable efforts toward the design of artificial protein assemblies⁸⁻¹³, aiming to take advantage of the chemical and structural versatility of proteins. Synthetic protein assemblies have become sophisticated in terms of their architectures and functional attributes¹⁴⁻¹⁷, yet they still lack the facile structural programmability of NA-based assemblies. The complementary advantages and limitations of NA and protein building blocks prompt the obvious question: can proteins and NAs be combined modularly to form well-defined supramolecular assemblies?

Natural nucleoprotein assemblies such as ribosomes¹⁸, telomerases¹⁹, nucleosomes²⁰, and gene-editing complexes²¹ are inspiring examples in which protein and NA components are integrated into highly intricate yet monodisperse architectures. The disparate and non-uniform structures of these hybrid architectures highlight the complexity of the problem in hand: chemically and structurally, NAs and proteins are vastly dissimilar polymers and there are no universal rules that dictate the interactions between them. In fact, synthetic approaches to create NA-protein hybrid architectures have generally relied on the immobilization of oligonucleotide-functionalized proteins or consensus NA-recognizing peptide sequences onto NA scaffolds²²⁻²⁷ or the modification of proteins with long oligonucleotide chains to enable their self-assembly via base-pairing interactions²⁸⁻³⁰. While these strategies are effective for the spatial organization of protein units, the resulting architectures are mostly dominated by the NA components in terms of composition and mode of self-assembly, and, unlike natural nucleoproteins, they possess little synergy involving protein-protein or protein-NA interactions. Recently, a hybrid system consisting of an engineered homodimer of a DNA-binding protein and cognate double-stranded DNA

(dsDNA) sequences was reported³¹, but the resulting assemblies were limited to linear fibers devoid of substantial structural order.

3.3 Results and Discussion

3.3.1 Design and self-assembly of DNA-protein chimeras

In this study, we set out to create a synthetic nucleoprotein system whose self-assembly is governed by a cooperative interplay of protein- and DNA-mediated interactions. As the protein component, we chose RIDC3 (an engineered variant of the monomeric, four-helix bundle protein cytochrome *cb*₅₆₂, **Figure 3.1a**) which can self-assemble in a chemically tunable fashion into 1-, 2-, and 3D crystalline arrays through a combination of designed Zn²⁺-mediated and non-covalent interactions¹². RIDC3 was modified with a single cysteine residue at position 21 (C21) on a surface loop to enable covalent conjugation to single-stranded (ss) DNA sequences (**Figure 3.1a**)³². We selected the DNA components to be (a) short, (<15 bases, melting temperatures (T_m) of 30 °C or below) such that they would approximately match the dimensions of the protein component and yield weak base-pairing interactions to promote self-assembly under thermodynamic control, and (b) non-self-associating, such that self-assembly is heterogenic, requiring two DNA-protein chimeras with complementary sequences.

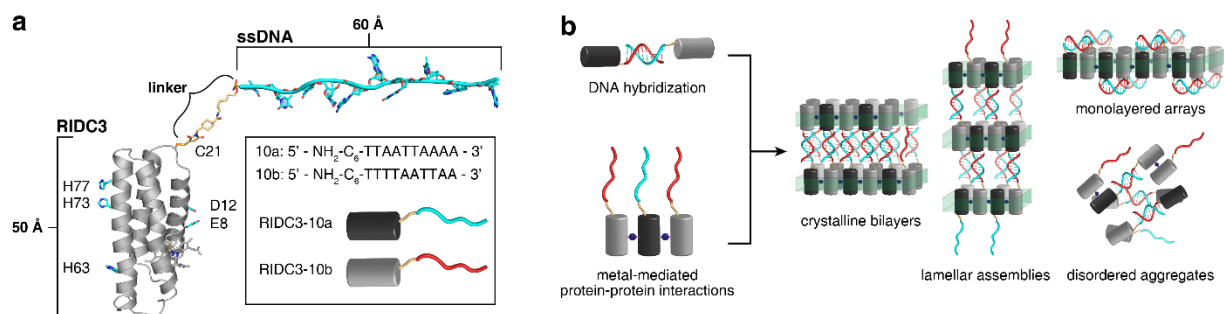


Figure 3.1 Design of RIDC3-10a/10b. **(a)** Schematic of an RIDC3-DNA hybrids. Metal-binding residues are shown as cyan sticks on the RIDC3 surface. **(b)** Potential self-assembly outcomes for the RIDC3-DNA hybrids.

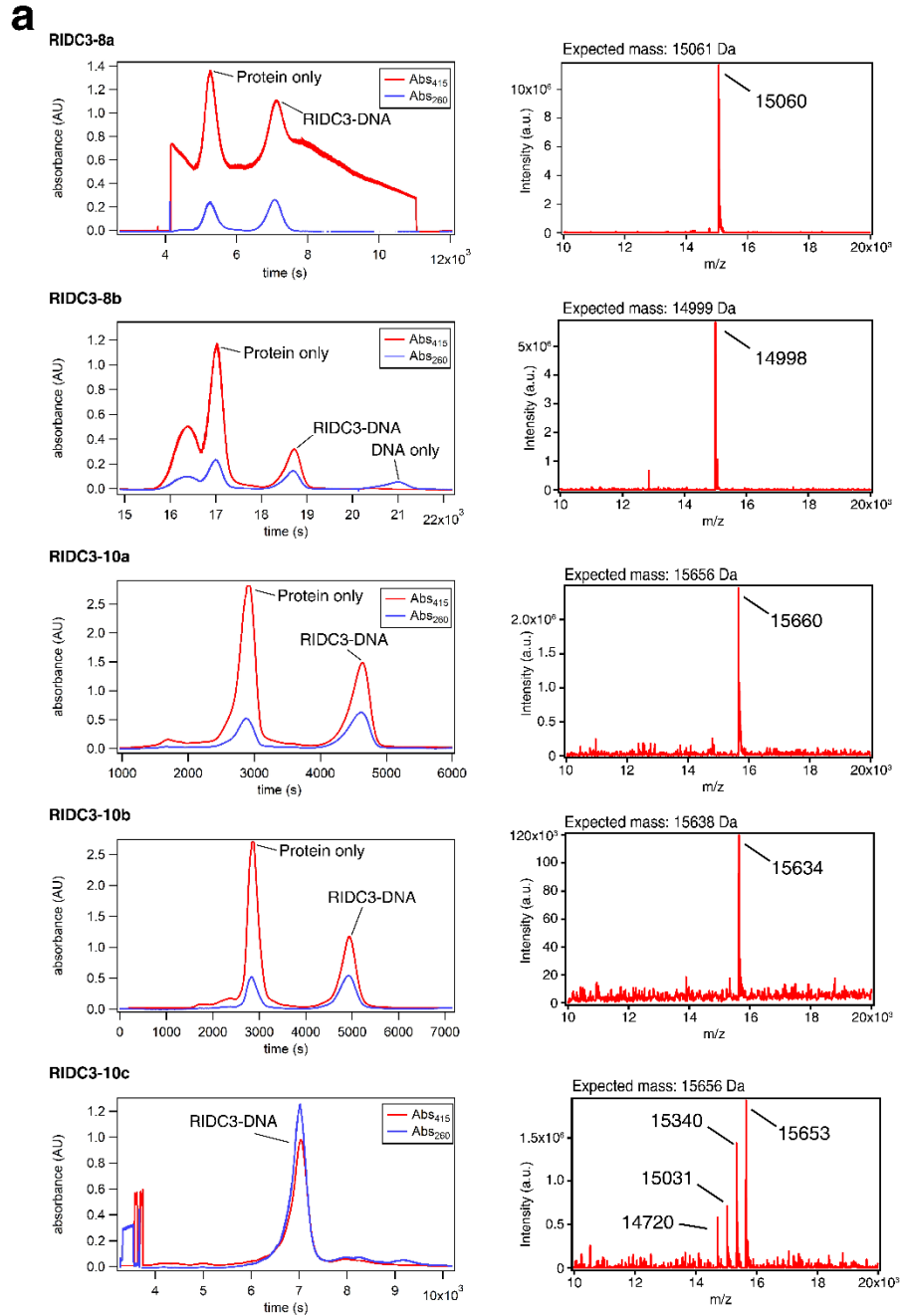


Figure 3.2 FPLC chromatograms and ESI-MS spectra of RIDC3-DNA conjugates. **(a)** Data for conjugates with varying DNA strand lengths. Spectra are omitted for samples which were undetected by MS analysis. **(b)** Data for alanine scanning point mutants of RIDC3-10a/10b.

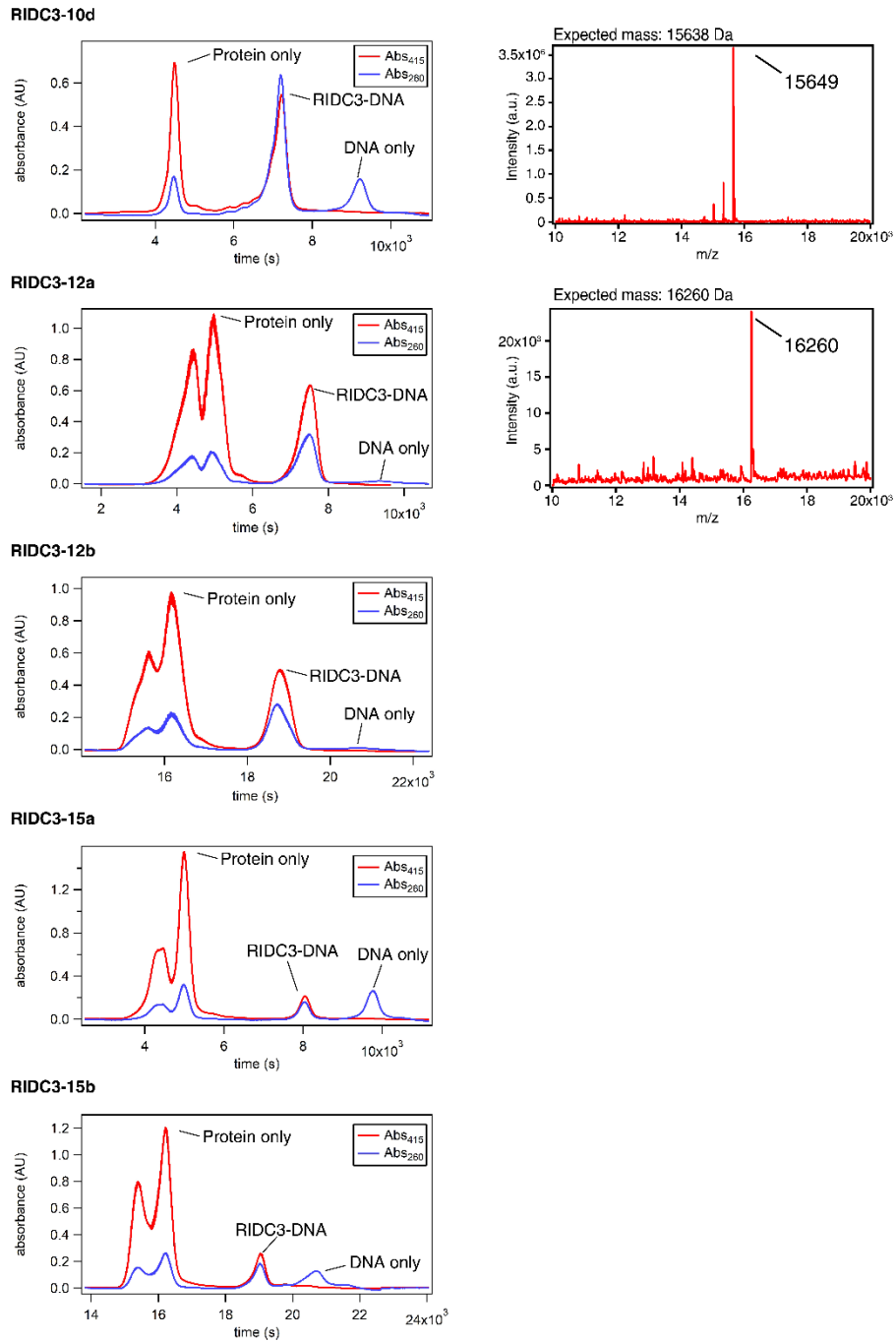


Figure 3.2 FPLC chromatograms and ESI-MS spectra of RIDC3-DNA conjugates, continued

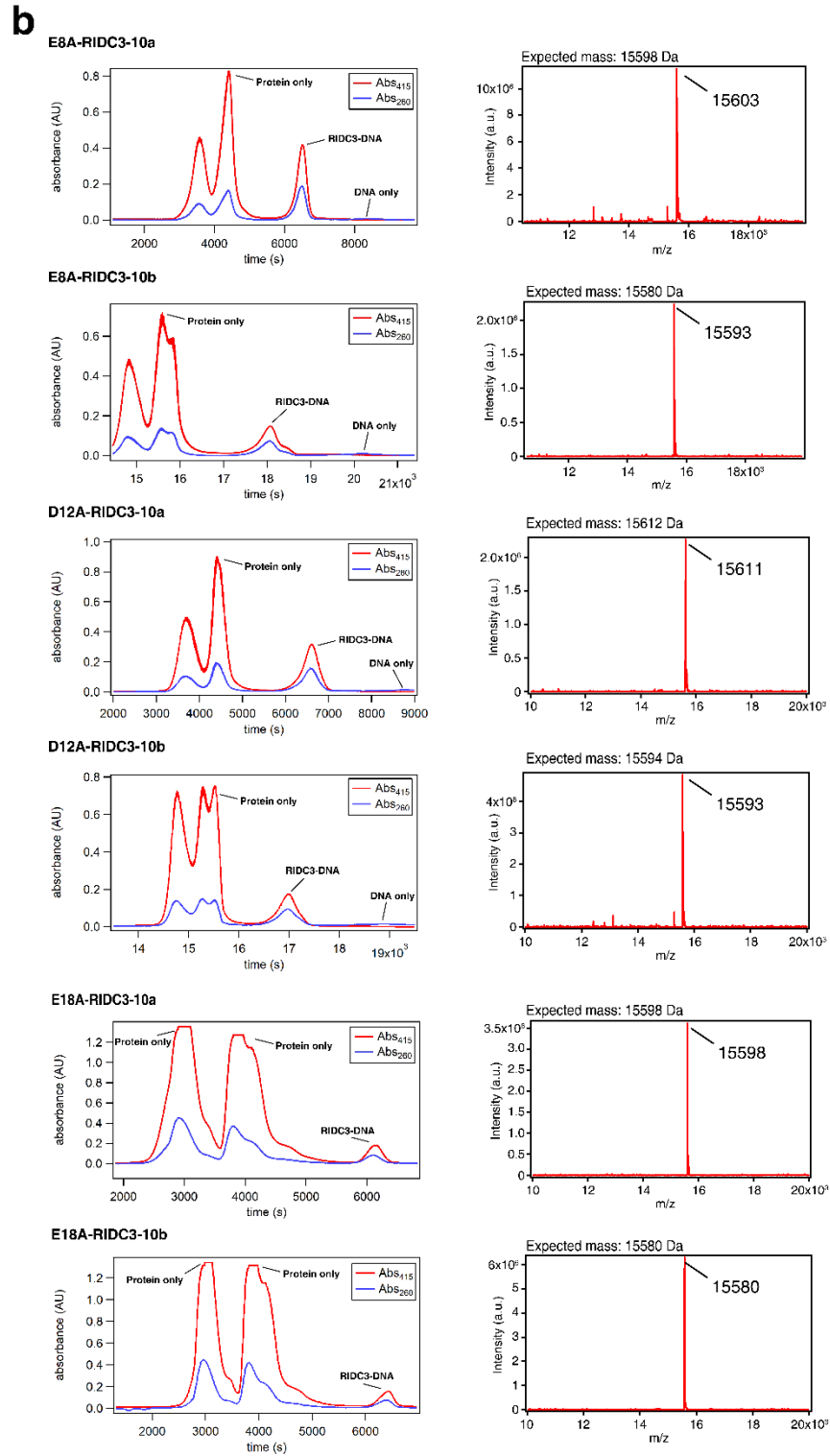


Figure 3.2 FPLC chromatograms and ESI-MS spectra of RIDC3-DNA conjugates, continued

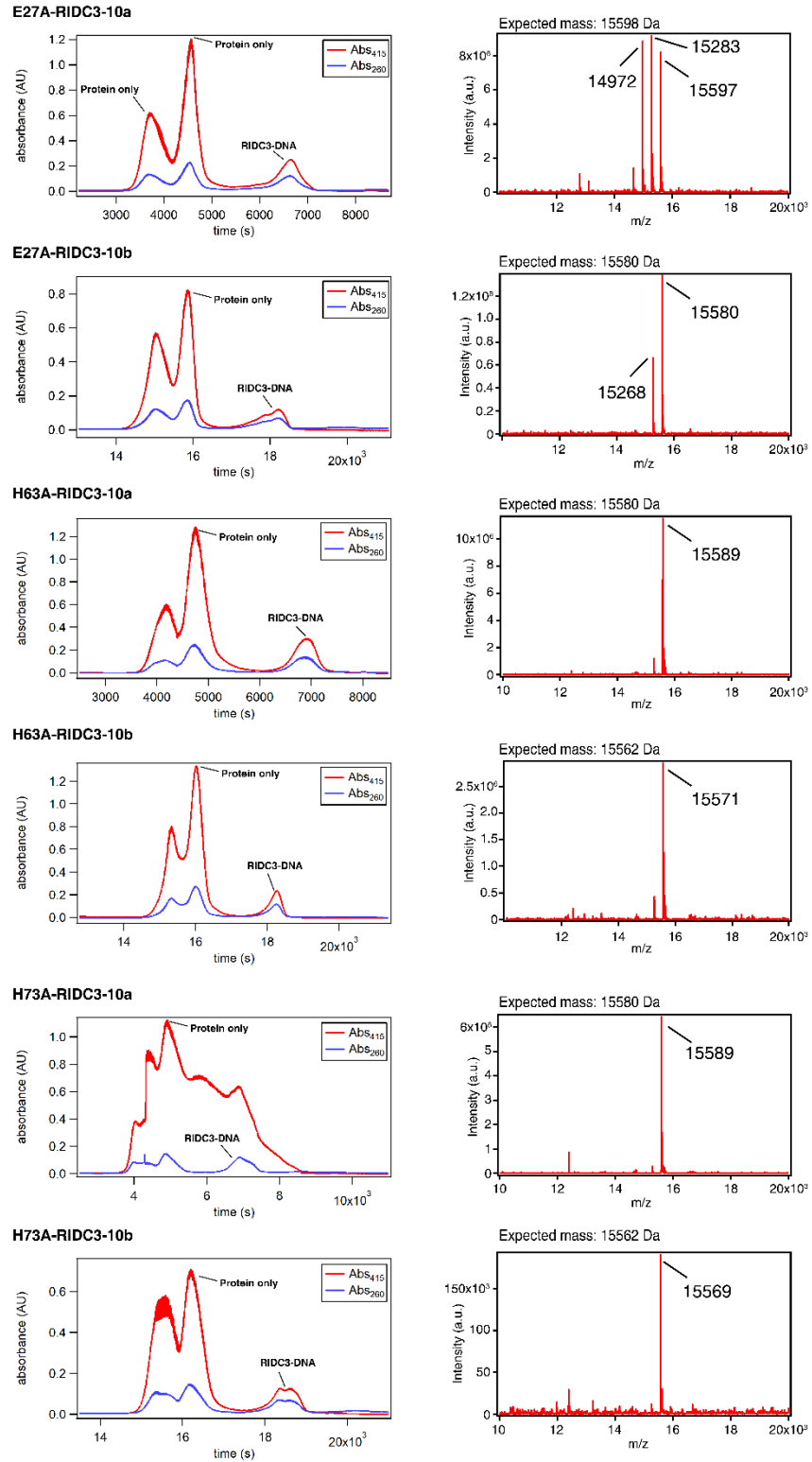


Figure 3.2 FPLC chromatograms and ESI-MS spectra of RIDC3-DNA conjugates, continued

^{c21}RIDC3 was quantitatively modified with complementary, non-slipping 10-bp DNA strands (10a and 10b; $T_{m, \text{pred}} \approx 7 \text{ }^\circ\text{C}$) to generate two hybrids, RIDC3-10a and RIDC3-10b (**Figure 3.2**). As depicted in Figure 3.1, RIDC3-10a and RIDC3-10b are essentially diblock copolymers (**Figure 3.1a**) with two domains whose self-assembly can be modulated by distinct stimuli and design parameters: DNA components primarily by oligonucleotide sequence and temperature, and protein components primarily by Zn^{2+} concentration and pH (to control the protonation of metal-coordinating residues). Given the proclivity of the RIDC3 components to form 2D layers^{12,33}, we envisioned that there would be multiple potential outcomes for self-assembly based on the interplay of DNA- and protein-mediated interactions, including crystalline bilayers, lamellar

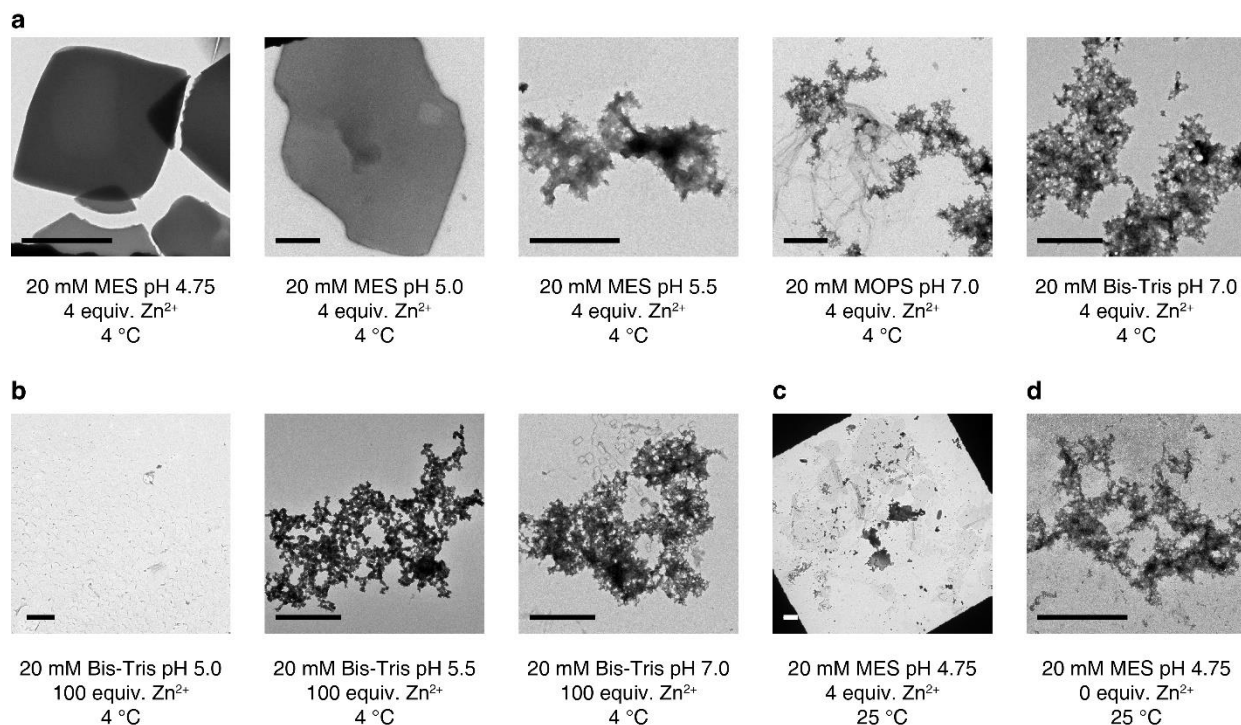


Figure 3.3 TEM characterization of RIDC3-10a/b self-assembly under various solution conditions. Representative ns-TEM micrographs are shown for equimolar (25 μM) concentrations of RIDC3-10a and RIDC3-10b after incubation in the above solution conditions for at least one day. **(a)** Ordered lattices were only observed at pH 4.75 – 5, with 4 equivalents Zn^{2+} , 4 °C. Disordered aggregates were observed at higher pH values. **(b)** Excessive quantities of Zn^{2+} led invariably to aggregation. **(c)** Temperatures significantly higher than the T_m of 10a/10b (ca. 7 °C) failed to yield ordered lattices both under ideal solution conditions, and **(d)** in the absence of metal. Scale bars are 2.5 μm .

assemblies, DNA-stapled monolayered arrays, and disordered aggregates (**Figure 3.1b**). To investigate the formation of ordered assemblies, we screened a variety of conditions (temperature, pH, absolute metal and protein concentrations, metal:protein molar ratios) using equimolar RIDC3-10a and RIDC3-10b (**Figure 3.3**). Ordered protein-DNA conjugate arrays (thin, μm -sized crystals) were observed only in a very narrow window of conditions: pH ranging from 4.75 to 5, temperatures of 4 to 10 $^{\circ}\text{C}$, and a molar ratio of $[\text{Zn}^{2+}]:[\text{total protein-DNA chimeras}]$ between 2 and 10 (**Figure 3.4a**). Alternative conditions lying outside these windows typically gave rise to small (<100 nm in dimensions) clusters of disordered aggregates which did not scatter visible light (**Figure 3.4b**). Solutions lacking either chimera or Zn^{2+} did not yield any self-assembly products. Likewise, addition of the metal chelator ethylenediaminetetraacetic acid (EDTA) to suspensions of the RIDC3-10a/b crystals or their incubation at >40 $^{\circ}\text{C}$ immediately led to the dissolution of the crystals (**Figure 3.5**), implicating the involvement of both Zn^{2+} and DNA-mediated interactions in self-assembly. A Job's analysis indicated that lattice formation was maximized at equal

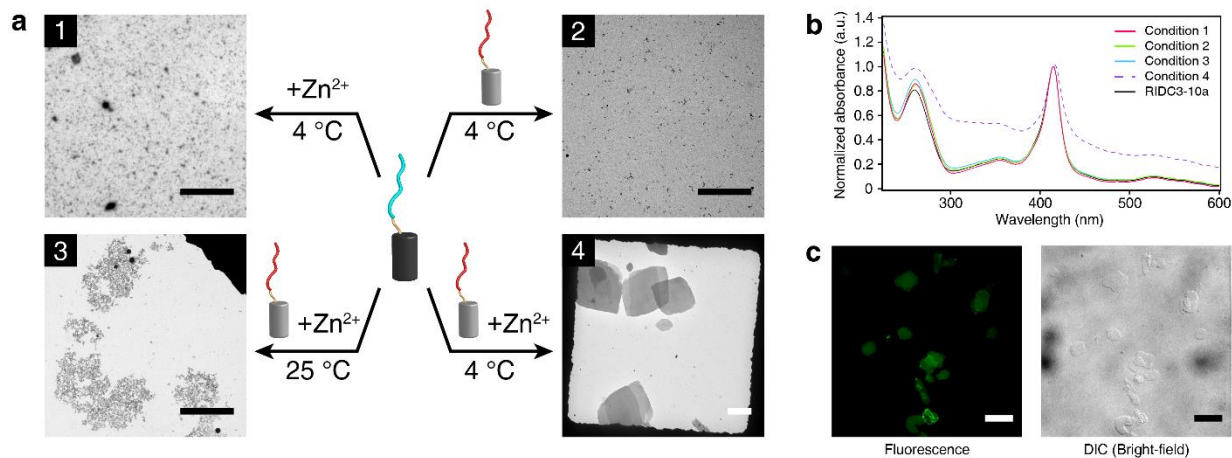


Figure 3.4 Initial characterization of self-assembled RIDC3-10a/b lattices. **(a)** TEM characterization of RIDC3-10a/10b self-assembly under different conditions. **(b)** UV-visible characterization of RIDC3-10a/10b self-assembly under conditions shown in a. Scattering of UV-visible light is only observed for condition 4, indicating the formation of μm -scale structures. **(c)** Confocal microscopy images of RIDC3-10a/b crystals upon intercalation by SYBR Green. Scale bars in a and c are 5 μm and 10 μm , respectively.

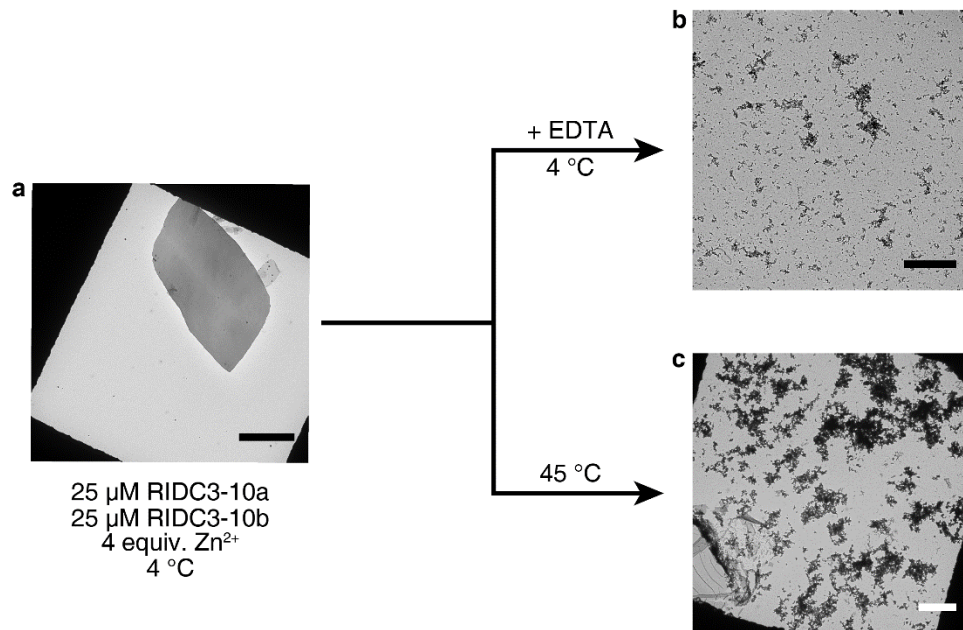


Figure 3.5 Thermal and EDTA-mediated disassembly of RIDC3-10a/b lattices. RIDC3-10a/b crystals (**a**) can be disassembled through the elimination of metal-mediated protein-protein interactions by the addition of EDTA (**b**) or through inhibition of 10a/10b complexation by incubation at 45 °C (**c**). Scale bars: (**a**) 10 μm , (**b**) 1 μm , (**c**) 5 μm .

concentrations of RIDC3-10a and RIDC3-10b, as calculated by a quantitative analysis of protein incorporation into lattices (**Figure 3.6**). RIDC3-10a/b lattices became intensely fluorescent upon treatment with a cyanine dye that specifically binds dsDNA, providing strong evidence that dsDNA is an integral component of the architecture (**Figure 3.4c**, **Figure 3.7**).

Zn- and DNA co-mediated self-assembly behavior of RIDC3-10a/b differs substantially from the Zn-mediated self-assembly of unmodified RIDC3, which yields crystalline arrays in a wide range of conditions, including pH values ranging from 5 to 8.5, temperatures ranging from 4 to 25 °C and Zn:protein molar ratios between 2 and 100¹². Whereas unmodified RIDC3 crystals took at least a day to emerge and a week to fully mature¹², RIDC3-10a/b crystals formed quite rapidly with ordered assemblies appearing within 10 min and growing to μm -dimensions within 4

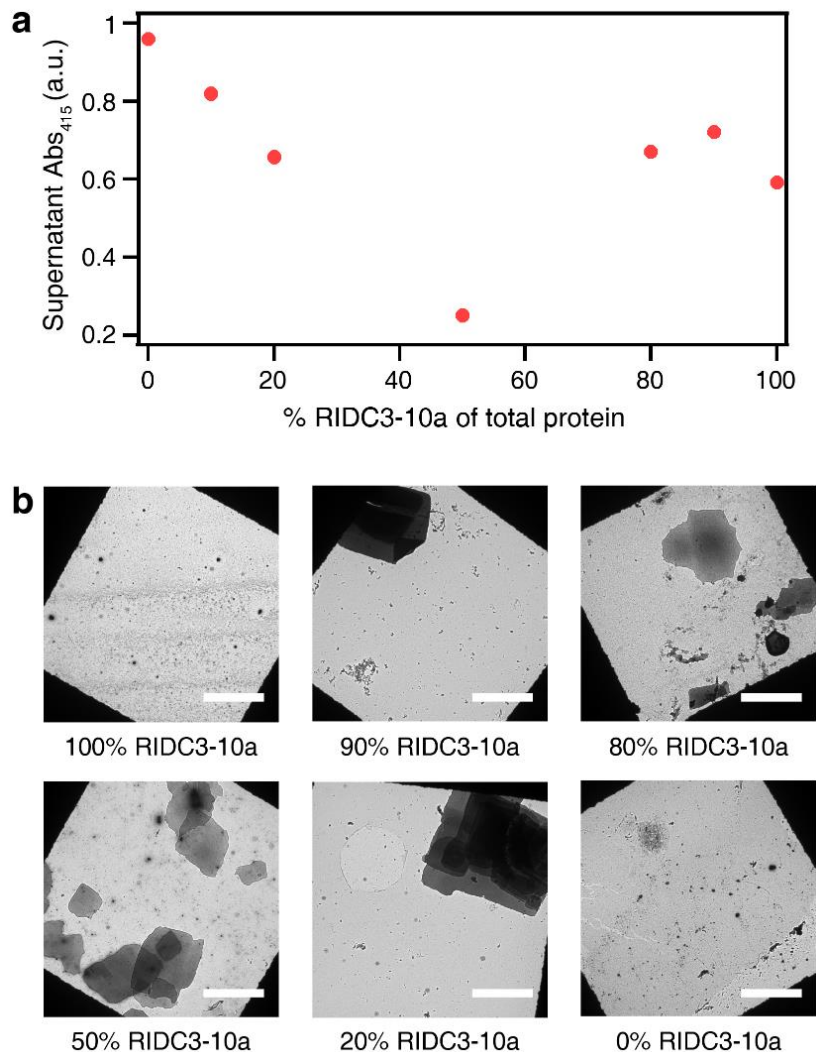


Figure 3.6 Determination of RIDC3-DNA stoichiometry in RIDC3-10a/b assemblies. **(a)** Determination of the maximal incorporation of RIDC3-DNA monomers into RIDC3-10a/10b crystals by Job's method, quantified by UV-Vis spectroscopy. Solutions were prepared at varying ratios of RIDC3-10a:RIDC3-10b under ideal assembly conditions (50 μ M total protein, 20 mM MES (pH 4.75), 4 equiv. Zn^{2+} , 4 $^{\circ}C$). Samples were allowed to incubate for 1 day prior to precipitation of RIDC3-10a/b crystals by centrifugation. Absorbance measurements of unincorporated protein remaining in solution after self-assembly are plotted against the % RIDC3-10a of total protein in solution. This quantity is lowest at equimolar ratios of RIDC3-10a and RIDC3-10b, indicating a 1:1 stoichiometry within the crystals. **(b)** Representative ns-TEM micrographs at varying ratios of RIDC3-10a:RIDC3-10b. Scale bars are 10 μ m.

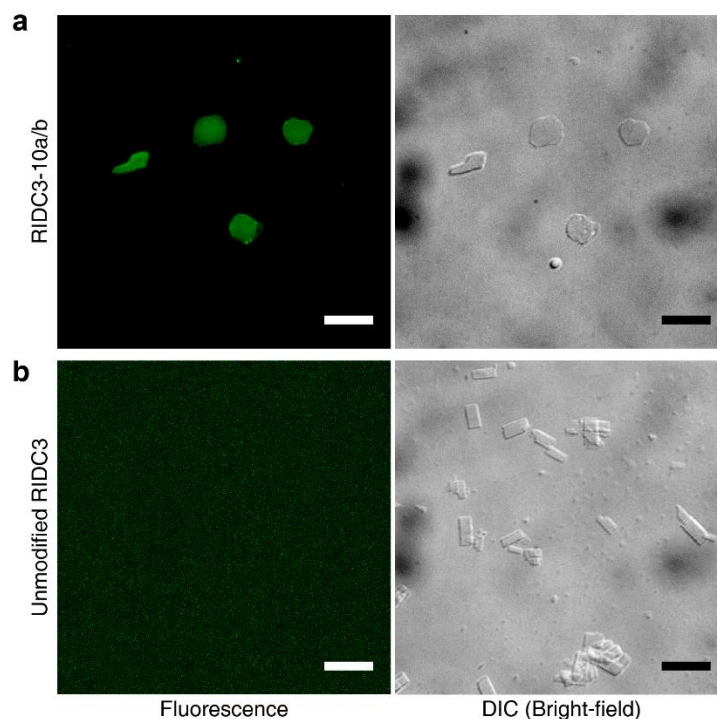


Figure 3.7 Fluorescent dye intercalation into RIDC3-10a/b assemblies. False-color confocal fluorescence images and DIC images are shown above. **(a)** RIDC3-10a/b crystals incubated with the DNA intercalator SYBR Green fluorescence brightly. **(b)** In contrast, unmodified RIDC3 crystals incubated with SYBR Green show no fluorescence. Scale bars are 10 μm .

h (**Figure 3.8a**). Negative-stain TEM (ns-TEM) measurements of RIDC3-10a/b lattices revealed a pgg plane group symmetry with unit cell parameters $a = 60 \text{ \AA}$, $b = 59 \text{ \AA}$, $\alpha = 90^\circ$ in contrast to the pg symmetric RIDC3 crystals with $a = 37 \text{ \AA}$, $b = 137 \text{ \AA}$, $\alpha = 90^\circ$ (**Figure 3.9**)¹². Taken together, these observations pointed to a fundamental alteration and increased chemical stringency in RIDC3 self-assembly through the inclusion of DNA interactions, and a new underlying molecular structure.

3.3.2 Structural characterization of RIDC3-10a/b lattices

Scanning electron microscopy (SEM) and atomic force microscopy (AFM) images of RIDC3-10a/b crystals showed that they consisted of ~ 4 -nm thick 2D layers that stacked up to a height of ca. 100 nm in 3D (**Figure 3.8b and 3.8c**; **Figure 3.10**). The small size and volume of the

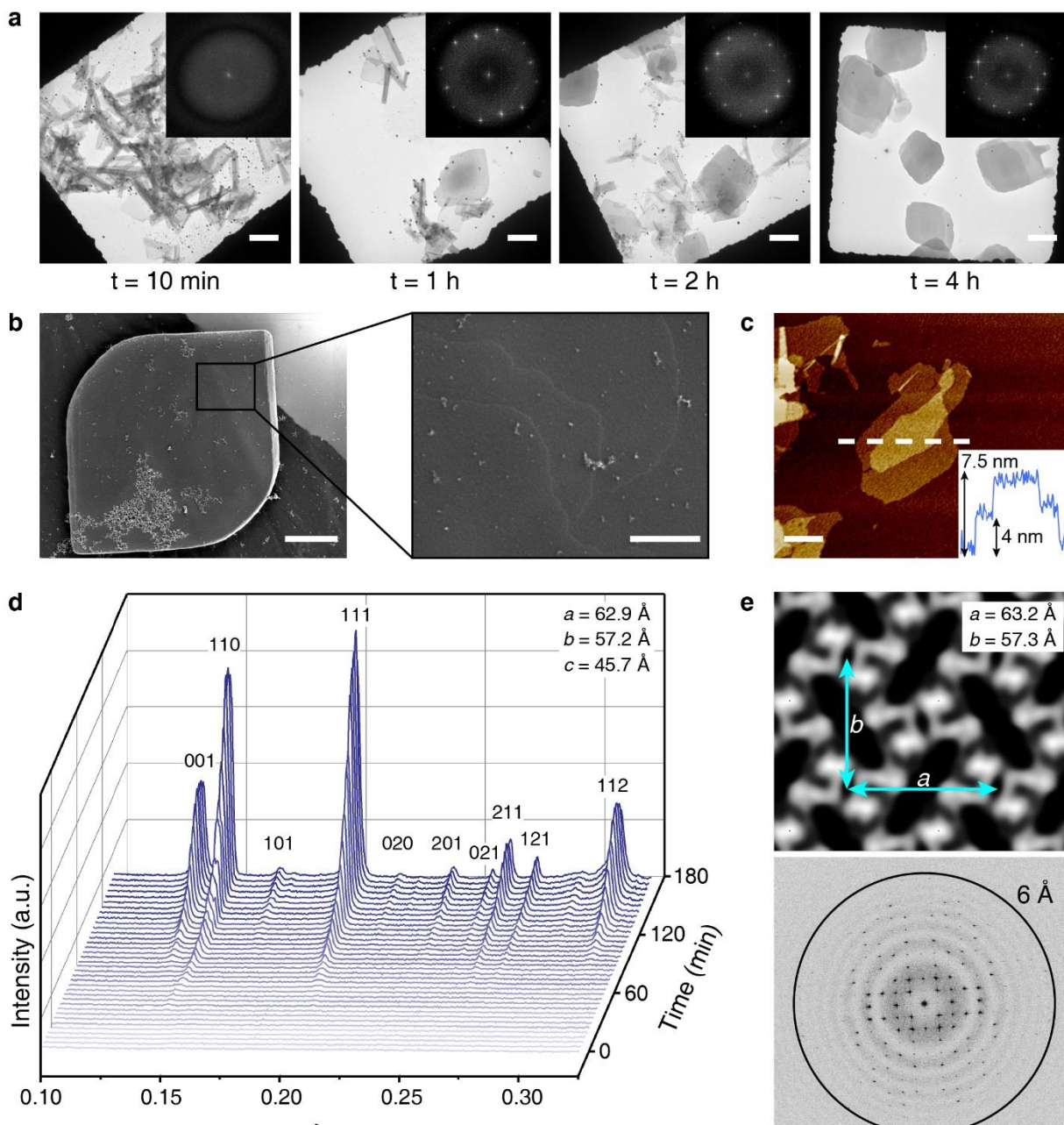


Figure 3.8 Characterization of the crystalline RIDC3-10a/b architecture. **(a)** Time-dependence of RIDC3-10a/b self-assembly as monitored by ns-TEM. Computed FFTs are shown as insets. **(b)** SEM micrograph of an RIDC3-10a/b crystal. Inset: magnification of boxed region to highlight the stacking of individual layers. Scale bars: 2 μm (image) and 500 nm (inset). **(c)** AFM image and height profile of two RIDC3-10a/b layers. Scale bar: 500 nm. **(d)** Time-dependent SAXS profiles of RIDC3-10a/b self-assembly. Miller indices are shown above the Bragg peaks. **(e)** Reconstructed 2D cryo-EM map (top) and computed FFT (bottom) of the RIDC3-10a/b architecture.

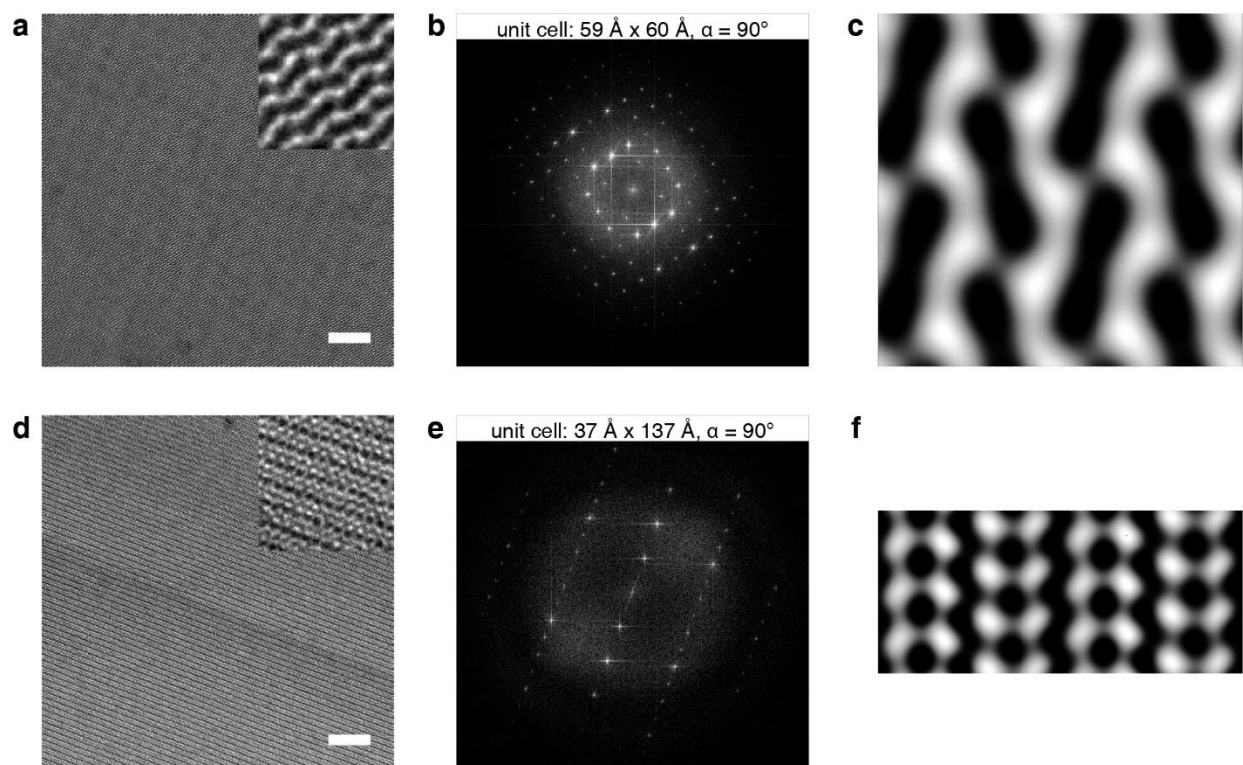


Figure 3.9 ns-TEM characterization of RIDC3-10a/b and RIDC3 assemblies. **(a)** A representative high-magnification ns-TEM micrograph of an RIDC3-10a/b crystal. **(b)** A computed fast Fourier transform (FFT) of **(a)** reveals a well-defined diffraction pattern. Lattice parameters calculated from this FFT are 59 Å x 60 Å, $\alpha = 90^\circ$. **(c)** The 15-Å resolution 2D projection map of RIDC3-10a/10b crystals calculated from ns-TEM micrographs. **(d)** A representative high-magnification image of a native RIDC3 lattice, with its corresponding FFT and lattice parameters **(e)**, and 15-Å resolution 2D projection map **(f)**. Visual inspection reveals significant differences in unit cell and overall connectivity between protein building blocks. Insets in an additional 20x **(a)** and 10x **(d)** magnification of underlying image. Scale bars are 50 nm.

crystals led to weak X-ray diffraction, precluding 3D structure determination by crystallography. Nevertheless, small-angle X-ray scattering (SAXS) measurements on suspensions of RIDC3-10a/b crystals allowed us to determine the 3D symmetry and unit cell dimensions ($P2_12_12_1$; $a = 62.9$ Å, $b = 57.2$ Å, $c = 45.7$ Å) (**Figure 3.8d**). These parameters are notable in several regards. First, the a and b dimensions are in close agreement with those deduced from ns-TEM analysis, and the c dimension matches the step height (~4 nm) observed in AFM experiments, confirming that

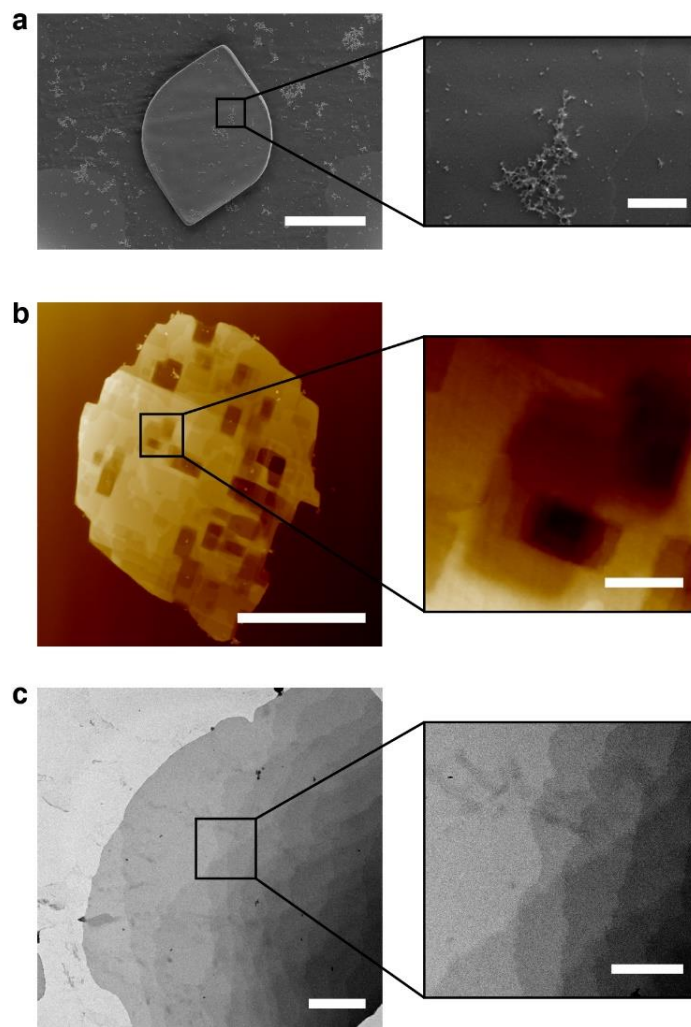


Figure 3.10 3D stacking in RIDC3-10a/b assemblies. **(a)** SEM, **(b)** AFM and **(c)** ns-TEM images of RIDC3-10a/b crystals all demonstrate stacking of individual 2D molecular layers. Insets: additional magnification of boxed regions within each image. Scale bars of images and insets are 5 μm and 500 nm, respectively.

RIDC3-10a/b crystals consist of discrete 2D molecular layers that stack up in 3D. Second, the $P2_12_12$ symmetry of the 3D unit cell is entirely consistent with the pgg symmetry obtained from 2D TEM analysis of the crystals, whereby the 2_1 screw axes along a and b dimensions (*i.e.*, the xy plane) translate into two perpendicular glide planes in a 2D projection. Importantly, the lack of a screw axis along the c dimension (*i.e.*, the z direction) indicates that the 2D RIDC3-10a/b molecular layers are in perfect alignment and are not rotated with respect to one another along the

c axis. It thus follows that 2D TEM projection maps of RIDC3-10a/b crystals (aligned flatly on the grids) should correspond to the electron densities of individual 2D molecular layers.

Using first the ns-TEM images, we obtained 15-Å resolution projection maps that displayed an arrangement of continuous interconnected chains quite distinct from those of the Zn-mediated crystals of unmodified RIDC3 (**Figure 3.9**). Next, we turned to cryogenic electron microscopy (cryo-EM) measurements, which yielded ~6-Å resolution images of RIDC3-10a/b lattices in a hydrated state (**Figure 3.8e**), with unit cell parameters ($a = 63 \text{ \AA}$, $b = 57 \text{ \AA}$, $\alpha = 90^\circ$) that closely approximate SAXS-determined values. Importantly, the cryo-EM projection maps were sufficiently detailed to allow us to build a plausible 3D structural model in combination with crystallographic constraints, the chemical knowledge of the system in hand and molecular dynamics (MD) simulations, as described below.

3.3.3 Construction of a 3D structural model for RIDC3-10a/b lattices

To assess the molecular arrangement of proteins and DNA in the 2D crystals, we developed a workflow that enabled us to compare potential molecular orientations to the experimental 2D reconstruction map and generate model candidates (**Figure 3.11**). We first focused on the basic repeat units of the lattice, which are dimers of L-shaped densities related to each other by the crystallographic two-fold rotation axis normal to the 2D plane (*i.e.*, the *z*-axis). Since the individual protein molecules do not have any internal symmetry, they cannot be placed on the symmetry axis, restricting each of the L-shaped halves to contain one or two protein units. 2D projected dimensions of the L-shaped densities (23 x 24 Å) can only accommodate a single four-helix bundle unit where the long axis of the protein is oriented approximately perpendicular to the 2D plane and therefore roughly parallel with the two-fold symmetry axis; this establishes that each dimeric unit contains two protein molecules that are parallel—and not antiparallel—to one another (**Figure 3.12**).

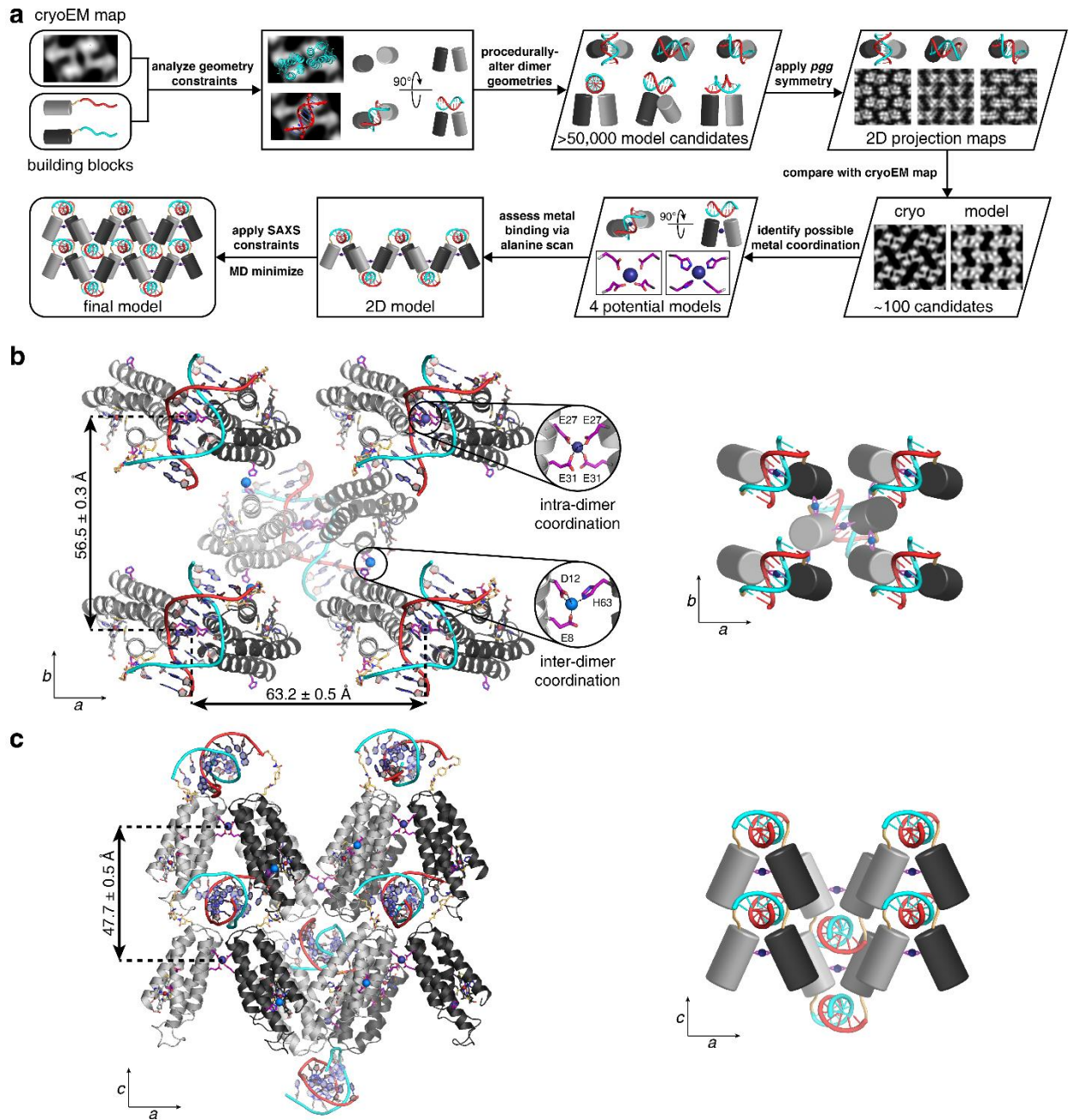


Figure 3.11 Structural modeling of the RIDC3-10a/b architecture. **(a)** Workflow detailing the generation of a 3D structural model of RIDC3-10a/b lattices. **(b)** MD-minimized model and cartoon illustration of a single 2D RIDC3-10a/b layer. **(c)** MD-minimized model and cartoon illustration showing the 3D stacking of two RIDC3-10a/b layers.

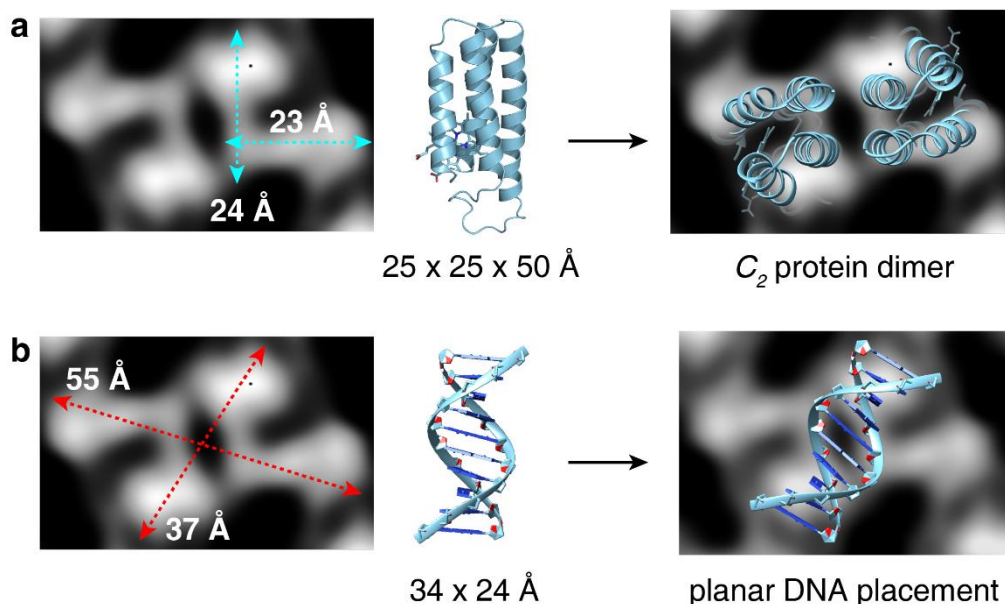


Figure 3.12 Geometric considerations for initial rigid-body fitting of protein and DNA structures into experimental cryo-EM density. Measurements of a C_2 symmetric repeating unit were used to initially define the placement of RIDC3 and 10a/10b within the structural model. **(a)** Comparison of the dimensions of an individual “L”-shaped density feature and those of an RIDC3 monomer suggest each feature contains a single protein with its longest dimension oriented perpendicular to the map. **(b)** Diagonal measurements across the dimeric unit suggest one particular orientation of DNA which can simultaneously satisfy both length and C_2 symmetry constraints.

The glide reflections inherent in pgg symmetry further dictate that the neighboring dimeric units must be oriented in opposite directions to each other with respect to the z-axis, dictating an antiparallel arrangement of protein molecules across the interdimer interface.

Given this protein arrangement, the position of the DNA-anchoring Cys21 residues on each protein monomer, and the dimensions of the dimeric unit (37 Å across one diagonal and 55 Å across the other), we can conclude that a) 10-bp dsDNA components (with approx. dimensions of 34 Å x 24 Å) must lie above and below the 2D lattice plane, and b) their “projected” two-fold rotational symmetry axis must coincide with the crystallographic two-fold rotation axis, suggesting that the dsDNA strands must be oriented parallel to the 2D plane (**Figure 3.12**). With these constraints placed on each dimeric building block (which contains protein A, protein B and an

associated dsDNA), we generated >50,000 unique model structures, in which the protein A, protein B, and the dsDNA units were rotated and translated with respect to one another in x, y and z directions (**Figure 3.13a and 3.13b**, see **Appendix 1** for python scripts). *Pgg* symmetry was further applied to generate a 2D lattice arrangement of the dimeric building blocks and to calculate a 2D projection map for each structural model (**Figure 3.13c**). These model electron densities were then examined to match features present in the experimentally measured maps. In particular, we noted that the projected map of a dsDNA unit lying parallel to the 2D crystal plane would give rise to a characteristic pattern with two parallel lines of electron density that overlays well with the long edges of the L-shapes within the dimer (**Figure 3.14**). Further, informed by previous experiments that established the necessity of Zn^{2+} coordination for protein-protein interactions,^{10,12,34} we parsed the structural models for protein orientations that placed potential metal-binding residues in close proximity (side chain distances <5 Å).

This procedure yielded four candidates whose calculated projection maps captured the salient features of the experimentally determined map and which provided a set of plausible metal- and DNA-mediated protein-protein interactions to produce a continuously connected 2D lattice (**Figure 3.15**). A set of metal coordination motifs consisting of the following residues were identified in the models (**Figure 3.15**): Glu8-Asp12-His63 (Models 1 and 2), Glu18-Glu92 (Model 1), Glu27-Glu31 (Model 2); Asp60-His63 (Model 3) and His73-His77 (Model 4). In order to identify which of these motifs were required for self-assembly, a series of point mutations were generated to change residues of interest to alanine (Ala), thereby abolishing metal binding (**Figure 3.16**). After two rounds of mutations, we established Model 2 as the only system that fulfills all chemical and crystallographic requirements, while also yielding a computed 2D projection map that most closely approximates the experimental map. This 2D arrangement fits the

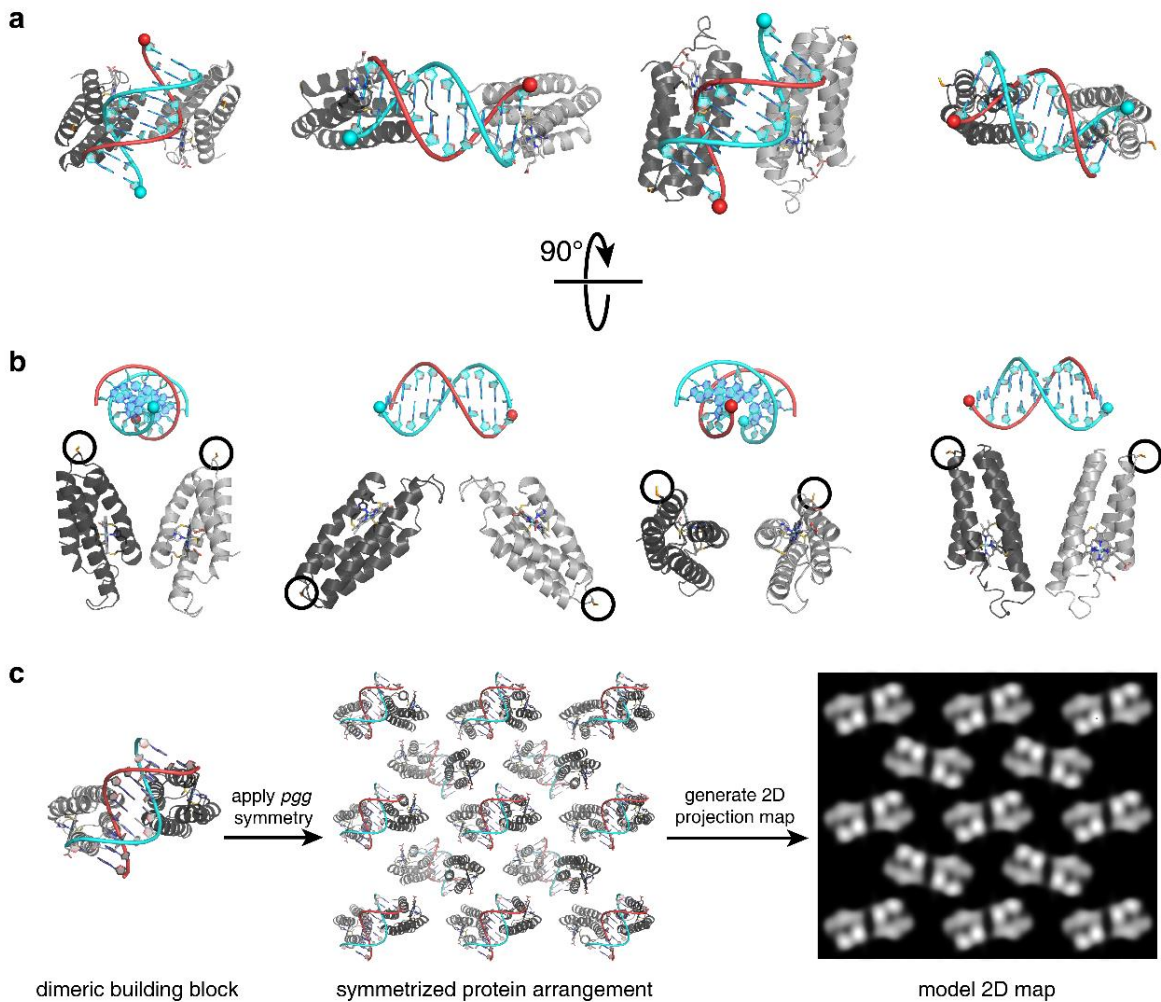


Figure 3.13 Generation of candidate structural models and their calculated projection maps. (a) Top-down and (b) side views of example RIDC3-DNA dimer orientations procedurally generated by a Python script. To highlight the attachment points for the sSMCC linker, the 5' ends of the DNA strands are shown as spheres, and the C21 residues of RIDC3 are shown as sticks and circled in (b). As these example dimers demonstrate, a wide variety of orientations were evaluated as candidate models with minimal input bias. (c) Each dimer candidate was symmetrized using the experimental *p*gg symmetry and unit cell to create a hypothetical 2D crystal, from which a calculated map could be computed and compared with the experimental density. This step was performed for all generated models.

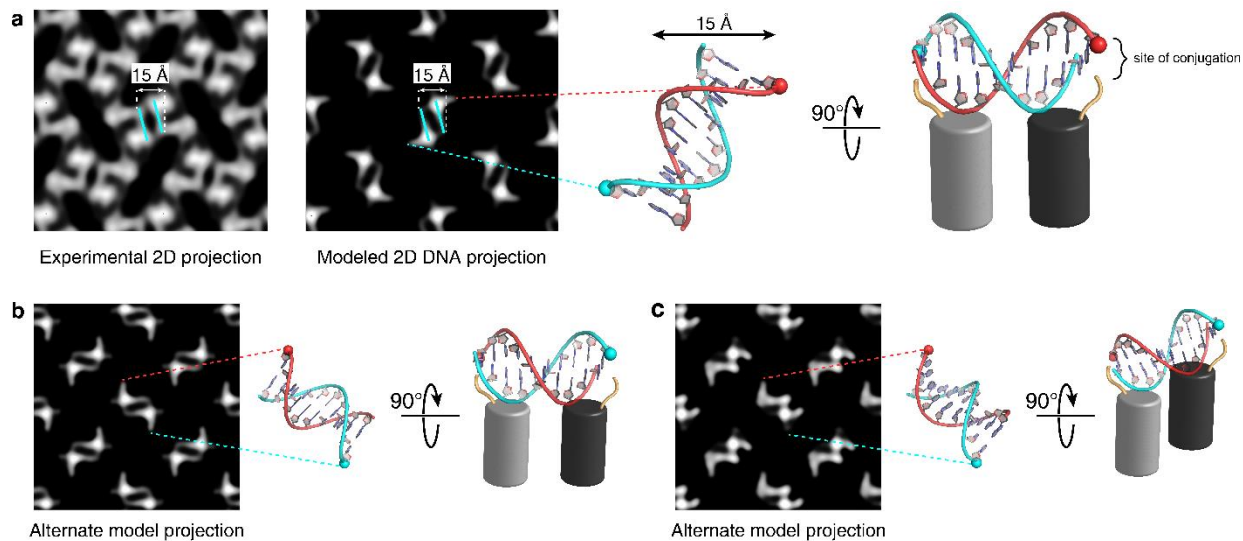


Figure 3.14 Determination of DNA orientation using calculated electron density maps. **(a)** The dimer densities within the experimental map display a set of parallel bars 15 Å apart at their center. From several random orientations of the 10a/10b DNA duplex, only one reproduces similar parallel bars in the calculated projection map. Overlays of the DNA structure on both the experimental and model map demonstrate how the turns of the helix coincide with the appearance of parallel bars. Regions of maximum intensity arise where the two ssDNA backbones cross over, which overlay well with high intensity features in the experimental map (left). Combining this structure with the predicted protein orientation, the 5'-ends (shown as spheres) are sufficiently close to the sites of conjugation (right). **(b)** A rotation of the DNA orientation in **(a)** within the plane of the crystal can be placed appropriately to conjugate to the protein, however the features in the calculated map do not match those in the experimental density. **(c)** Rotating the DNA about its long axis results in asymmetric densities and fails to recreate the parallel bars, though the 5' ends remain within crosslinking distance of the proteins.

experimentally determined cryo-EM and SAXS lattice constraints in *a* and *b*, so *z*-directional stacking (as determined by SAXS measurements) was applied to complete the model. The final 3D model was subject to MD minimization and equilibration to assess the existence of specific protein-protein and/or protein-DNA interactions.

The RIDC3-10a/b architecture consists of dimeric modules of RIDC3-10a and RIDC3-10b, linked to one another through dsDNA and a four-coordinate Zn binding motif comprising two Glu27-Glu31 pairs (**Figure 3.11b, Supplementary Movie 3.1**). Each of these dimeric, chevron-shaped modules are connected to four neighboring modules (in an antiparallel fashion) through a

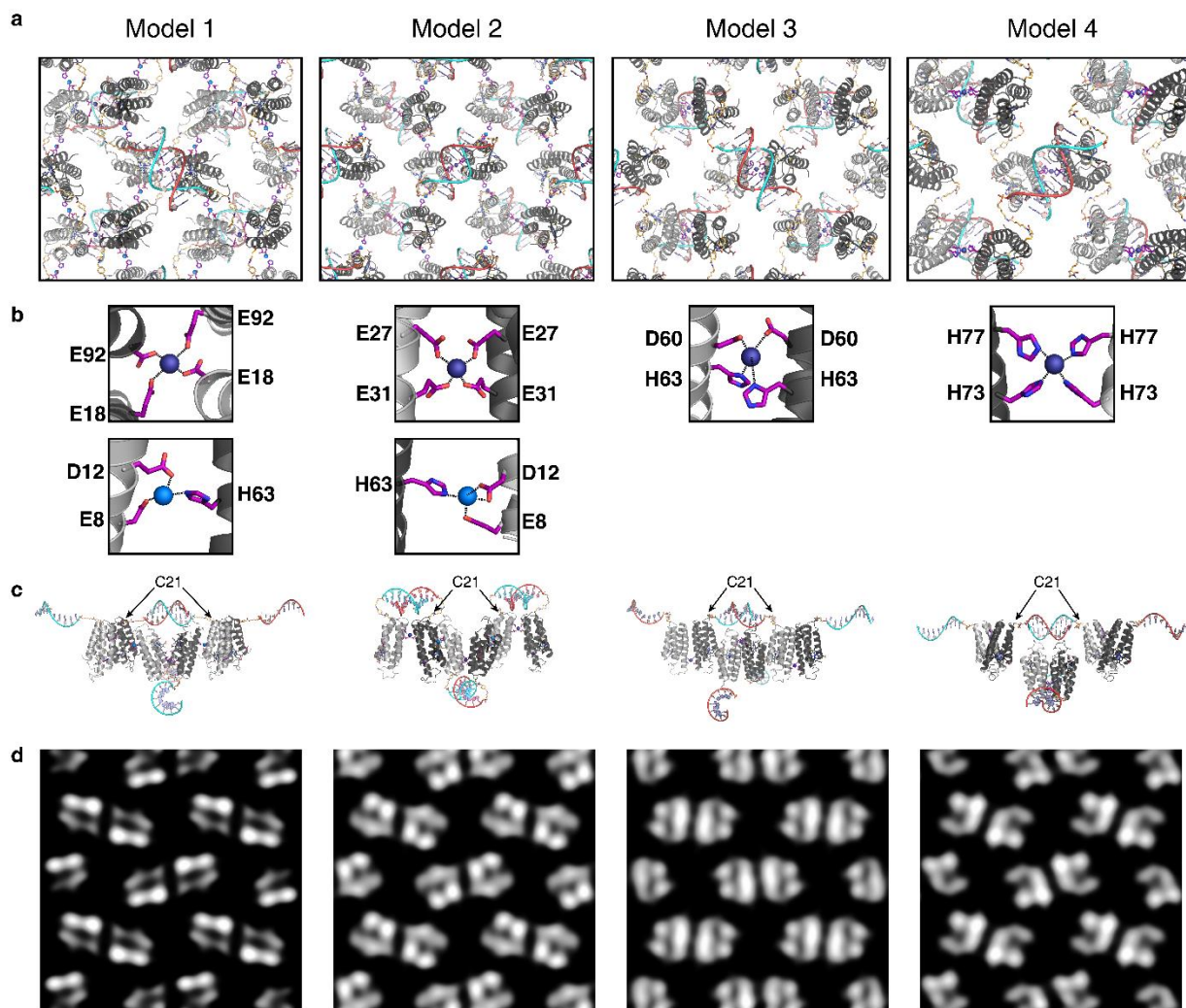


Figure 3.15 Top four candidate RIDC3-DNA models based on calculated projection maps. **(a)** Top-down views of the theoretical lattices for each model. Common to all top candidates is the placement of DNA above and below the protein dimers. **(b)** The models are differentiated by the metal-binding observed within and between dimers. All models possess intra-dimer metal-binding sites at the C_2 -symmetry axis (top), while only Models 1 and 2 also predict inter-dimer metal-binding (bottom). Intra-dimer and inter-dimer Zn^{2+} atoms are displayed as dark-blue and light-blue spheres, respectively. **(c)** Cutaway side-view of each model, showing only four dimers to demonstrate the connectivity of DNA characteristic of each one. Most notably, Models 1, 3, and 4 all contain DNA connected to adjacent dimers whereas Model 2 possesses self-contained DNA duplexes within each dimer. **(d)** Calculated 2D projection maps for each modeled crystal.

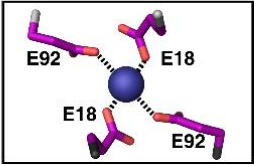
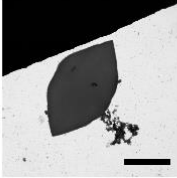
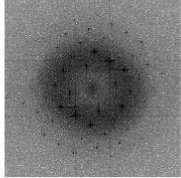
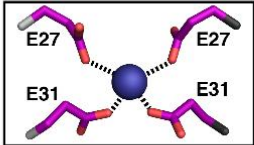
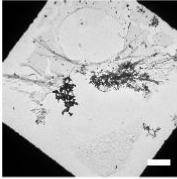
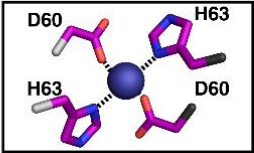
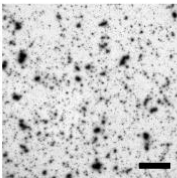
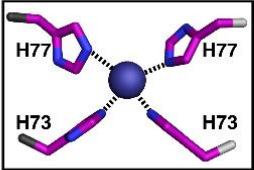
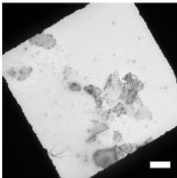
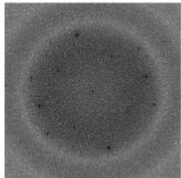
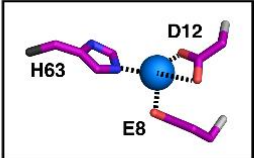
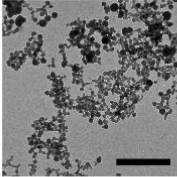
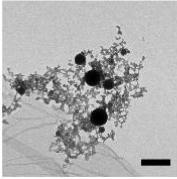
Modeled coordination environment	Site of mutation	Outcome of mutation	
Round 1			
	E18A		
	E27A		aggregates
	H63A		aggregates
	H73A		
Round 2			
	E8A		aggregates
	D12A		aggregates

Figure 3.16 TEM analysis of the effects of alanine point mutations on RIDC3-10a/b self-assembly. Each modeled coordination motif (left column) was experimentally probed by alanine scanning mutagenesis of possible metal-binding residues (central column). Intra-dimer and inter-dimer Zn^{2+} atoms are displayed as dark-blue and light-blue spheres, respectively. In Round 1, singular residues were removed from symmetric sites, thus removing 2 of 4 residues in each motif. In Round 2, all metal-binding residues in the 3-coordinate E8-D12-H63 site were individually removed. Self-assembly of all mutants into 2D crystals was evaluated by ns-TEM (right column). Computed FFTs are shown where diffraction was observed; only E18A and H73A remained competent for self-assembly. Scale bars are 5 μm .

tridentate Zn coordination motif consisting of Glu8, Asp12 and His63 side chains. This generates a 2D plane of proteins interconnected by Zn coordination with dsDNA “staples” above and below the plane, giving rise to a corrugated 2D sheet (**Figure 3.11c**). The metal content of the RIDC3-10a/b lattice was determined by using the fluorescent indicator 4-(2-pyridylazo)resorcinol (PAR), a strong Zn chelator. The measured value of ~1.25 tightly bound Zn ions per protein monomer closely approximates the ratio of 1.5 expected from the structure. The corrugated arrangement allows the stacking of 2D sheets in register along the z-direction to yield $P2_12_12$ symmetry and an interplanar spacing of ~47 Å as determined by SAXS measurements. At this spacing, the dsDNA components of each sheet wedge into the open protein-protein interfaces of the neighboring sheets and form close non-covalent contacts with the protein surfaces (**Figure 3.11c**). Thus, a fully integrated, 3D structural network is created through the synergy of DNA-, protein- and metal-mediated interactions.

This synergy is manifested in the temperature-dependent behavior of RIDC3-10a/b lattices. Although temperatures less than 10 °C were required to initiate self-assembly (in accord with the predicted $T_m \approx 7$ °C of the 10a/10b duplex), once the lattices were formed, they maintained their crystallinity until ~35 °C, at which point they rapidly disintegrated in a highly cooperative transition typical of the melting behavior of dsDNA (**Figure 3.17**). Upon recooling of the solution and incubation at <10 °C, RIDC3-10a/b lattices reformed. However, if excess (>8-molar equiv.) single-stranded 10a or 10b was added as a competitor to the self-assembly mixture during recooling, no lattices were observed, confirming the involvement of DNA hybridization in RIDC3-10a/b lattice formation (**Figure 3.18**).

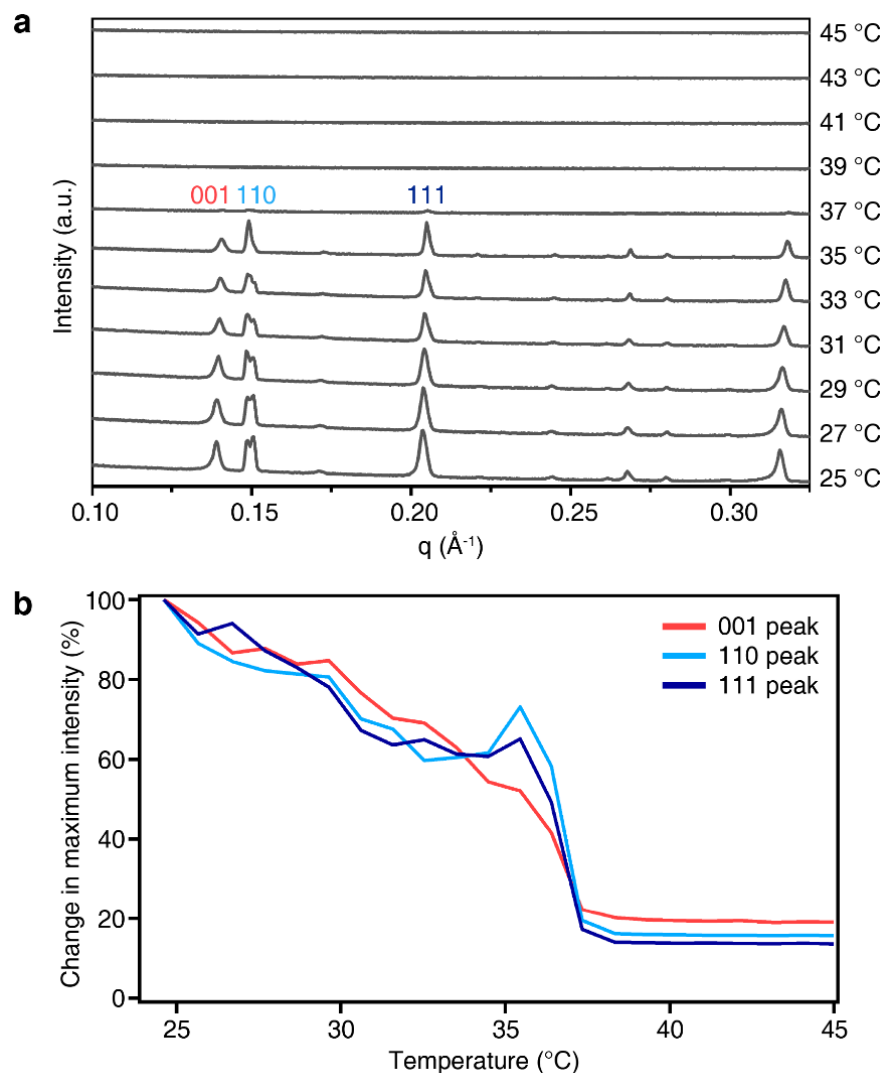


Figure 3.17 Temperature-dependent SAXS profiles of RIDC3-10a/b assemblies. **(a)** SAXS profiles of RIDC3-10a/b lattices as the temperature is raised from approximately 25 to 45 °C, plotted against the scattering vector length q . There is a sharp drop in scattering intensity between 35 and 37 °C across all peaks, suggesting decomposition of the RIDC3-10a/b crystals. **(b)** This transition is more readily visualized by monitoring the change in maximum intensity for three scattering peaks (within specific ranges of q values) as a function of temperature.

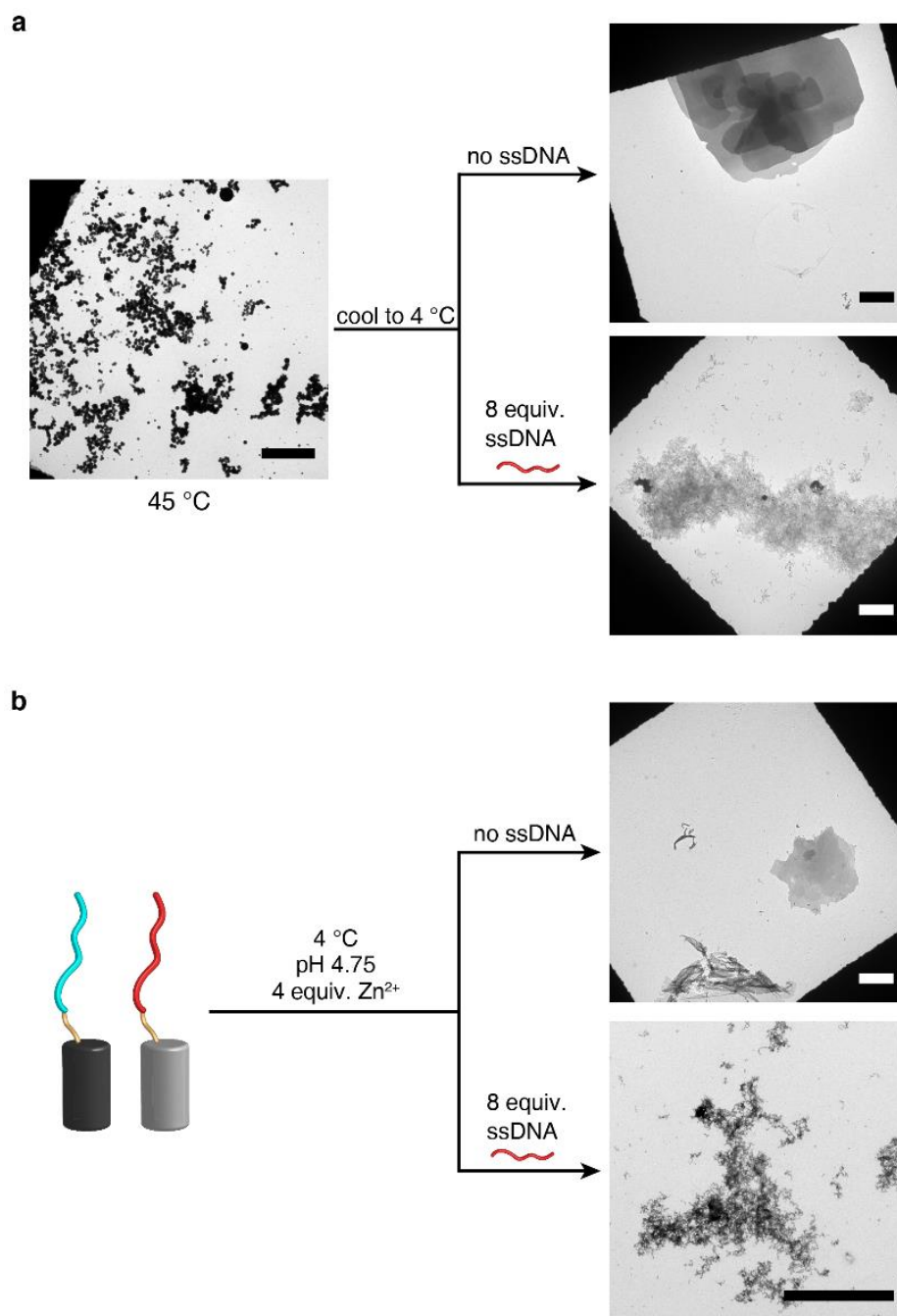


Figure 3.18 Inhibition of RIDC3-10a/b self-assembly process by the inclusion of complementary ssDNA. **(a)** Preformed RIDC3-10a/b crystals were heated to 45 °C to induce their disassembly (left), to which no or ca. 8 equivalents of ssDNA were added prior to cooling to 4 °C. The addition of ca. 8 equivalents ssDNA prevented the re-formation of crystals (right). **(b)** The inclusion of no or ca. 8 equivalents of ssDNA prior to the addition of Zn²⁺ to induce the self-assembly of new crystals yielded identical results. Scale bars are 5 μ m.

3.3.4 Computational analysis of protein-DNA interactions within the RIDC3-10a/b architecture

The preponderance of acidic residues participating in Zn^{2+} coordination is consistent with the low pH (4.75) required for the self-assembly of RIDC3-10a/b and distinct from the Zn-mediated self-assembly of unmodified RIDC3, which invariably involved Zn coordination to surface His residues (H63, H73, H77; **Figure 3.1a**) at higher pH's¹². In addition, an inspection of the protein surfaces that form interplanar contacts with the negatively charged dsDNA units indicates that they are also predominantly negatively charged at neutral pH (**Figure 3.19**). Therefore, the 3D stacking of RIDC3-10a/b layers can only be stabilized at acidic pH upon mitigation of the protein negative charge and activation of specific protein-DNA interactions, as

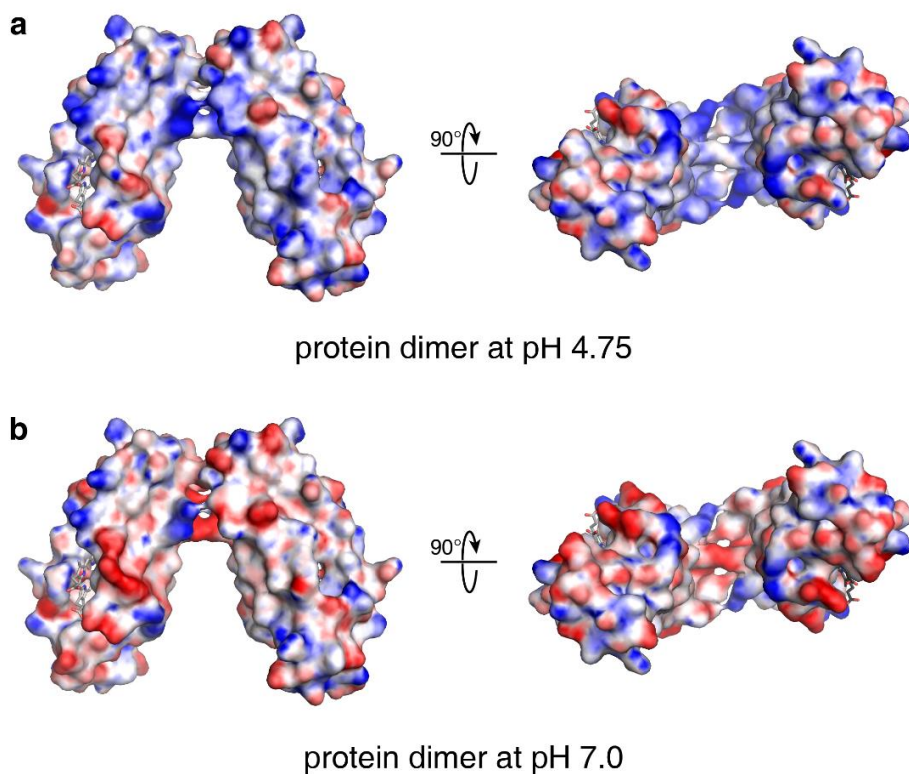


Figure 3.19 pH dependence of the calculated surface charges of RIDC3 dimers. Electrostatic maps of RIDC3 dimers at (a) pH 4.75 and (b) pH 7.0, as calculated using APBS 1.5^{57, 58}. As anticipated, the overall surface charge of the proteins is indeed more positive at pH 4.75, most notably on the underside of the dimers into which the DNA of another dimer would dock, supporting the role of low pH on RIDC3-10a/b self-assembly.

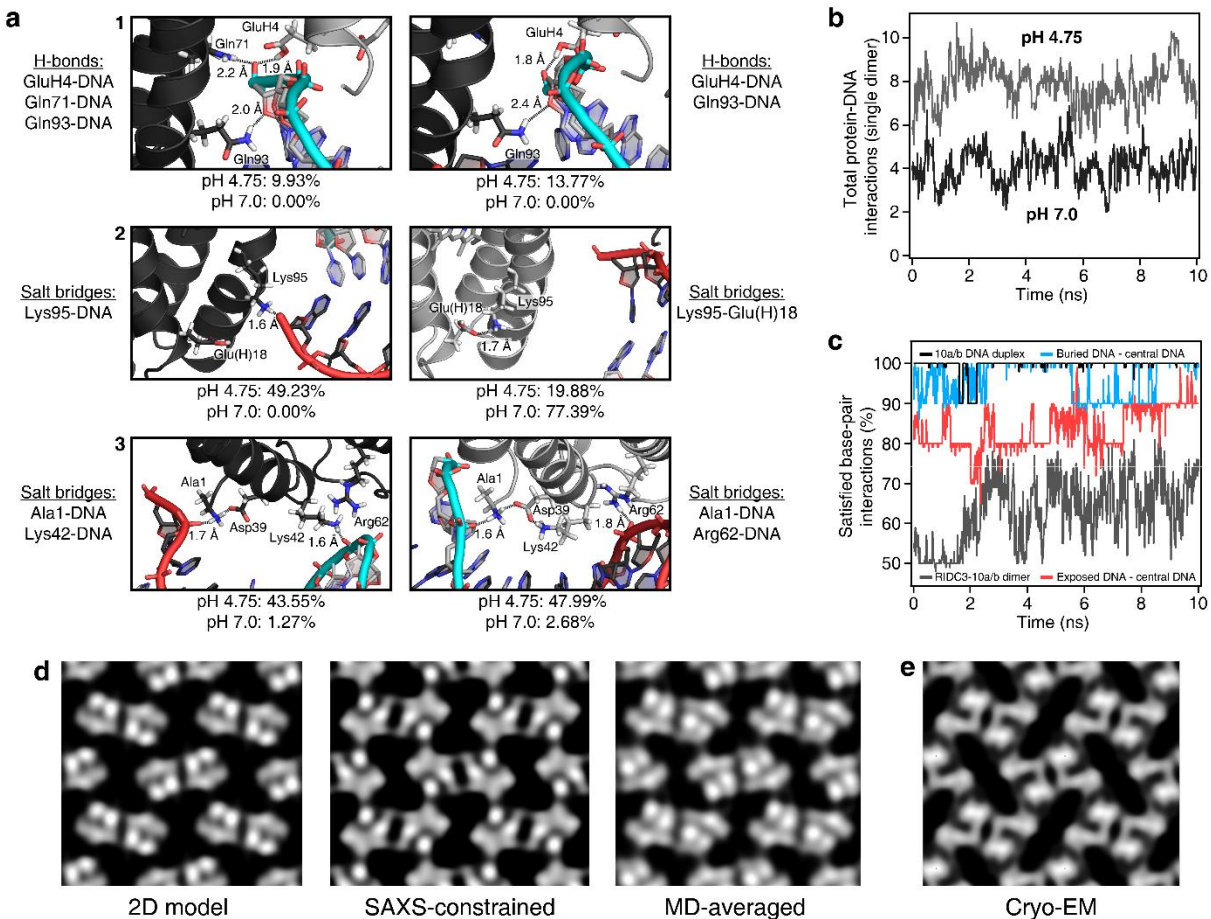


Figure 3.20 Molecular dynamics simulations of the RIDC3-10a/b lattice and comparison of calculated and experimental 2D projection maps. **(a)** pH-dependent protein-DNA interactions at pH 4.75 and 7.0 promoting 3D stacking. Percentages listed underneath each box indicate the relative occupancy (per RIDC3 monomer) of the depicted interactions during 10 ns of sampling at equilibrium for both pH 4.75 and pH 7.0. **(b)** Total protein-DNA interactions for a single dimer in the crystal bilayer plotted over 10 ns of simulation. **(c)** Fraction of satisfied base-pair interactions over time for an isolated (black trace) or RIDC3-conjugated (colored traces) 10a/10b duplex. **(d)** Calculated projection maps for RIDC3-10a/b Model 2 before (left) and after (center) adjustment to account for SAXS constraints, and from equilibrium MD simulations (right). **(e)** Experimental cryo-EM density map for the RIDC3-10a/b lattice.

corroborated by MD calculations (**Figure 3.20**). Simulations of a minimal bilayer crystal containing 10 dimeric units (*i.e.*, two stacked sheets, each containing 5 dimers) revealed an intricate network of interactions arising from the protonation of surface-exposed Glu residues at pH 4.75, increasing contact with the DNA backbone both directly through H-bonding and

indirectly by the release of basic residues from intra-protein salt bridges (**Figure 3.20a**). Interactions **1** and **2**, as shown in **Figure 3.20a**, highlight some direct consequences of sidechain protonation at low pH, resulting in a severe reduction in interaction strength at pH 7.0 by favoring intra-protein salt bridging (Lys95-Glu18) and charge repulsion (Glu4-DNA), respectively. Common to both pH values is interaction network **3**, in which the protonated N-terminus of the protein associates with the DNA backbone and Lys42/Arg62 provide redundant contacts to ensure a basal level of interaction; pH 7.0 most frequently exhibits singular Arg62-DNA interactions in place of the multivalent network observed at pH 4.75. This pattern is reflected in a nearly two-fold increase in the number of total protein-DNA contacts at low pH (**Figure 3.20b**). MD simulations further showed that dimeric RIDC3-10a/b units could only maintain their structural integrity when placed in the context of the lattice and were not stable in isolation, as evidenced by a ca. 20% increase in DNA base-pairing relative to a free dimer (**Figure 3.20c**, **Figure 3.21**). Notably, a superior approximation to the experimental cryo-EM map, relative to static models, was obtained from an averaged calculated 2D projection from the production trajectory (**Figure 3.20d**).

3.3.5 Effect of DNA length and sequence on RIDC3-DNA self-assembly

When equimolar RIDC3-10a and RIDC3-10b were incubated at 4 °C and pH 4.75, we only observed monomeric species in solution and no evidence of a stable dimer (**Figure 3.22**), showing that 10-bp DNA hybridization alone is not sufficiently strong to maintain a stable protein dimer. These findings again illustrate that the formation of the 3D structural network of RIDC3-10a/b lattices requires the cooperativity of three different types of interactions (metal-mediated protein-protein, DNA hybridization, protein-DNA). While the orthogonality of these interactions affords self-assembly with high chemical specificity (*i.e.*, in a very narrow window of conditions), their

individual weakness allows for the reversibility and structural fluidity that is needed for the observed rapid assembly kinetics.

a

Simulation component	Average fraction base-paired
10a/b duplex	99.53%
RIDC3-10a/b dimer	63.04%
5 dimer monolayer crystal	77.04%
10 dimer bilayer crystal - pH 7.0	82.66% (buried) 70.72% (exposed)
10 dimer bilayer crystal - pH 4.75	83.03% (buried) 75.22% (exposed)

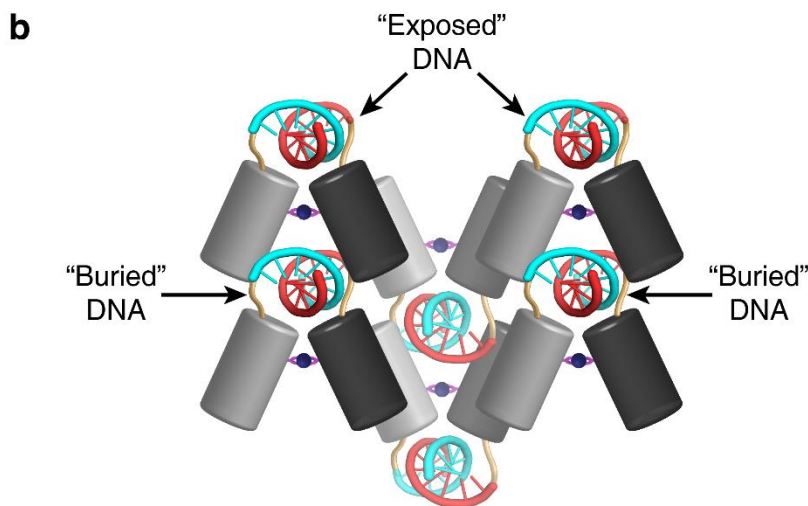


Figure 3.21 Structural integrity of DNA duplexes in various contexts as determined by MD simulations. **(a)** Tabulated averages of DNA base-pairing interactions (for an isolated 10a/10b duplex, isolated RIDC3-10a/10b conjugate, and monolayer/bilayer crystals) averaged over 10 ns (10,000 frames) of simulation. The 10a/10b duplex tethered to a free RIDC3 dimer exhibits a severe reduction in base-pairing relative to a free 10a/10b duplex. Base-pairing recovers upon the integration of the RIDC3-DNA dimer into a crystal monolayer and maximizes with the addition of stacking interactions. In particular, “buried” DNA (with protein interactions on all sides) exhibit the highest degree of base-pair interactions relative to solvent-exposed (“exposed”) DNA, indicative of the stabilization afforded by stacking of crystal layers. **(b)** Side view cartoon of the RIDC3-10a/b stacked layers visually depicting the “exposed” and “buried” DNA terminology used in **(a)**.

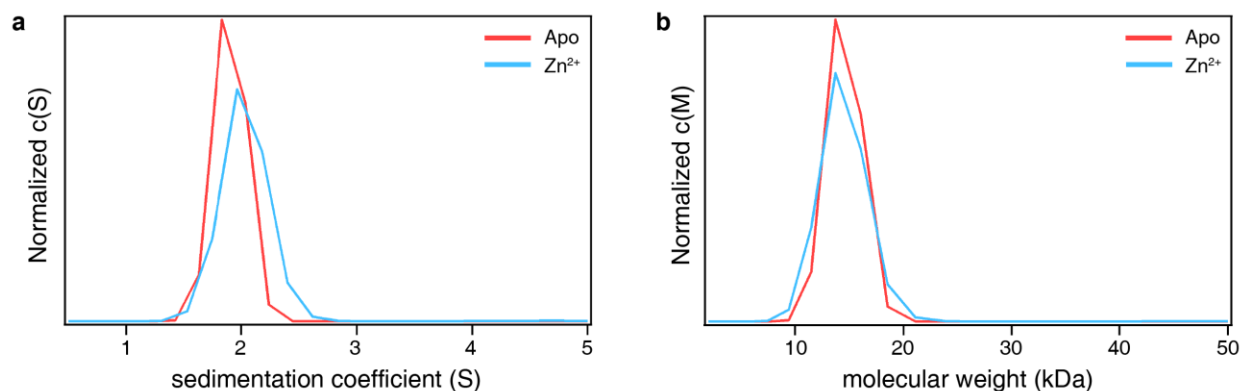


Figure 3.22 Determination of the oligomerization state of RIDC3-10a/10b in solution by AUC. **(a)** Sedimentation velocity profiles and **(b)** molecular weight distributions for equimolar (2.5 μM) RIDC3-10a and RIDC3-10b in the presence of 1 mM EDTA (red traces) or 5 μM (1 equivalent) Zn^{2+} (blue traces) at pH 4.75 and 4 $^{\circ}\text{C}$. Both sets of distributions reveal only monomeric RIDC3-DNA conjugates present in solution, indicating their inability to dimerize either by DNA hybridization alone or in conjunction with sufficient Zn^{2+} for intra-dimer coordination.

Finally, we investigated whether the hybrid lattices can accommodate variations in the DNA components to modulate self-assembly. To this end, we prepared four additional pairs of complementary RIDC3-DNA chimeras: one pair (RIDC3-10c and -10d) which differs from the 10a and 10b chimeras in the order of DNA sequences, and three pairs (8a-8b, 12a-12b and 15a-15b) which differ in both sequence and length (**Figure 3.23a**). As anticipated, RIDC3-10c and RIDC3-10d formed Zn-mediated crystalline arrays when paired with one another, but not when paired with RIDC3-10a or RIDC3-10b (**Figure 3.24**). RIDC3-12a/RIDC3-12b also formed crystalline arrays under the same solution conditions, but now at temperatures up to 20 $^{\circ}\text{C}$, consistent with the higher $T_{\text{m, pred}}$ (~ 25 $^{\circ}\text{C}$) of the 12a/12b duplex (**Figure 3.23b and 3.23c**). The experimentally determined lattice parameters of RIDC3-10c/d and RIDC3-12a/b crystals were slightly different than those of the RIDC3-10a/b crystals, but a comparison of the 2D electron density projection maps of the two lattices showed a nearly identical connectivity pattern (**Figure 3.23b**). This observation is consistent with the invariance of the Zn-mediated protein-protein interactions in the 2D plane and the ability of this lattice arrangement to accommodate a small

DNA sequence pairs	Predicted T_m (°C)	2D crystal formation	ns-TEM lattice parameters
8a: TTGGTTAA 8b: TTAACCAA	6.3	none observed	N/A
10a: TTAATTAAAA 10b: TTTTAATTAA	7.1	pH 4.75 at 4 °C	$a = 60 \pm 1 \text{ \AA}$, $b = 58 \pm 1 \text{ \AA}$ $\alpha = 90 \pm 1^\circ$
10c: TATATAAAAT 10d: ATTTTATATA	3.9	pH 4.75 at 4 °C	$a = 59 \pm 1 \text{ \AA}$, $b = 56 \pm 1 \text{ \AA}$ $\alpha = 93 \pm 1^\circ$
12a: AAATCATCTAAG 12b: CTTAGATGATTT	24.5	pH 4.75 at 25 °C	$a = 65 \pm 1 \text{ \AA}$, $b = 55 \pm 1 \text{ \AA}$ $\alpha = 90 \pm 0.5^\circ$
15a: GCATCCTCTAAGATT 15b: AATCTTAGAGGATGC	40.4	none observed	N/A

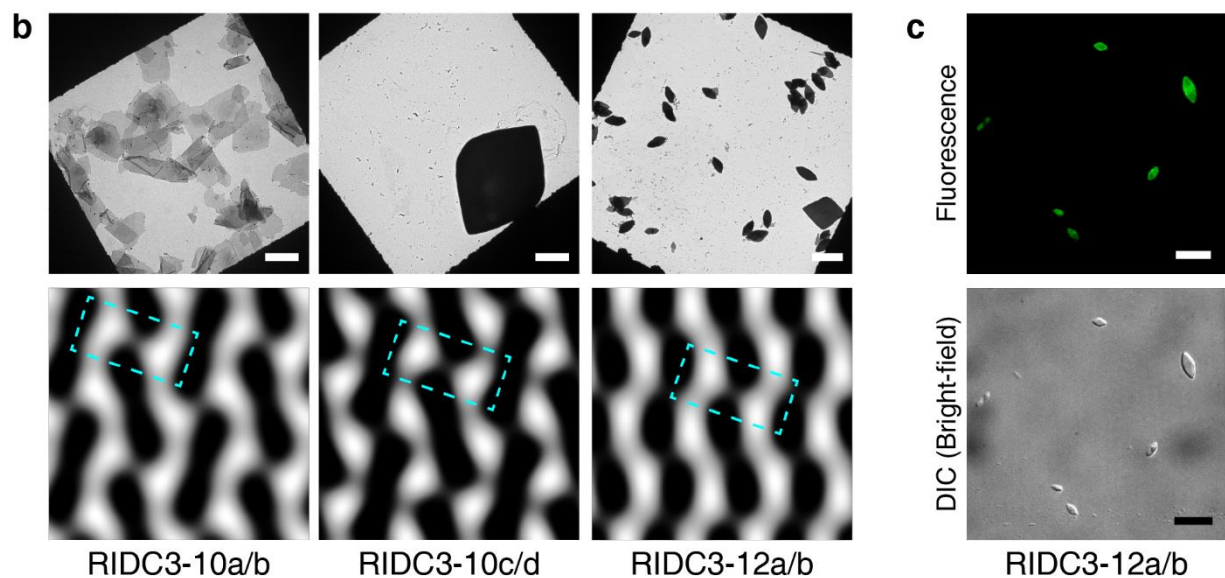


Figure 3.23 Characterization of various RIDC3-DNA constructs evaluated for self-assembly. **(a)** Tabulated parameters for all RIDC3-DNA conjugates. The predicted T_m of each duplex was calculated using OligoAnalyzer 3.1. **(b)** Representative ns-TEM micrograph and approximately 15-Å resolution 2D projection maps of RIDC3-10a/b, 10c/d, 12a/b. Similar features in projection maps are boxed in cyan. **(c)** SYBR Green-incubated RIDC3-12a/b crystals brightly fluoresce, as seen in a fluorescence channel image and DIC (bright-field) image, indicating the presence of dsDNA. Scale bars for **(b)** and **(c)** are 5 μm and 10 μm , respectively.

elongation of the DNA staples afforded by the flexibility of the linkers between the protein and the DNA components (**Figure 3.25**). In contrast, we could not identify any experimental conditions under which the RIDC3-8a/8b ($T_{m, \text{pred.}} < 6.3 \text{ }^\circ\text{C}$) and RIDC3-15a/15b ($T_{m, \text{pred.}} > 40 \text{ }^\circ\text{C}$) pairs formed ordered assemblies. This is likely not only due to the steric incompatibility of very short (8 base pairs) or very long (15 base pairs) DNA staples with the lattice, but also due to significant alterations in the delicate energetic balance between DNA hybridization and Zn-mediated protein interactions necessary for crystalline self-assembly.

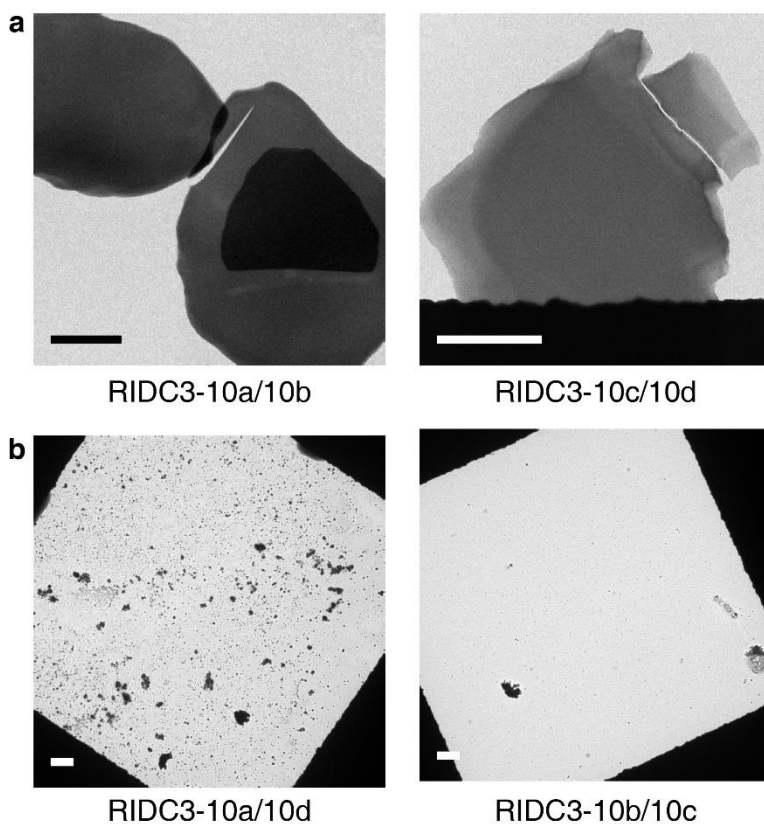


Figure 3.24 Role of DNA sequence specificity in RIDC3-DNA self-assembly. All samples contain equimolar (25 μM) each RIDC3-DNA conjugate, 4 equivalents Zn^{2+} , and were incubated at 4 $^\circ\text{C}$. Ordered crystals were observed with for the complementary sequences (**a**) RIDC3-10a/10b and RIDC3-10c/10d, but the non-complementary sequences (**b**) RIDC3-10a/10d or RIDC3-10b/10c only produced aggregates. Scale bars are 5 μm .

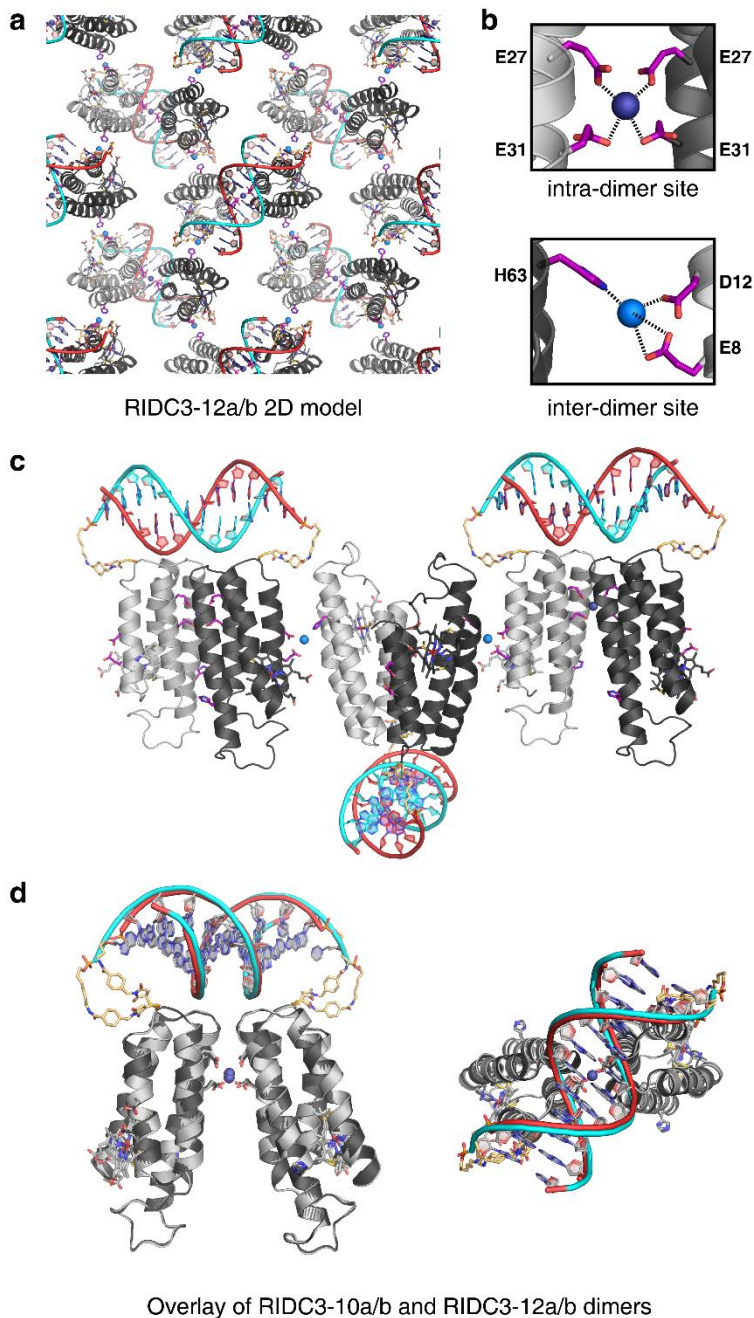


Figure 3.25 Structural model of RIDC3-12a/b lattices. **(a)** A top-down view of the modelled RIDC3-12a/b crystal reveals an identical bonding arrangement to that of RIDC3-10a/b. **(b)** Insets of the same metal-binding motifs identified in the RIDC3-10a/b model. **(c)** Side view of the assembly, demonstrating that the SMCC linker can still properly connect C21 of RIDC3 and the 5' end of each DNA strand. **(d)** Overlaying and aligning the RIDC3-10a/b and RIDC3-12a/b models reveals essentially no differences between the two systems. RIDC3-10a/b is shown in light gray with red DNA; RIDC3-12a/b is shown in dark gray with cyan DNA.

3.4. Conclusions

The functioning of a cell hinges upon the self-assembly of and communication between complex biopolymers in a crowded environment. The extremely high fidelity of these processes, in turn, necessitates a sophisticated chemical language that emerges from various combinations of biologically available non-covalent interactions. Along these lines, we have presented here a synthetic nucleoprotein assembly that combines three prominent classes of intermolecular interactions (Watson-Crick base pairing, DNA-protein interactions, protein-metal coordination) to self-assemble with high structural order and specificity in a manner that is reminiscent of natural nucleoproteins like the ribosome. While the modular nature of such multi-component systems should offer distinct advantages in the construction of structurally tunable materials, the intricate architecture of the RIDC3-DNA assembly also highlights the opportunities and challenges inherent in designing artificial nucleoprotein complexes that arise from the distinct structural and chemical properties of proteins and NAs.

3.5 Materials and Methods

3.5.1 General Considerations

Unless otherwise stated, reagents and solvents were purchased from Fisher Scientific and used without further purification. DNA oligos were ordered from Integrated DNA Technologies or Eurofins Genomics with a 5' six carbon amino linker and used without further purification.

3.5.2 Site-directed mutagenesis, protein expression, and purification

All protein constructs were housed in a pET20b vector (Novagen). RIDC3 mutants were introduced into the pET20b expression vector containing the genes encoding ^{C21}RIDC3 using QuikChange (Stratagene) site-directed mutagenesis using primers synthesized by Integrated DNA

technologies. Protein expression and purification was performed as described in published procedures^{35, 36}.

3.5.3 Preparation of RIDC3-DNA conjugates

All DNA sequences for protein conjugation were obtained from Integrated DNA Technologies or Eurofins Genomics with a 6-carbon 5'-amine linker and used without further purification. Modification of the 5' amine was performed as follows: Each ssDNA strand was dissolved in water to a final concentration of 1 mM. 7.5 mg sulfosuccinimidyl-4-[N-maleimidomethyl]cyclohexane-1-carboxylate (sulfo-SMCC) was dissolved in 100 μ l dimethylformamide (DMF) to yield a 172 mM stock solution. 400 μ l of the DNA stock solution and 500 μ l of the conjugation buffer (CB; 16.7 mM potassium phosphate monobasic, 83.3 mM potassium phosphate dibasic, 150 mM sodium chloride, pH 7.3) was combined with 100 μ l of the sulfo-SMCC stock solution and mixed for 10 s via vortexing³². The solution was incubated at 35 °C in the absence of light for 1.5 – 2 h. The SMCC-functionalized DNA was purified by reverse-phase HPLC on a Zorbax SB-C18 PrepHT column (Agilent) and eluted in 30% buffer B (Buffer A: 5 mM sodium acetate, pH 5.0; Buffer B: acetonitrile) using a linear gradient. The desired fraction was collected, frozen, and lyophilized to dryness overnight.

Approximately 20 equivalents of dithiothreitol (DTT) were added to 1.5 ml of 0.7 mM ^{C21}RIDC3 (or ^{C21}RIDC3 mutant) and incubated at room temperature for 30 min to fully reduce the reactive thiol group on the protein. An Econo-Pac 10DG desalting column (BioRad) was used to remove excess DTT and the protein was eluted in 2 ml CB. Once dry, lyophilized DNA was dissolved in 1 ml CB and added to the protein, which was present in large excess. The conjugation reaction proceeded for 12 h at 4 °C in the absence of light.

The reaction mixture was loaded onto an Econo-Pac 10 DG desalting column (BioRad) and eluted in 4 ml of a buffered solution of 10 mM sodium phosphate (pH 8.0). The RIDC3-DNA conjugate was purified using a DuoFlow fast protein liquid chromatography (FPLC) workstation (BioRad) equipped with a Bio-Scale Mini Macro-Prep High Q Cartridge (BioRad) anion exchange column equilibrated in a 10 mM sodium phosphate buffer solution (pH 8.0). Unreacted protein and the protein-DNA conjugate were eluted using a gradient of 0 – 1 M NaCl. Unlabeled protein eluted at approximately 200 mM NaCl, RIDC3-DNA conjugate eluted at approximately 500 mM NaCl, and free DNA eluted at approximately 650 mM NaCl. Fractions containing pure (described below) RIDC3-DNA conjugate were combined, concentrated, and stored in deionized water at – 80 °C.

3.5.4 Characterization of RIDC3-DNA conjugates

Ultraviolet-visible (UV-Vis) absorbance measurements were used to assess the purity of FPLC-purified RIDC3-DNA conjugates. The ratio of the Soret band of the protein at 415 nm ($\epsilon_{415} = 148,000 \text{ M}^{-1}\text{cm}^{-1}$) to the DNA absorbance at 260 nm was determined to be approximately 1.15 for pure DNA-protein conjugate. Purities of the conjugates and incorporation of a single ssDNA strand per protein monomer were confirmed by non-reducing SDS-PAGE.

Electrospray ionization mass spectrometry (ESI-MS) was performed to verify the mass of the RIDC3-DNA conjugate. Mass spectrometry was performed at the Molecular Mass Spectrometry Facility at UC San Diego on a Micromass Quattro Ultima Triple Quadrupole MS. RIDC3-DNA conjugate samples were exchanged into water and diluted to a concentration of 0.1-1.0 mg/ml using a solution of 50% methanol in water.

3.5.5 Self-assembly of 2D RIDC3-10a/b lattices

RIDC3-10a and RIDC3-10b, each at a concentration of 25 μM , were combined in buffered solutions of 20 mM 2-(N-morpholine)ethanesulfonic acid (MES) (pH 4.75) to a total protein concentration of 50 μM . A 5 mM Zn^{2+} stock solution was prepared in water, and the required amount of Zn^{2+} (4-10 equivalents) was added to the protein solution. All samples were prepared at room temperature and moved into a cold room at 4 °C with the exception of RIDC3-12a/b, which formed ordered arrays at room temperature. All samples that formed crystalline arrays were opaque within 10 min and formed insoluble precipitates that settled to the bottom of the tube, which could be easily re-suspended by shaking or pipetting. All other RIDC3-DNA conjugate variants were prepared under identical conditions to assess array formation.

3.5.6 Negative-stain transmission electron microscopy (ns-TEM) analysis

A 3- μl solution of an RIDC3-DNA conjugate sample was pipetted onto carbon-coated Cu grids (Electron Microscopy Sciences, 400 square mesh) or formvar/carbon-coated Cu grids (Ted Pella, Inc.) glow discharged within 30 min of sample addition. After sample incubation on the grid for 5 min, grids were blotted with filter paper to remove excess liquid. Samples were washed by submerging in a 250 μl drop of MilliQ water, blotted with filter paper to remove excess liquid, and stained by the addition of 5 μl of 1% uranyl acetate (UA) solution in water to the grid. After a 5-min incubation, the grid was blotted to remove excess UA solution, and a second aliquot of UA was added. After the second 5-min incubation, the grid was blotted dry using filter paper. Grids were imaged using a FEI Sphera transmission electron microscope operating at 200 keV, equipped with a LaB_6 filament and a Gatan 4K charged-coupled device (CCD). Micrographs were collected using objective lens underfocus settings ranging from 250 nm to 2 μm and analyzed using Fiji (<http://fiji.sc/Fiji>).

3.5.7 Job's analysis of RIDC3-10a/b self-assembly

The stoichiometry of RIDC3-DNA complexes in fully formed crystals was assessed by Job's method. While maintaining the total protein concentration at 50 μM for all samples, self-assembly was carried out using the following ratios of RIDC3-10a:RIDC3-10b: 1:0; 9:1; 4:1; 1:1; 1:2; 1:4; 1:9; 0:1. All samples were prepared as previously described to facilitate the formation of 2D crystals. After a 12 h incubation period at 4 $^{\circ}\text{C}$, grids were prepared for each sample and imaged by TEM as previously described. Images were collected at a nominal magnification of 1700x with a pixel size of 60.8 \AA . RIDC3-10a/b suspensions were centrifuged to separate the pelleted arrays from free protein in the supernatant. The absorbance of the supernatant at 415 nm was measured for each sample to determine the amount of protein remaining in solution and therefore not incorporated into RIDC3-10a/b arrays.

3.5.8 Confocal microscopy

A 0.5 μl aliquot of a 20x SYBR Safe stock diluted in a buffered solution of 20 mM MES (pH 4.75) was added to a 5- μl sample of RIDC3-10a/b crystals, nominally at 50 μM total protein concentration. After 10 min, the solution was pipetted onto a glass slide and was immediately covered with a cover slip, and the edges were sealed with clear nail polish. Samples were imaged with a 100x oil objective on a spinning-disk confocal Zeiss Axio Observer inverted microscope equipped with a pair of Roper Quantum 5125C cameras using a filter to collect light at 500-550 nm (green channel). Differential interference contrast and fluorescence (488 nm excitation) images were captured at 1-s and 100-ms exposures, respectively. Images were collected in Slidebook 6 (Intelligent Imaging Innovations) and analyzed using Fiji (<http://fiji.sc/Fiji>).

3.5.9 Scanning electron microscopy (SEM)

RIDC3-10a/b crystals were deposited onto glow discharged, formvar/carbon-coated Cu grids (Ted Pella, Inc.) as described previously. Grids were mounted onto a STEM 12x v2 sample holder and imaged using a Zeiss Sigma 500 scanning electron microscope (Zeiss) at an accelerating voltage of 1 kV using a 30 μm aperture. Images were analyzed using Fiji (<http://fiji.sc/Fiji>).

3.5.10 Atomic force microscopy (AFM)

A 10 μl solution containing RIDC3-10a/b crystals was centrifuged at 3500 RPM for 1 min. The supernatant was removed, and the crystals were resuspended in 10 μl MilliQ water. The sample was then deposited onto freshly cleaved mica (Ted Pella, Inc.), and incubated for 10 min. At that time, the mica was dried using a stream of nitrogen without washing.

AFM measurements were performed on two microscopes. For images shown in **Figure 3.8c**, mica discs were imaged on a Veeco Scanning Probe Microscope using Silicon AFM probes with aluminum reflex coating at a resonance frequency of 300 kHz (Ted Pella, Tap300Al-G). Images were obtained at a field size consisting of 512 x 512 pixels. Image analysis was done using WSxM 5.0³⁷. For images shown in **Figure 3.10**, mica discs were imaged on a Bruker Dimension Icon ScanAsyst atomic force microscope using a ScanAsyst-Air tip (Bruker) operating in tapping mode. Images were processed and analyzed using NanoScope Analysis (v.1.5, Bruker).

3.5.11 Monitoring RIDC3-10a/b self-assembly with Small Angle X-ray Scattering (SAXS)

A 60- μl buffered solution of 20 mM MES (pH 4.75) containing 25 μM RIDC3-10a, 25 μM RIDC3-10b and 100 μM ZnCl_2 was prepared and transferred to a 1.5 mM quartz capillary tube (Hampton). This capillary tube was placed in a custom-built thermal stage and rapidly cooled to 4

°C. Data were collected at beamline 4-2 of the Stanford Synchrotron Radiation Lightsource (SSRL) using collimated X-ray radiation (1.1271 Å, 11 keV) calibrated with a silver behenate standard. A 1-s X-ray exposure was taken every 5 min (62 total exposures) and the scattered radiation was captured on a Pilatus 1M detector (Detectris). One-dimensional scattering data were obtained through the azimuthal averaging of the two-dimensional data to produce plots of scattering intensity as a function of the scattering vector length, $q = 4\pi\sin(\theta/\lambda)$, where θ is one-half of the scattering angle and λ is the wavelength of the X-rays used. Analysis of the one-dimensional data was performed using the powder diffraction processing software JADE (MDI).

3.5.12 Monitoring thermal disassembly of RIDC3-10a/b lattices with SAXS

The capillary tube containing the RIDC3-10a/b crystals from the previous SAXS experiment was removed and stored at ca. 22 °C for 3 h. This tube was loaded into a custom-built thermal stage and heated at ca. 1 °C/min. A 1 s X-ray exposure was taken every minute for 25 minutes and the X-ray radiation data was collected and processed as described above.

3.5.13 Cryo-EM data acquisition and processing

A 3- μ l solution of an RIDC3-DNA sample was deposited onto either a freshly glow-discharged, lacey carbon grid or a Quantifoil grid (Electron Microscopy Sciences, Quantifoil R1.2/1.3 holey carbon on 200 mesh copper). Plunge freezing was performed with a Leica EM GP instrument (Leica Microsystems), the grid was blotted from behind for 4 s with a blot force of 2 using filter paper, and rapidly plunged into a liquid ethane bath cooled with liquid nitrogen. Frozen specimens were stored under liquid nitrogen. The grids were imaged in a Titan Krios (FEI) operated at 300 kV equipped with a Gatan Quantum-LS energy filter, applying 20 eV zero loss filtration. Image data were recorded with a K2 Summit Direct Electron Detector (DED, Gatan) in counting mode, at a magnification resulting in a pixel size at the sample level of 1.098 Å. Dose-

fractionated images were recorded to yield 40 frame movies recorded over 16 s at 0.4 s per frame with a total electron dose of $\sim 40 \text{ e}/\text{\AA}^2$ per movie. Objective-lens underfocus settings varied between 0.5 μm and 3 μm . Data collection was performed using SerialEM³⁸, corrected for drift with MotionCorr2.1³⁹ and processed in real time in the Focus⁴⁰ software suite using the 2dx^{41, 42} software package.

3.5.14 Structural modeling

In order to assess the molecular arrangement of proteins and DNA within the RIDC3-10a/b lattices, a workflow was devised to compare plausible molecular arrangements to experimental data. Analysis of the asymmetric unit of the experimentally derived cryo-EM projection map (C_2 symmetric “L”-shaped densities) resulted in a model building block consisting of two protomers and one dsDNA. Protein coordinates were derived from the crystal structure of RIDC3 (PDB ID: 3TOM) and duplex DNA was generated using Nucleic Acid Builder⁴³ via the make-na server. With these constraints placed on each dimeric building block (which contains protein A, protein B and an associated dsDNA), an input model containing protein and DNA was manipulated within Pymol⁴⁴, in which the protein A, protein B, and the dsDNA units were systematically rotated and translated with respect to one another in x, y and z directions to vary the orientation of the proteins (**Figure 3.13**). Each unique arrangement was saved as a separate coordinate file (see **Appendix 1** for more details). Proteins were rotated/translated such that C_2 symmetry was preserved in each model. The DNA was untethered to the proteins at this stage to facilitate manipulation of the components. This process was repeated iteratively using an *ad hoc* Python script to procedurally generate >50,000 unique coordinate files.

Once each set of models had been generated, dimeric building blocks were arranged in accordance with $p6g$ symmetry and experimental cryo-EM lattice parameters (63.2 \AA x 57.3 \AA) to

generate extended 2D lattices using a Python script within Pymol. The EMAN2 processing suite was used to create a projection map of the *pgg* model by first converting the coordinate files to MRC files using *e2pdb2mrc* and creating a computed 2D projection with *e2proc2d* (**Figure 3.13**)⁴⁵. Maps and models were visualized and analyzed using Chimera⁴⁶ and Pymol respectively. The generated model maps were visually inspected for features characteristic of the experimentally measured maps.

Promising models were inspected manually to more precisely position protein and DNA monomers to ensure the absence of clashes between components. Models were further analyzed to determine whether the 5' amine of the DNA strands was sufficiently close ($< 17 \text{ \AA}$) to C21 of each protein to permit the crosslinking with sulfo-SMCC. Additionally, we parsed the structural models for protein orientations that placed potential Zn-binding residues in close proximity (side chain distances $< 5 \text{ \AA}$). Previous crystallographic characterization of Zn-bound *cb*₅₆₂ variants allowed us to prioritize specific residues that were likely to bind Zn; however, we examined all possible metal-binding residues (Asp, Glu, His) displayed on the protein surface that could coordinate both between and within C_2 -symmetric dimers. A pseudoatom, representing Zn^{2+} , was placed at the center of mass any potential metal-binding sites to provide a rough estimate of Zn-protein distances. In order to generate final models, the SMCC linker and Zn atoms were added to model PDBs. A SMILES string of SMCC was used to generate a coordinate file, which was conjugated to both protein and DNA within Pymol. Pseudoatoms were created at metal-binding sites to populate Zn into the models.

3.5.15 Molecular modeling and simulations

The final 3D model of an RIDC3-10a/b dimeric unit (as described above) was used for the construction of all protein-DNA systems, which comprised two ^{C21}RIDC3 proteins, the idealized

10a/10b DNA duplex, and the SMCC linkers. PSFGEN (within VMD⁴⁷) was used to add missing hydrogens, assign atom types, and patch RIDC3/SMCC and RIDC3/c-type heme covalent linkages. The protonation state of all RIDC3 sidechains was calculated using PROPKA 3.0^{48, 49}. At pH 4.75, the sidechains E4, E18, E49, H73, and H77 were non-metal-binding sidechains consistently evaluated as protonated and were patched as such using PSFGEN for all pH 4.75 simulations. CHARMM27 parameters^{50, 51} were used for all simulations to ensure compatibility with published c-type heme parameters⁵², and parameters for the SMCC linker were assigned by analogy with molecular fragments present in the CHARMM27 forcefield. Due to the localization of negative charge induced by the DNA backbone, the *cionize* plugin in VMD (which takes into account the local electric field of the structure) was used for neutralization of the system. Together, these steps comprise the construction of a single RIDC3-10a/b dimeric building block. This constructed dimer could then be translated and rotated via the crystallographic parameters to obtain crystals of arbitrary size. Systems containing 1, 5, and 10 dimeric units were all investigated computationally. TIP3P water was used for all simulations, added using the SOLVATE plugin of VMD, and ensuring at least 12 Å of water on all sides of the solute in all cases. Periodic boundary conditions were employed for all simulations.

Energy minimization and MD simulations were carried out using NAMD 2.12 with multi-core and CUDA support⁵³ on a home-built workstation possessing two NVIDIA GTX 970 GPUs. All simulations were performed in the isothermal-isobaric (NPT) ensemble, employing a Langevin thermostat with 1 ps⁻¹ damping coefficient and Nosé-Hoover barostat (period of 200 fs and decay of 100 fs) to maintain the temperature at 277 K and 1 atm, respectively. Long-range electrostatics were evaluated with the Particle mesh Ewald method, using a 12.0 Å cutoff for non-bonded interactions with a switching function activated at 10.0 Å. Hydrogen bond lengths were kept

constant using the SHAKE algorithm, and the trajectory was propagated using the velocity Verlet integration method with a timestep of 2 fs.

In all simulations, metal coordination geometry was enforced by restricting the distances between E27 and E31 carboxylate sidechains and the intra-dimer zinc atom via the use of harmonic potentials centered at 2.0 Å, maintained using a 100 kcal/mol/Å² force constant. Likewise, systems with lateral interactions with neighboring dimers had analogous restraints to maintain coordination between E8, E12, and H63 sidechains and inter-dimer zinc atoms. In all cases, metal-coordination restraints were initiated with a 10 kcal/mol/Å² force constant which was increased linearly to 100 kcal/mol/Å² over the first 1 ns of equilibration. To incorporate the effects of stacking of individual layers, two layers of 5 RIDC3-10a/b dimers were initially separated by a distance of 48.0 Å in the z-direction, and were slowly pulled together to a distance of 45.0 Å over 1 ns via 5 pairwise 100 kcal/mol/Å² harmonic distance restraints (between each dimeric building block and its corresponding partner directly above it). At the end of the pulling, these restraints were removed, leaving only metal-coordination restraints for the remainder of each run. We note that this step led to major distortions in the layer geometries when sidechains were deprotonated (for pH 7.0 simulations), so the 10-dimer simulations at pH 7.0 were created from the last frame of this equilibration step at pH 4.75 by restoring the protonation state and sidechain charges of residues 4, 18, 49, 73, and 77 and balancing the system charge with additional sodium counterions using the Autoionize plugin of VMD. While this was done in an attempt to mitigate the possibility of artifactually low protein-DNA interactions for the pH 7.0 state, we interpret this as indicative of the role that low pH plays in stabilizing the full RIDC3-10a/b lattice. All simulations were equilibrated after this step for 10 ns prior to an additional 10 ns of production sampling for structural analysis.

3.5.16 DNA structure analysis

Analysis of the double helical structure of dsDNA was performed using the `nastruct` analysis method implemented within the `cpptraj`⁵⁴ program of AmberTools 16⁵⁵. The CHARMM PSF file for a single DNA duplex was converted to `prmtop` format using the `ParmEd` utility for use with `cpptraj`, and analysis was performed on individual helices extracted from all systems. The idealized 10a/b duplex was used as a reference structure to define all expected base pairs, and `nastruct` was run with the `calcnob` flag to ensure calculation of the parameters even for non-base-paired nucleotides.

3.5.17 Quantification of Zn²⁺ in assembled arrays

A standard curve to quantify Zn²⁺ concentration was generated using 4-(2-pyridylazo)resorcinol (PAR) absorbance on an Agilent 8453 UV-Vis spectrometer and fit to the equation $y = mx + b$. Solutions containing RIDC3-10a/b crystals were centrifuged for 1 min at 13,300 rpm, washed with a buffered solution of 20 mM MES (pH 4.75), and subsequently dissolved in 300 μ l of a buffered solution of 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS) with 150 mM NaCl (pH 7.0). 100- μ l aliquots of protein solution were removed and the protein concentration was calculated using the measured absorbance at the Soret maximum (415 nm). 100 μ l of a buffered solution of 20 mM MOPS (pH 7) with 150 mM NaCl and 5 M guanidium HCl was then added to the protein solution. After a 5-min incubation at room temperature, PAR was added at an identical set of concentrations to that of the standard curve to quantify the amount of Zn²⁺ in solution. Finally, EDTA was added to chelate any free Zn²⁺ ions and obtain the PAR background absorbance.

3.5.18 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

Solutions of 2.5 μM RIDC3-10a and 2.5 μM RIDC3-10b in a buffered solution of 20 mM MES (pH 4.75) was treated with 1 equivalent Zn^{2+} or 1 mM EDTA. Sedimentation velocity measurements were made on a XL-1 Analytical Ultracentrifuge (Beckman-Coulter) as described previously². Scans were processed in Sedfit⁵⁶ using buffer viscosity (0.01002 poise), density (1.007 g/mL) and partial specific volume (0.6765 ml/g) parameters. The final c(S) and c(M) distributions are reported at a confidence level of 0.95.

3.6 Acknowledgments

We would like to thank the following colleagues for assistance: R. Anderson for assistance with SEM data collection; N. Gianneschi for use of AFM instrumentation; W-J Rappel for use of confocal microscope; K. Goldie, L. Kovacic and R. Righetto for assistance with cryo-EM data collection and processing; T. Weiss and I. Rajkovic for assistance with SAXS measurements; D. Shi, J. De La Cruz and T. Gonen for assistance with MicroED measurements; S. Joseph and A. Komor for helpful discussions. This work was primarily supported by the National Science Foundation (Division of Materials Research; DMR-1602537 to F.A.T.). Additional funding was provided by the U.S. Department of Energy (Division of Materials Sciences, Office of Basic Energy Sciences; DE-SC0003844 to F.A.T.; for MD simulations and partially for EM studies). Rohit Subramanian was supported by the National Institute of Health Chemical Biology Interfaces Training Grant UC San Diego (T32GM112584-01). ns-TEM data were collected at the UCSD EM facilities supported by funding to T. S. B. from the NIH (R01-GM033050) and the Agouron Foundation. SAXS data were collected at Stanford Synchrotron Radiation Lightsource (SSRL),

supported by the DOE Office of Science, Office of Basic Energy Sciences under contract No. DE-AC02-76SF00515.

Chapter 3 is reproduced, in part, with permission, from Subramanian, R. H. *; Smith, S. J. *; Alberstein, R. G.; Bailey, J. B.; Zhang, L.; Cardone, G.; Suominen, L.; Chami, M.; Stahlberg, H.; Baker, T., Tezcan, F.A. Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions, *ACS Cent. Sci.* **2018**, *4* (11), 1578-1586.

The dissertation author is primary author on all reprinted materials.

3.7 References

- (1) Seeman, N. C.; Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3*, 17068.
- (2) Hong, F.; Zhang, F.; Liu, Y.; Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **2017**, *117* (20), 12584-12640.
- (3) Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **2006**, *440*, 297-302.
- (4) Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **2012**, *338* (6111), 1177-1183.
- (5) Wagenbauer, K. F.; Sigl, C.; Dietz, H. Gigadalton-scale shape-programmable DNA assemblies. *Nature* **2017**, *552*, 78-83.
- (6) Geary, C.; Rothemund, P. W. K.; Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **2014**, *345* (6198), 799-804.
- (7) Han, D.; Qi, X.; Myhrvold, C.; Wang, B.; Dai, M.; Jiang, S.; Bates, M.; Liu, Y.; An, B.; Zhang, F. et al. Single-stranded DNA and RNA origami. *Science* **2017**, *358* (6369), eaao2648.
- (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. USA* **2001**, *98* (5), 2217-2221.
- (9) Ringler, P.; Schulz, G. E. Self-Assembly of Proteins into Designed Networks. *Science* **2003**, *302* (5642), 106-109.

- (10) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **2010**, *43* (5), 661-672.
- (11) 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*, 558-562.
- (12) Brodin, J. D.; Ambroggio, X. I.; Tang, C.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **2012**, *4*, 375-382.
- (13) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, 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.
- (14) Song, W. J.; Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014**, *346* (6216), 1525-1528.
- (15) 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*, 369-373.
- (16) Votteler, J.; Ogohara, C.; Yi, S.; Hsia, Y.; Nattermann, U.; Belnap, D. M.; King, N. P.; Sundquist, W. I. Designed proteins induce the formation of nanocage-containing extracellular vesicles. *Nature* **2016**, *540*, 292-295.
- (17) 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. et al. Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **2017**, *552* (7685), 415-420.
- (18) Cate, J. H.; Yusupov, M. M.; Yusupova, G. Z.; Earnest, T. N.; Noller, H. F. X-ray Crystal Structures of 70S Ribosome Functional Complexes. *Science* **1999**, *285* (5436), 2095-2104.
- (19) Lin, J.; Ly, H.; Hussain, A.; Abraham, M.; Pearl, S.; Tzfati, Y.; Parslow, T. G.; Blackburn, E. H. A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (41), 14713-14718.
- (20) Eitoku, M.; Sato, L.; Senda, T.; Horikoshi, M. Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. *Cell. Mol. Life Sci.* **2008**, *65* (3), 414-444.
- (21) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816-821.

- (22) Niemeyer, C. M. Semisynthetic DNA–Protein Conjugates for Biosensing and Nanofabrication. *Angew. Chem. Int. Ed.* **2010**, *49* (7), 1200-1216.
- (23) Fu, J.; Yang, Y. R.; Johnson-Buck, A.; Liu, M.; Liu, Y.; Walter, N. G.; Woodbury, N. W.; Yan, H. Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotech.* **2014**, *9*, 531-536.
- (24) Malo, J.; Mitchell, J. C.; Vénien-Bryan, C.; Harris, J. R.; Wille, H.; Sherratt, D. J.; Turberfield, A. J. Engineering a 2D Protein–DNA Crystal. *Angew. Chem. Int. Ed.* **2005**, *44* (20), 3057-3061.
- (25) Praetorius, F.; Dietz, H. Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes. *Science* **2017**, *355* (6331), eaam5488.
- (26) Delebecque, C. J.; Lindner, A. B.; Silver, P. A.; Aldaye, F. A. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science* **2011**, *333* (6041), 470-474.
- (27) Song, J.; Li, Z.; Wang, P.; Meyer, T.; Mao, C.; Ke, Y. Reconfiguration of DNA molecular arrays driven by information relay. *Science* **2017**, *357*, eaan3377.
- (28) Brodin, J. D.; Auyeung, E.; Mirkin, C. A. DNA-mediated engineering of multicomponent enzyme crystals. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (15), 4564-4569.
- (29) McMillan, J. R.; Brodin, J. D.; Millan, J. A.; Lee, B.; Olvera de la Cruz, M.; Mirkin, C. A. Modulating Nanoparticle Superlattice Structure Using Proteins with Tunable Bond Distributions. *J. Am. Chem. Soc.* **2017**, *139* (5), 1754-1757.
- (30) Kashiwagi, D.; Sim, S.; Niwa, T.; Taguchi, H.; Aida, T. Protein Nanotube Selectively Cleavable with DNA: Supramolecular Polymerization of “DNA-Appended Molecular Chaperones”. *J. Am. Chem. Soc.* **2018**, *140* (1), 26-29.
- (31) Mou, Y.; Yu, J.-Y.; Wannier, T. M.; Guo, C.-L.; Mayo, S. L. Computational design of co-assembling protein–DNA nanowires. *Nature* **2015**, *525*, 230-233.
- (32) Kukulka, F.; Niemeyer, C. M. Synthesis of fluorescent oligonucleotide–EYFP conjugate: Towards supramolecular construction of semisynthetic biomolecular antennae. *Org. Biomol. Chem.* **2004**, *2* (15), 2203-2206.
- (33) Brodin, J. D.; Carr, J. R.; Sontz, P. A.; Tezcan, F. A. Exceptionally stable, redox-active supramolecular protein assemblies with emergent properties. *Proc. Nat. Acad. Soc. U.S.A.* **2014**, *111* (8), 2897-2902.
- (34) Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A. Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **2015**, *137* (33), 10468-10471.

- (35) Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A. Controlling Protein–Protein Interactions through Metal Coordination: Assembly of a 16-Helix Bundle Protein. *J. Am. Chem. Soc.* **2007**, *129* (44), 13374-13375.
- (36) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. In *Methods in Enzymology*; Pecoraro, V. L., Ed.; Academic Press, **2016**; Vol. 580.
- (37) Horcas, I.; Fernández, R.; Gomez-Rodriguez, J.; Colchero, J.; Gómez-Herrero, J.; Baro, A. WSXM: a software for scanning probe microscopy and a tool for nanotechnology. *Rev. Sci. Instrum.* **2007**, *78* (1), 013705.
- (38) Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **2005**, *152* (1), 36-51.
- (39) Zheng, S. Q.; Palovcak, E.; Armache, J.-P.; Verba, K. A.; Cheng, Y.; Agard, D. A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **2017**, *14*, 331.
- (40) Biyani, N.; Righetto, R. D.; McLeod, R.; Caujolle-Bert, D.; Castano-Diez, D.; Goldie, K. N.; Stahlberg, H. Focus: The interface between data collection and data processing in cryo-EM. *J. Struct. Biol.* **2017**, *198* (2), 124-133.
- (41) Gipson, B.; Zeng, X.; Zhang, Z. Y.; Stahlberg, H. 2dx—User-friendly image processing for 2D crystals. *J. Struct. Biol.* **2007**, *157* (1), 64-72.
- (42) Scherer, S.; Kowal, J.; Chami, M.; Dandey, V.; Arbeit, M.; Ringler, P.; Stahlberg, H. 2dx_automator: Implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline. *J. Struct. Biol.* **2014**, *186* (2), 302-307.
- (43) Macke, T. J.; Case, D. A. In *Molecular Modeling of Nucleic Acids*; American Chemical Society, 1997; Vol. 682.
- (44) Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.3*. **2010**.
- (45) Tang, G.; Peng, L.; Baldwin, P. R.; Mann, D. S.; Jiang, W.; Rees, I.; Ludtke, S. J. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **2007**, *157* (1), 38-46.
- (46) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605-1612.
- (47) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33-38.
- (48) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525-537.

- (49) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284-2295.
- (50) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586-3616.
- (51) Foloppe, N.; MacKerell, J., Alexander D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. **2000**, *21* (2), 86-104.
- (52) Autenrieth, F.; Tajkhorshid, E.; Baudry, J.; Luthey-Schulten, Z. Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comput. Chem.* **2004**, *25* (13), 1613-1622.
- (53) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781-1802.
- (54) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084-3095.
- (55) D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman AMBER 2016. University of California, San Francisco. **2016**.
- (56) Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **2000**, *78* (3), 1606-1619.
- (57) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (18), 10037.
- (58) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. **2007**, *35* (suppl_2), W522-W525.

Chapter 4: Constructing Protein Polyhedra via Orthogonal Chemical Interactions

4.1 Abstract

Many proteins exist naturally as symmetrical homooligomers or homopolymers¹. The emergent structural and functional properties of such protein assemblies have inspired extensive efforts in biomolecular design²⁻⁵. As synthesized by ribosomes, proteins are inherently asymmetric. Thus, they must acquire multiple surface patches that selectively associate to generate the different symmetry elements needed to form higher-order architectures^{1,6} — a daunting task for protein design. Here we address this problem using an inorganic chemical approach, whereby multiple modes of protein–protein interactions and symmetry are simultaneously achieved by selective, ‘one-pot’ coordination of soft and hard metal ions. We show that a monomeric protein (protomer) appropriately modified with biologically inspired hydroxamate groups and zinc-binding motifs assembles through concurrent Fe³⁺ and Zn²⁺ coordination into discrete dodecameric and hexameric cages. Our cages closely resemble natural polyhedral protein architectures^{7,8} and are, to our knowledge, unique among designed systems⁹⁻¹³ in that they possess tightly packed shells devoid of large apertures. At the same time, they can assemble and disassemble in response to diverse stimuli, owing to their heterobimetallic construction on minimal interprotein-bonding footprints. With stoichiometries ranging from [2 Fe:9 Zn:6 protomers] to [8 Fe:21 Zn:12 protomers], these protein cages represent some of the compositionally most complex protein assemblies—or inorganic coordination complexes—obtained by design.

4.2 Introduction

Cage-like architectures have featured prominently in supramolecular design, owing to their aesthetically appealing structures and isolated interiors, which enable them to encapsulate molecular cargo and to perform selective chemical transformations^{14–18}. Inspired by naturally occurring polyhedral assemblies, protein engineers have combined principles of symmetry with the proper design and arrangement of noncovalent interfaces to build diverse supramolecular architectures^{9–13}. However, some of the key structural features of natural protein cages have been difficult to emulate (**Figure 4.1a**). First, each cage is invariably composed of asymmetric protomers, which possess multiple self-associative patches to simultaneously satisfy the symmetry requirements necessary to build polyhedral assemblies (that is, concurrent generation of at least C_2 and C_3 symmetries, in addition to C_4 or C_5 symmetries for octahedra or icosahedra)^{1,6}. Second, these self-associative patches collectively occupy a large fraction of the surface area on each protomer, enabling the formation of tightly packed shells with small apertures to enable the influx and efflux of select species⁸. Third, although the inter-protomer interfaces in natural protein cages are extensive, to ensure stable and selective association, they are also often conformationally flexible and chemically tunable, allowing the cages to undergo cooperative motions or disassembly in response to external cues^{7,19}.

Given the difficulty of designing multiple, selectively associative surfaces on a protomer, construction of artificial cages has relied exclusively on using natively oligomeric proteins or designed peptides with $C_{n \geq 2}$ symmetries as building blocks and the design of a single type of binary protein–protein interaction (PPI) through computation¹⁰, genetic fusion^{9,11}, disulfide bond formation¹² or metal coordination^{11,13}. Although these strategies can yield polyhedral symmetries, the resulting architectures are highly porous, do not display externally controllable assembly or

disassembly (with two exceptions)^{11,13} and cannot be easily modified to adopt alternative structures (that is, they are not modular or flexible). Inspired by previous work on bimetallic supramolecular coordination cages^{20,21}, we investigated whether these design problems could be addressed using an inorganic chemical approach, wherein a protomer is equipped with chemically orthogonal coordination motifs to self-assemble into polyhedral architectures.

4.3 Results and Discussion

4.3.1 Design of bimetallic protein cages

Previously, we have taken advantage of the simultaneous strength, lability and directionality of metal coordination bonds (particularly those formed by late first-row, low-valent transition metal ions) to effect the self-assembly of discrete protein complexes⁵ and extended one-, two- and three-dimensional arrays^{22,23}. Typically, selective nucleation sites for metal-mediated PPIs are formed by pairs of metal-binding amino acids (mainly His, Asp and Glu residues) (**Figure 4.1b**) or non-native bidentate functionalities (for example, 2,2'-bipyridine, 1,10-phenanthroline and 8-hydroxyquinoline)^{5,24}. However, all of these natural or synthetic coordination motifs can be considered as soft (or intermediate-soft) according to the Hard-Soft Acid-Base (HSAB) classification²⁵ and have considerable overlap in terms of their coordination preferences for soft, low-valent transition metal ions. Owing to this lack of chemical discrimination, it has not been possible to design a heterometallic protein complex for which the self-assembly is selectively guided by multiple metal ions that mediate different PPIs.

To achieve this goal, we turned to a bidentate chelating motif, hydroxamate (HA, the conjugate base of hydroxamic acid), a common functional group found in bacterial siderophores to enable exceptionally stable coordination of Fe³⁺ ions^{26,27}. HA groups preferentially form

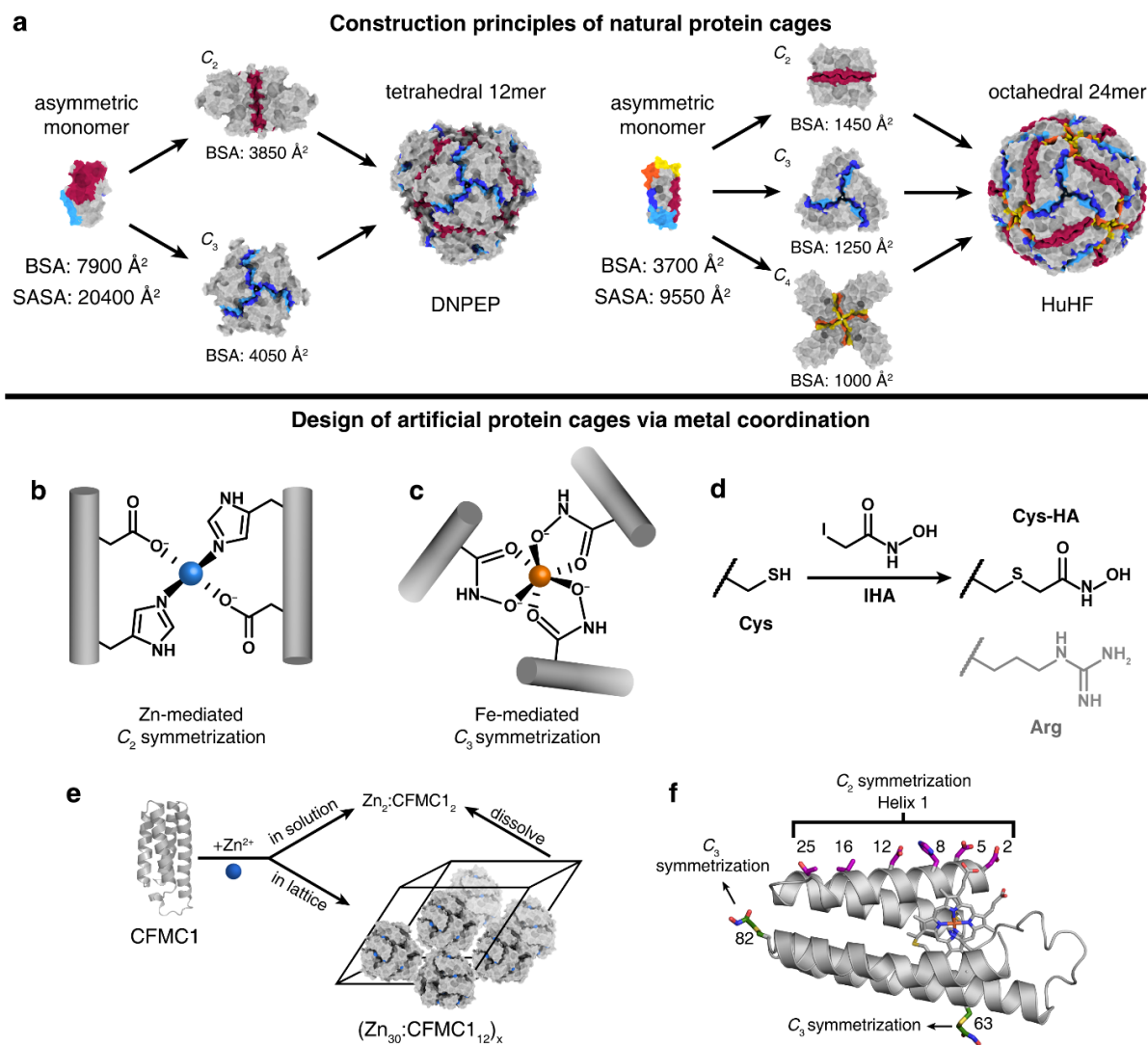


Figure 4.1 Design of protein cages. **(a)** Representative examples of natural protein cages (DNPEP – aspartyl aminopeptidase, HuHF – human heavy-chain ferritin) and their assembly from asymmetric protomers. Per-protomer solvent-accessible surface areas (SASA) and buried surface areas (BSA) are indicated. Associative surfaces on the protomers are colored in red for homologous interactions and in orange/yellow or blue/cyan for heterologous interactions. **(b)** C_2 -symmetric protein dimerization induced by tetrahedral Zn^{2+} coordination of native amino acid sidechains. **(c)** C_3 -symmetric protein trimerization induced by octahedral Fe^{3+} -tris-hydroxamate coordination. **(d)** Scheme showing modification of native Cys sidechains with IHA to yield Cys-HA, which is isosteric with arginine (light grey). **(e)** Zn-mediated solution dimerization and crystallization of CFMC1. **(f)** Structural overview of the cytochrome cb_{562} scaffold. Salient structural elements are shown as sticks.

octahedral Fe^{3+} complexes with an inherent C_3 symmetry that we sought to impose on protein oligomerization (**Figure 4.1c**). Notably, the formation constants of $\text{Fe}^{3+}:(\text{HA})_3$ complexes ($>10^{28} \text{ M}^{-3}$) are vastly higher than those of other metal–HA complexes, such that they can be considered as orthogonal to the aforementioned soft metal–ligand combinations^{26,27}. For protein derivatization, we synthesized a small reagent, iodo-hydroxamic acid (IHA), which selectively reacts with Cys residues (**Figure 4.1d**, **Figure 4.2**). The resulting Cys–HA side chain is isosteric with that of arginine and devoid of bulky aromatic moieties, furnishing a pseudo-natural amino acid functionality with the ability to chelate hard metal ions and induce C_3 symmetric oligomerization on a single-residue footprint.

As a model system, we used cytochrome *cb*₅₆₂, a monomeric four helix-bundle protein that has proved to be a versatile building block for metal-directed protein self-assembly⁵. A variant of cyt *cb*₅₆₂ (CFMC1), which was designed and observed to form Zn-mediated dimers in solution, crystallizes into rhombohedral lattices in which the protomers arrange into dodecameric, cage-like units via Zn-mediated crystal packing interactions²⁸ (**Figure 4.1e**). Whereas Zn-mediated interactions were not sufficiently strong to maintain the tetrahedral dodecamers upon crystal dissolution, we envisioned that these lattice units could serve as a structural model to engineer the protomers such that they would form self-standing cages. Looking first to stabilize the C_2 symmetric interfaces, we incorporated a bidentate His8–Asp12 motif to mediate the antiparallel association of two protomers along their helices 1 via tetrahedral Zn^{2+} coordination (**Figure 4.1f**). Given that C_3 symmetric interfaces are small and heterologous (that is, they involve two different patches on each protomer; **Figure 4.3a**), they were unsuitable for stabilization by noncovalent interactions. Therefore, we focused on the central pores in each C_3 symmetric substructure and identified positions 63 and 82 as suitable locations for installing Cys–HA

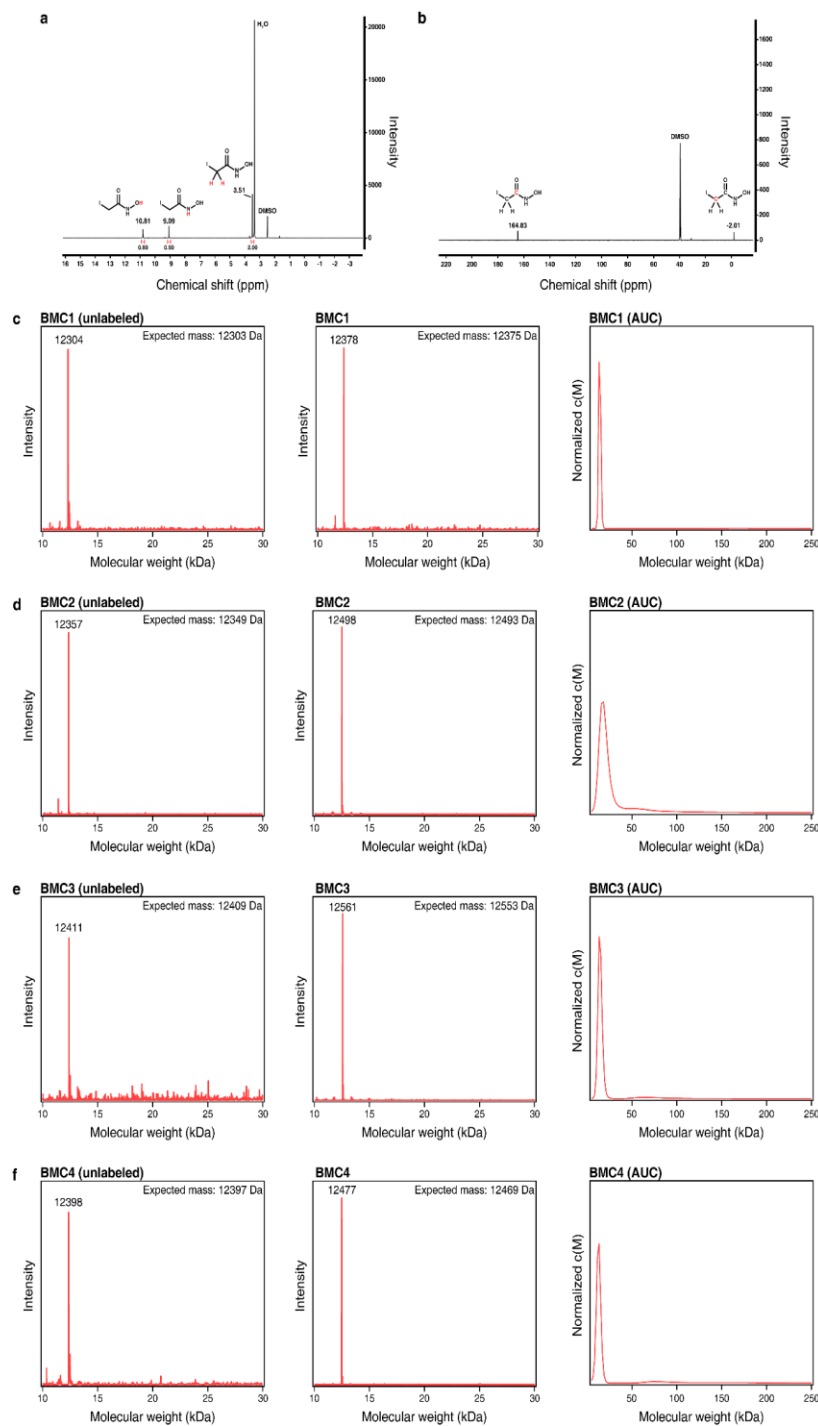


Figure 4.2 Characterization of the IHA ligand and the BMC constructs. NMR spectra of *N*-hydroxy-2-iodoacetamide in DMSO- d_6 : (a) ^1H (b) ^{13}C . ESI-MS of as-isolated and HA-functionalized BMC constructs, and AUC profiles of HA-functionalized protomers for (c) BMC1 (d) BMC2 € BMC3 and (f) BMC4. The calculated masses for each unlabeled protein are determined by summing the mass of the polypeptide sequence and the *c*-type heme (618 Da) covalently linked to the cytochrome.

functionalities, which would stabilize trimeric substructures by forming $\text{Fe}^{3+}:(\text{HA})_3$ centers (**Figure 4.1f**). Thus, we prepared two CFMC1 variants designated bimetallic cage 1 and bimetallic cage 2 (BMC1 and BMC2; **Figure 4.3b**). Both BMC1 and BMC2 bear the His8–Asp12 motif on helix 1 and Cys63–HA along with the native peripheral Zn coordination sites (Ala1N-term, Asp39 and His77) of the parent CFMC1 structure. BMC2 additionally contains Cys82–HA (**Figure 4.2, 4.3b**).

Table 4.1 Amino acid sequences of protein variants employed in the current research. Mutated amino acids in reference to parent CFMC1 are bolded and underlined.

	10	20	30	40	50	60	70	80	90	100		
CFMC1	ADLED NMETL	NDNLK VIEKA	DNAAQ VKDAL	TKMAA AAADA	WSATP PKLED	KSPDS PEMHD	FRHGF WCLIG	QIHAA LHLAN	EGKVK EAQAA	AEQLK TTCNA	CHQKYR	
BMC1	ADLED NM <u>HTL</u>	NDNLK VIEKA	DNA <u>TT</u>	VKDAL TKM <u>QA</u>	AA <u>QDA</u>	WSATP PKLED	KSPDS PEM <u>SD</u>	FRC <u>GF</u>	WELIG QI <u>NAA</u>	LHL <u>AK</u>	<u>Q</u> GKVK EAQAA AEQLK TTCNA CHQKYR	
BMC2	ADLED NM <u>HTL</u>	NDNLK VIEKA	DNA <u>TT</u>	VKDAL TKM <u>QA</u>	AA <u>QDA</u>	WSATP PKLED	KSPDS PEM <u>SD</u>	FRC <u>GF</u>	WELIG QI <u>NAA</u>	LHL <u>AK</u>	<u>Q</u> GKVK EAQAA AEQLK TTCNA CHQKYR	
BMC3	ADLE <u>H</u> NM <u>HTL</u>	NDNLK <u>H</u> IEKA	DNA <u>TT</u>	VKDAL TKM <u>QA</u>	AA <u>QDA</u>	WSATP PKLED	KSPDS PEM <u>SD</u>	FRC <u>GF</u>	WELIG QI <u>NAA</u>	LHL <u>AK</u>	<u>Q</u> GKVK EAQAA AEQLK TTCNA CHQKYR	
BMC4	<u>A</u> E <u>L</u> E <u>E</u>	NM <u>HTL</u>	NDNLK <u>H</u> IEKA	DNA <u>A</u> E	VKDAL TKM <u>QA</u>	AA <u>QDA</u>	WSATP PKLED	KSPDS PEM <u>SD</u>	FR <u>S</u> GF	WELIG QI <u>NAA</u>	LHL <u>AK</u>	<u>Q</u> GKVK EAQAA AEQLK TTCNA CHQKYR

Crystals of BMC1 and BMC2 were obtained in the presence of near equimolar ZnCl_2 and FeSO_4 . These crystals were isomorphic ($R32$ space group; $a = b = 126 \pm 1 \text{ \AA}$, $c = 167 \pm 1 \text{ \AA}$) with those of CFMC1²⁸, indicating that they possessed the same underlying lattice structure composed of dodecameric units (**Table 4.1, Figure 4.3c, d**). Crystals were dissolved in a solution lacking the precipitating agent (PEG-400) and then analyzed by negative-stain transmission electron microscopy (ns-TEM; **Figure 4.4a, 4.5**). The images revealed uniform particles with a diameter of $8.4 \pm 0.8 \text{ nm}$ in the case of BMC2 but not BMC1, implying that two HA coordination motifs are necessary for cage stability. Analysis of the same BMC2 solution by analytical ultracentrifugation (AUC) indicated a predominant species with a molecular weight (MW_{obs}) of about 140 kDa (**Figure 4.4b**), approximating the calculated value (MW_{calc}) of 150 kDa for a dodecamer. BMC2 particles dissociated upon treatment with ethylenediamine tetraacetic acid (EDTA), confirming their metal-dependent self-assembly (**Figure 4.4b, 4.5**).

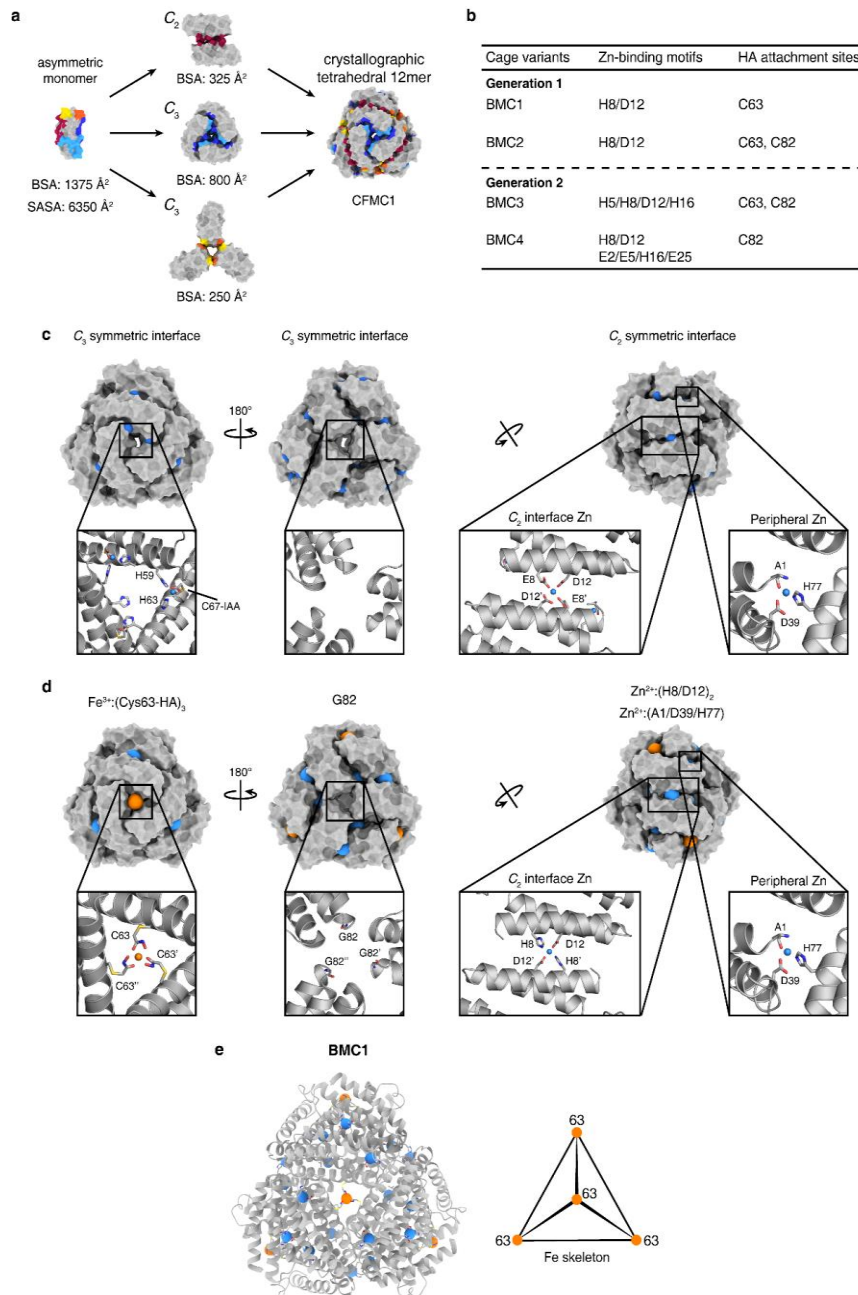


Figure 4.3 Structural comparison of CFMC1 and BMC1 cages. **(a)** The symmetric substructures of the CFMC1 dodecameric unit and its per-protomer SASA and BSA values. Associative surfaces on the protomers are colored in red for homologous interactions and in red/orange or blue/cyan for heterologous interactions (right). **(b)** Summary of engineered metal-coordination motifs for BMC constructs (see **Table 4.1** for all mutations). Comparison of C_2 and C_3 symmetric interfaces and corresponding metal binding sites for **(c)** CFMC1 and **(d)** BMC1. Full cages are shown as surfaces; insets show details of each interface. Fe and Zn ions are represented as orange and teal spheres, respectively. **(e)** Cartoon representation of a full-size BMC1 cage with all metal ions shown as spheres. PDB ID: 3M4B (CFMC1), 6OT9 (BMC1).

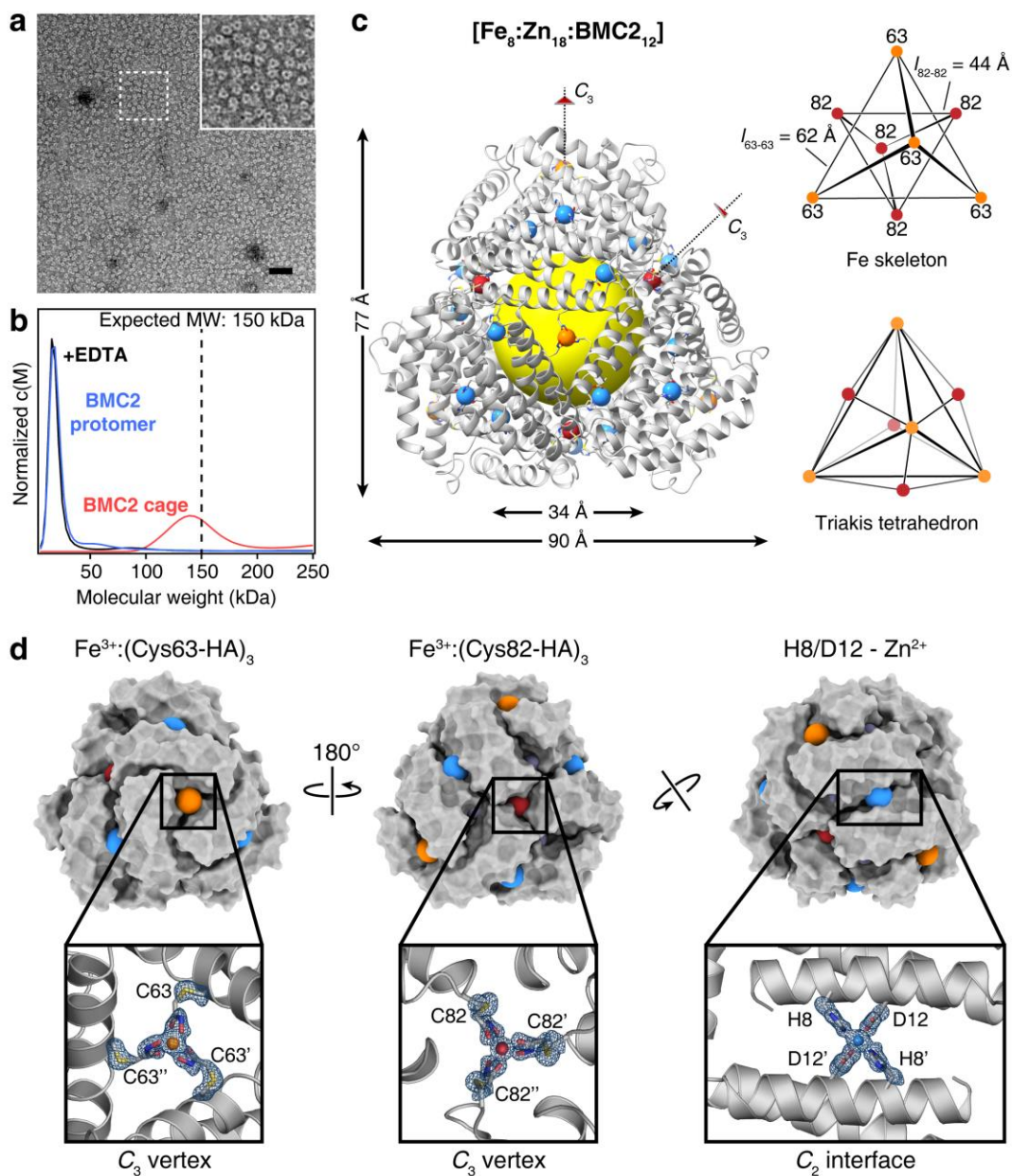


Figure 4.4 Characterization of BMC2 cages. **(a)** ns-TEM of BMC2 cages obtained by the dissolution of 3D crystals; the inset is a close-up of the boxed region. Scale bar = 50 nm. **(b)** AUC characterization of BMC2 protomers, BMC2 cages after crystal dissolution and after subsequent treatment with EDTA. **(c)** Crystal structure of the BMC2 cage. Fe and Zn ions are represented as orange/red and blue spheres, respectively. The central cavity is highlighted by a yellow sphere. Two types of C_3 vertices formed by $\text{Fe}:(\text{Cys63-HA})_3$ and $\text{Fe}:(\text{Cys82-HA})_3$ coordination motifs form two superimposed tetrahedra to generate a triakis tetrahedron. **(d)** Surface representations of the BMC2 cage, with metal ions shown as colored spheres. Atomic details of each metal coordination site are shown in the insets, with the $mF_o - DF_c$ electron density omit map (blue mesh) contoured at 3σ .

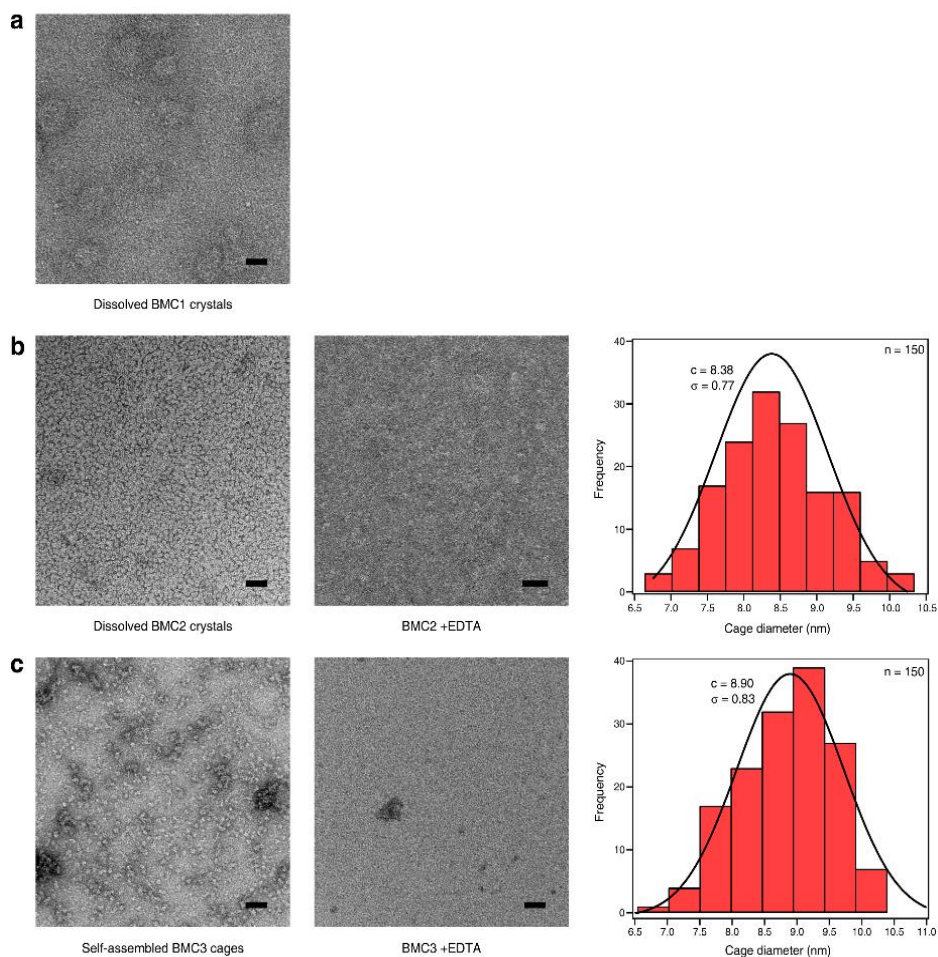


Figure 4.5 ns-TEM characterization of BMC constructs. Dissolved (a) Fe:Zn:BMC1 and (b) Fe:Zn:BMC2 crystals in a buffer containing 100 mM HEPES (pH 7.5), 200 mM MgCl₂ and 800 μM ZnCl₂. (c) Self-assembled Fe:Zn:BMC3 cages in a buffer containing 20 mM Tris (pH 8.5), 20 μM FeSO₄ and 60 μM ZnCl₂. Histograms in (b) and (c) reflect the size distributions of Fe:Zn:BMC2 and Fe:Zn:BMC3 cage diameters as measured from ns-TEM images. Gaussian fits to both distributions are drawn as solid lines along with their centers and standard deviations reported. Scale bars are 50 nm.

We determined the crystal structure of the BMC2 cage at 1.4 Å resolution (**Table 4.2**), revealing a compact structure with the shape of a truncated tetrahedron, outer dimensions of 80 × 90 Å and a cavity volume of 32,700 Å³ (**Figure 4.4c, 4.6**). Like natural protein cages, the shell is tightly packed and the largest opening measures less than 4 Å across. Nearly 30% of the surface area of each protomer (1,700 Å² out of 6,500 Å²) is buried in interfaces despite a design footprint of only four amino acids (His8, Asp12, Cys63 and Cys82).

Table 4.2 X-ray data collection, processing and refinement statistics. Numbers in parentheses correspond to the highest resolution shell.

	BMC1	BMC2	BMC3	BMC4
Data collection				
Space group	R 3 2	R 3 2	R 3 2	P 6 ₃ 2 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	125.6, 125.6, 166.4	126.1, 126.1, 168.2	126.7, 126.7, 167.8	87, 87, 63.3
α , β , γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	39.92 – 2.40 (2.46 – 2.40)	39.25 – 1.40 (1.44 – 1.40)	91.83 – 1.85 (1.90 – 1.85)	48.47 – 1.50 (1.54 – 1.50)
No. Reflections Observed	393414 (28956)	1863896 (88779)	885564 (66924)	734764 (18399)
No. Reflections Unique	38245 (2875)	195424 (14465)	85783 (6342)	37652 (1334)
<i>R</i> _{merge}	0.185 (2.986)	0.056 (1.830)	0.096 (2.377)	0.053 (1.936)
<i>I</i> / σ <i>I</i>	8.1 (0.96)	20.3 (0.84)	13.7 (1.04)	26.6 (1.23)
<i>CC</i> 1/2	0.998 (0.501)	1.000 (0.410)	0.999 (0.468)	1.000 (0.480)
Completeness (%)	99.81 (100.00)	99.37 (98.9)	100.00 (100.00)	87.66 (42.2)
Redundancy	10.29 (10.01)	9.54 (6.138)	10.32 (10.54)	19.51 (13.94)
Wilson B (Å) ²	52	19	36	29
Refinement				
Resolution (Å)	36.37 – 2.40 (2.43 – 2.40)	31.54 - 1.40 (1.42 - 1.40)	91.83 - 1.85 (1.88 - 1.85)	48.47 - 1.50 (1.53 - 1.50)
No. reflections	38194 (1286)	195270 (8377)	85755 (3974)	37642 (810)
<i>R</i> _{work} / <i>R</i> _{free}	0.2174/0.2718	0.1659/0.1909	0.1826/0.2106	0.1900/0.2181
No. atoms				
Protein	3292	3502	3459	925
Ligand/ion	200	244	240	58
Water	19	786	342	124
<i>B</i> -factors (Å) ²				
Protein	70	24	41	42
Ligand/ion	68	22	43	35
Water	63	37	46	49
R.m.s. deviations				
Bond lengths (Å)	0.010	0.011	0.009	0.013
Bond angles (°)	1.29	1.33	1.15	1.33
Clashscore	7	5	7	10
Ramachandran favored (%)	100	100	100	97
Ramachandran allowed (%)	0	0	0	3
Ramachandran outliers (%)	0	0	0	0
Rotamer outliers (%)	1	2	2	1

Numbers in parentheses correspond to the highest-resolution shell.

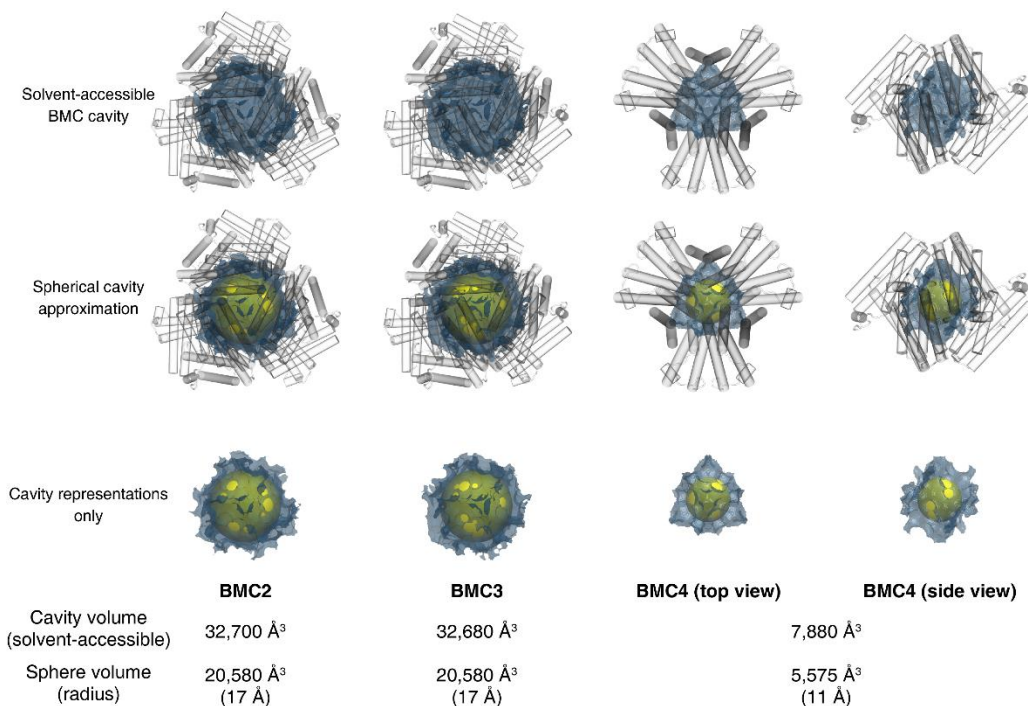


Figure 4.6 Cavity volumes of BMC cages. Solvent-accessible cavity volumes within BMC cages as calculated by a 1.4-Å rolling probe are shown visually as blue meshes and reported numerically below. Spherical cavities, shown as yellow spheres in **Figure 4.4** and **Fig. 4**, are reproduced for comparison to the calculated volumes. BMC proteins are represented as transparent cylinders.

The full complement of metal ions, comprising eight Fe ions (four each in the C_3 symmetric pores) and eighteen Zn ions (six in C_2 interfaces and twelve in peripheral sites) are clearly resolved (**Figure 4.4c**). Anomalous X-ray diffraction data collected at and below Fe and Zn K-edges indicate that the designed Fe- and Zn-coordination sites exclusively bind to their cognate ions with no evidence of crosstalk (**Figure 4.7, Tables 4.3-4.6**), which establishes that the metal-dependent self-assembly of BMC2 cages occurs with absolute chemical selectivity. The Fe centers form the eight C_3 vertices of a triakis tetrahedron, a Catalan solid with twelve equivalent faces (**Figure 4.4c**). It can be viewed as the superposition of two tetrahedra: four Fe centers that are coordinated by Cys63–HA motifs generate the larger of these two tetrahedra (with an edge length (l_{edge}) of 62 Å), and four Fe centers coordinated by Cys82–HA motifs produce the smaller one

($l_{\text{edge}} = 44 \text{ \AA}$) (**Figure 4.4c**). The BMC1 structure, in comparison, has a regular tetrahedral arrangement of four Fe centers as it lacks the Cys82–HA group (**Table 4.2, Figure 4.3**).

Figure 4.7 Anomalous densities of engineered metal binding sites and conformational flexibility of Cys82-HA site. Cartoon and stick representations of the **(a-b)** BMC1, **(c-e)** BMC2, **(f-h)** BMC3, and **(i-j)** BMC4 symmetric interfaces showing the engineered metal binding sites with the C63-HA ligands **(a, c, f)**, C82-HA ligands **(d, g, i)** and the Zn binding sites **(b,e,h,j)**. To discern between bound Zn or Fe, the difference of the anomalous signal between pairs of datasets above and below the K-shell energy of Zn and Fe respectively, are depicted as blue or orange meshes. A strong signal illustrates strong change in anomalous signal across the respective edge, in turn suggesting the presence of the respective metal. The upper right corner of each panel indicates the energies of the datasets used for the map of the respective color. All anomalous difference maps were contoured at 3σ . As datasets around the Fe-edge were not available for BMC1 and BMC3 (necessitating calculations using anomalous difference density of singular datasets), the calculated f'' for Zn at 7.3 and 9.3 keV are 0.82 and 0.52 (i.e. non-zero) and thus some residual anomalous signal of the lower energy maps around the Zn atoms is expected to result even from strictly selective Zn loading. For a more quantitative analysis of the nature of the bound metal, ratios of the anomalous signal to the expected values (lower left corner of each panel) were calculated as described in the Methods. **(k)** Stick representation of the BMC2 Cys82-HA binding site in both alternative conformations with the anomalous difference density over the “Fe-edge” shown as orange mesh and a simulated annealing omit map (omitting all C82-HA atoms and Fe) of the normal electron density as light blue mesh contoured at 2σ . For all Cys-HA binding sites, arrows indicate the handedness of the binding site as Δ (right handed) or Λ (left handed). The reversion of handedness in **k** with the respective view angle is indicated by arrows. Color code for atoms in all panels: Fe in orange, Zn in blue, S in yellow, O in red and N in dark blue.

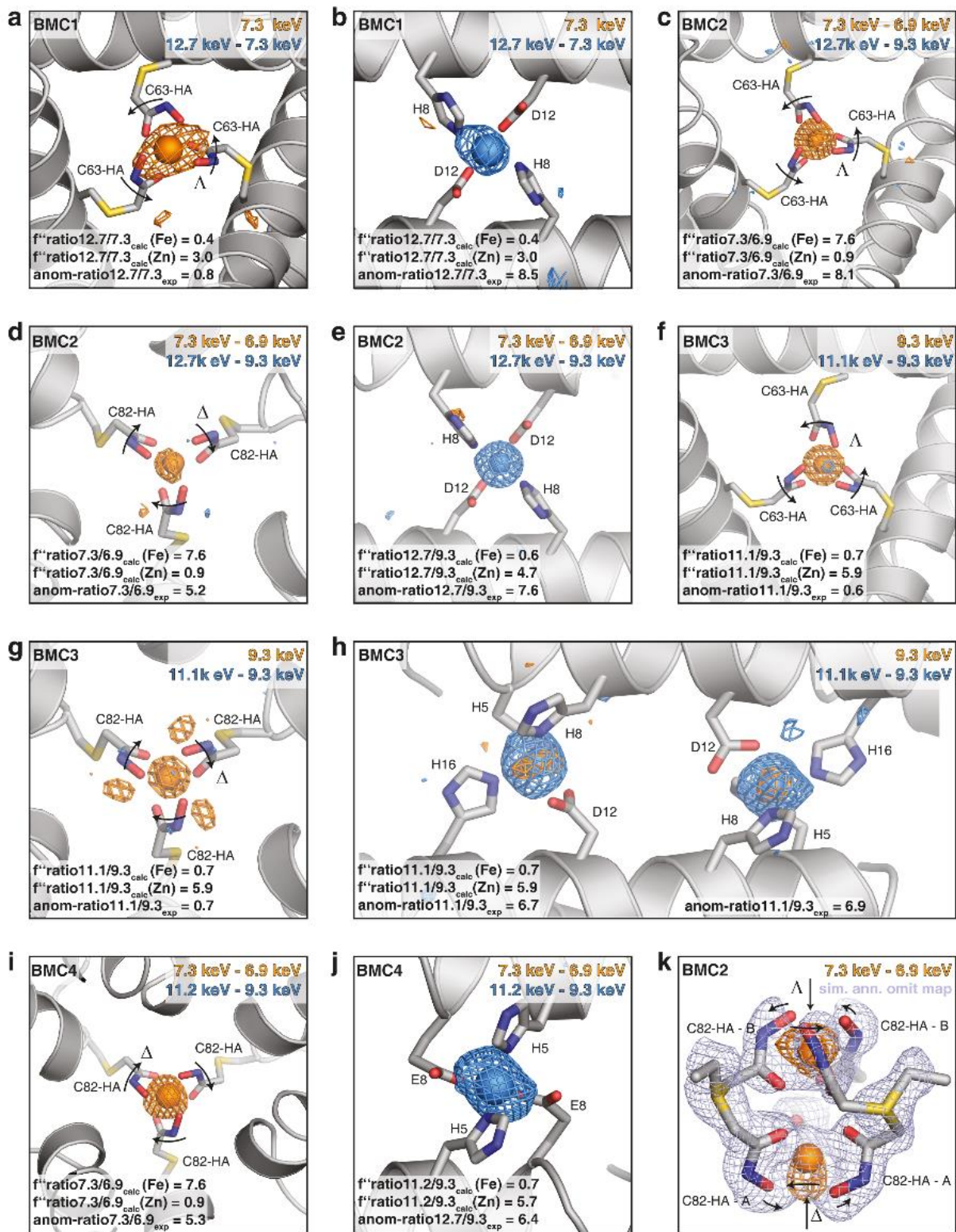


Table 4.3 Crystallographic quantification of metal content at BMC1 coordination sites.

Observed ratio 12.7 keV / 7.3 keV		Calculated approximated ratios				
		Molar Fraction Fe	Molar Fraction Zn	f' for 12.7 keV	f' for 7.3 keV	ratio f' 12.7 keV/7.3 keV
Heme-Fe	0.5					
Fe-63CHA	0.8	1	0	1.5	3.8	0.4
Zn-C ₂ -H8D12	8.5	0.8	0.2	1.7	3.2	0.5
Zn-A1D39H77	12.5	0.6	0.4	1.9	2.6	0.7
		0.4	0.6	2.1	2.0	1.0
		0.2	0.8	2.3	1.4	1.6
		0	1	2.5	0.8	3.0

Table 4.4 Crystallographic quantification of metal content at BMC2 coordination sites.

Observed ratio 12.7 keV / 9.3 keV		Calculated approximated ratios ("Zn-edge")				
		Molar Fraction Fe	Molar Fraction Zn	f' for 12.7 keV	f' for 9.3 keV	ratio f' 12.7 keV/9.3 keV
Heme-Fe	0.6					
Fe-63CHA	0.6	1	0	1.5	2.5	0.6
Fe-82CHA	0.6	0.8	0.2	1.7	2.1	0.8
Zn-C ₂ -H8D12	7.6	0.6	0.4	1.9	1.7	1.1
Zn-A1D39H77	6.2	0.4	0.6	2.1	1.3	1.6
		0.2	0.8	2.3	0.9	2.5
		0	1	2.5	0.5	4.7

Observed ratio 7.3 keV / 6.9 keV		Calculated approximated ratios ("Fe-edge")				
		Molar Fraction Fe	Molar Fraction Zn	f' for 7.3 keV	f' for 6.9 keV	ratio f' 7.3 keV/6.9 keV
Heme-Fe	9.6					
Fe-63CHA	8.1	1	0	3.8	0.5	7.6
Fe-82CHA	5.2	0.8	0.2	3.2	0.6	5.5
Zn-C ₂ -H8D12	0.5	0.6	0.4	2.6	0.7	3.9
Zn-A1D39H77	0.6	0.4	0.6	2.0	0.7	2.7
		0.2	0.8	1.4	0.8	1.7
		0	1	0.8	0.9	0.9

Table 4.5 Crystallographic quantification of metal content at BMC3 coordination sites.

Observed ratio 11.1 keV / 9.3 keV		Calculated approximated ratios				ratio f'' 11.1 keV/9.3 keV
		Molar Fraction Fe	Molar Fraction Zn	f'' for 11.1 keV	f'' for 9.3 keV	
Heme-Fe	0.6					
Fe-63CHA	0.6	1	0	1.9	2.5	0.7
Fe-82CHA	0.7	0.8	0.2	2.1	2.1	1.0
Zn-C ₂ -H8D12	6.7	0.6	0.4	2.3	1.7	1.4
Zn-A1D39H77	7.3	0.4	0.6	2.6	1.3	2.0
		0.2	0.8	2.8	0.9	3.1
		0	1	3.1	0.5	5.9

Table 4.6 Crystallographic quantification of metal content at BMC4 coordination sites.

Observed ratio 11.2 keV / 9.3 keV		Calculated approximated ratios ("Zn-edge")				ratio f'' 11.2 keV/9.3 keV
		Molar Fraction Fe	Molar Fraction Zn	f'' for 11.2 keV	f'' for 9.3 keV	
Heme-Fe	0.6					
Fe-82CHA	0.6	1	0	1.9	2.5	0.7
Zn-C ₂ -H5D8	6.4	0.8	0.2	2.1	2.1	1.0
Zn- D21E25H77	6.4	0.6	0.4	2.3	1.7	1.3
		0.4	0.6	2.5	1.3	1.9
		0.2	0.8	2.7	0.9	3.0
		0	1	3.0	0.5	5.7

Observed ratio 7.3 keV / 6.9 keV		Calculated approximated ratios ("Fe-edge")				ratio f'' 7.3 keV/6.9 keV
		Molar Fraction Fe	Molar Fraction Zn	f'' for 7.3 keV	f'' for 6.9 keV	
Heme-Fe	6.0					
Fe-82CHA	5.3	1	0	3.8	0.5	7.6
Zn-C ₂ -H5D8	0.3	0.8	0.2	3.2	0.6	5.5
Zn- D21E25H77	0.3	0.6	0.4	2.6	0.7	3.9
		0.4	0.6	2.0	0.7	2.7
		0.2	0.8	1.4	0.8	1.7
		0	1	0.8	0.9	0.9

As designed, the edges of the BMC2 tetrahedra are formed by six Zn ions located centrally in C_2 interfaces (**Figure 4.4d**). The $\text{Fe}^{3+}:(\text{Cys63-HA})_3$ and $\text{Fe}^{3+}:(\text{Cys82-HA})_3$ motifs display near-ideal octahedral geometries (**Figure 4.4d**), with the former in Λ (left handed) and the latter in Δ (right handed) configuration (**Figure 4.7**). Notably, the $\text{Fe}^{3+}:(\text{Cys82-HA})_3$ center also adopts an alternative conformation (20% abundance) owing to the flexibility of the Cys-HA side chain (**Figure 4.7k**). All Fe-O bond distances are in the range of 1.95–2.1 Å, which are typical of $\text{Fe}^{3+}:(\text{HA})_3$ complexes²⁹. Given that a Fe^{2+} precursor was used to initiate self-assembly, and as $\text{Fe}:\text{HA}_3$ centers have low reduction potentials ($E_{\text{red}} < -400$ mV)²⁶, this observation suggests that the protein self-assembly involves the initial formation of $\text{Fe}^{2+}:\text{HA}$ centers, followed by the thermodynamically favored oxidation of these species into Fe^{3+} either by the Fe^{3+} -heme centers embedded in each protomer (see Methods) or directly by ambient O_2 .

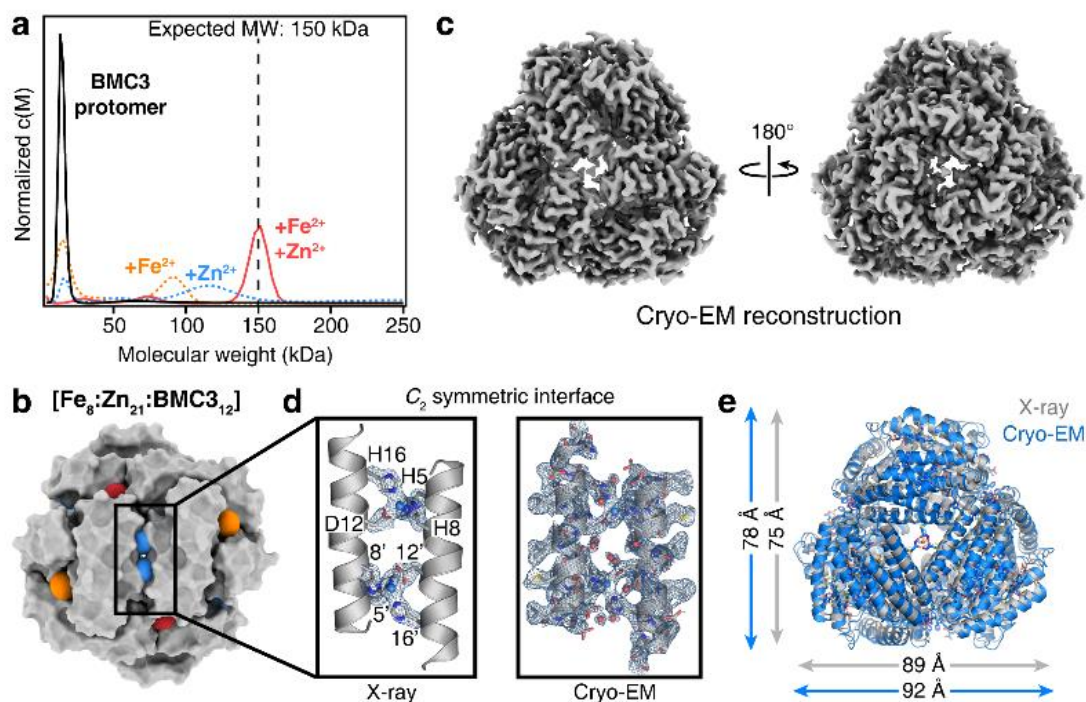


Figure 4.8 Characterization of BMC3 cages. **(a)** AUC characterization of BMC3 self-assembly. **(b)** Surface representation of the BMC3 cage (as derived from the crystal structure), oriented to show the incorporation of two Zn ions at the C_2 symmetric interface. **(c)** 2.6-Å density map for the BMC3 cage as determined by cryo-EM. **(d)** Atomic details of both Zn-binding sites of the 2-fold interface overlaid with the electron density *mFo-DFc* omit map from the crystal structure (left) and as observed for the cryo-EM structure (right). Additional sidechains and waters are shown for the cryo-EM structure to emphasize structural robustness of the interface. **(e)** Overlay of the BMC3 X-ray and cryo-EM structures to highlight the isotropic expansion of the cage in the absence of crystallographic packing interactions.

4.3.2 Reversible assembly of dodecameric cages

Despite the stability of isolated BMC2 cages, their formation required an initial crystallization step. We reasoned that slow crystal nucleation or growth kinetics and the high attendant protein and metal concentrations probably increased the fidelity and yield of the complex self-assembly process to produce a discrete supermolecule consisting of 12 protomers and 26 metal ions of two different kinds. Reasoning that strengthened Zn-mediated interactions across the C_2

interface could increase the efficiency of cage self-assembly in solution, we generated two second-generation variants based on BMC2: BMC3 and BMC4 (**Figure 4.2, 4.3b**). In BMC3, the helix 1 surface was engineered to form two Zn-coordination sites (composed of His5, His8, Asp12 and His16) across the C_2 interface, whereas in BMC4 three potential Zn-coordination sites were engineered (one central site composed of two His8–Asp12 pairs as in BMC3 and two peripheral sites composed of Glu2, Glu5, His16 and Glu25). In BMC4, we also removed the Cys63–HA group with the purpose of eliminating any potential undesired assembly products that involve heteromeric Fe^{3+} coordination by Cys63–HA and Cys82–HA.

BMC3 indeed formed dodecameric cages in solution with high yields (>80%) as determined by ns-TEM and AUC measurements (**Figure 4.5, 4.8**). The 1.85-Å resolution crystal structure confirmed the eight Fe centers in the vertices and the two Zn coordination sites in each C_2 interface (twelve in total) as well as nine of the twelve possible peripheral Zn sites, which complete a [8 Fe:21 Zn:12 protomers] architecture (**Figure 4.8b**). Notably, the self-assembly of BMC3 cages in solution was dependent on the presence of both Fe and Zn ions. The absence of either metal ion or the addition of various other first-row transition metal ions instead of Fe led to smaller oligomeric forms of BMC3 or non-specific assemblies (**Figure 4.8a, 4.9a**). BMC3 cages also formed with a Fe^{3+} precursor, $Fe(acetylacetonate)_3$ (**Figure 4.9b**). Consistent with self-assembly under thermodynamic control, the formation of dodecameric cages was independent of the order of addition of Fe or Zn ions. We determined the single-particle cryoelectron microscopy (cryo-EM) structure of isolated BMC3 cages at a resolution of 2.6 Å (**Figure 4.8c, Table 4.7**). A major portion of the assembly could be resolved at 2.0 Å or less (**Figure 4.10**). At this resolution, nearly all side chains, Zn coordination sites and some ordered water molecules are clearly distinguished (**Figure 4.8d**). Consistent with the crystallographically observed flexibility of

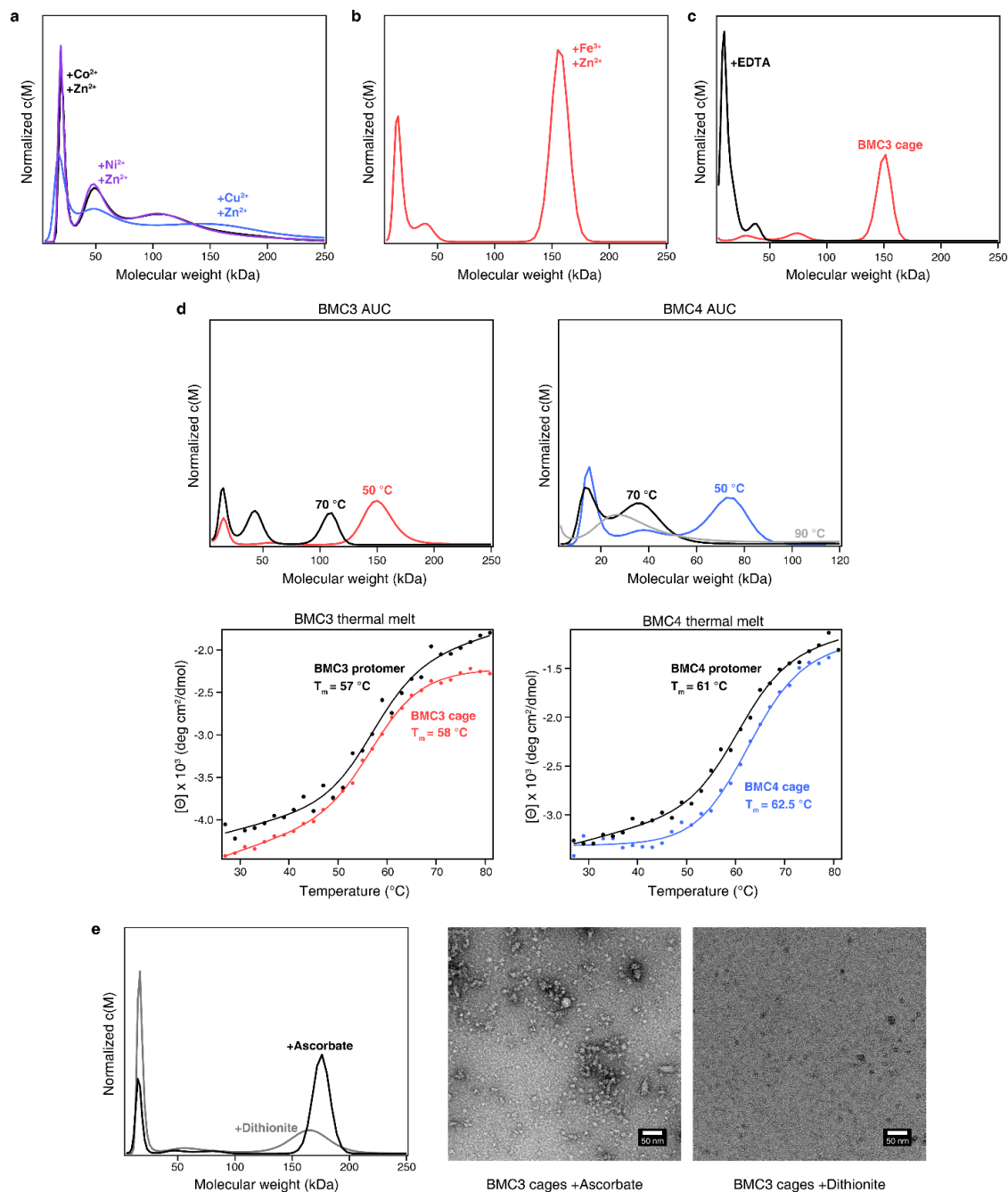


Figure 4.9 Solution characterization of self-assembled BMC3 and BMC4 cages. The oligomerization state of BMC3 cages as monitored by AUC measurements following: **(a)** incubation with various first-row transition metal ions **(b)** incubation with Zn^{2+} and Fe^{3+} ($\text{Fe}(\text{acetylacetonate})_3$) and **(c)** Disassembly via sequestration of metal ions by EDTA. **(d)** AUC profiles of BMC variants after equilibration for two hours at the indicated temperatures (top). Thermal unfolding of BMC variants as measured by circular dichroism spectroscopy at 222 nm (bottom). **(e)** Treatment with chemical reductants of different reduction potentials. ns-TEM micrographs are shown for cage samples incubated with chemical reductants.

Table 4.7 Cryo-EM data collection, processing, and refinement statistics.

BMC3 (EMDB ID: EMD-20212) (PDB ID: 6OVH)	
Data collection and processing	
Magnification	165,000x
Voltage (kV)	300
Electron dose (e ⁻ /Å ²)	60
Exposure rate (e ⁻ /Å ² /s)	6
Defocus range (μm)	0.84
Pixel size (Å)	0.3 – 2.7
Symmetry imposed	T
Total extracted particles (no.)	805,156
Final refined particles (no.)	25,391
Map resolution (Å)	2.57
FSC 0.143 (unmasked/masked)	2.62/2.57
Map resolution range (Å)	∞ – 2.57
Applied B-factor (Å ²)	-79
Refinement	
Initial model used (PDB code)	6OT7
Model resolution (Å)	2.57
FSC 0.5 (unmasked/masked)	2.74/2.72
FSC 0.143 (unmasked/masked)	2.62/2.57
Model resolution range (Å)	∞ – 2.57
Map sharpening B factor (Å ²)	-79
Model composition	
Non-hydrogen atoms	10895
Protein residues	1272
Ligand/ion	68
Water	171
B-factors (Å ²)	
Protein	51
Ligand/ion	59
Water	48
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	0.876
Validation	
MolProbity score	1.33
Clashscore	1.15
Rotamer outliers (%)	4.55
Ramachandran plot	
Favored (%)	99.04
Allowed (%)	0.96
Disallowed (%)	0.00

$\text{Fe}^{3+}:(\text{Cys-HA})_3$ coordination sites, electron densities in the C_3 vertices are diffuse and some side chains that display high temperature factors in the crystal structure are found in alternative conformations in the cryo-EM structure (**Figure 4.10**). These observations confirm that the solution architecture of the BMC3 cage closely reflects the solid-state structure. Probably owing to lattice packing, the latter is isotropically compressed by around 2–3 Å compared to the former (**Fig. 3e**), which can be accommodated by slight changes in interfacial metal coordination.

Next, we examined the assembly and disassembly behavior of BMC3 cages in response to different stimuli. BMC3 cages readily disassemble upon treatment with EDTA (**Figure 4.9c**). They were stable at 50 °C but dissociated upon incubation at 70 °C (**Figure 4.9d**). A key feature of siderophores is that their cellular release of Fe is promoted by the destabilization and labilization of their $\text{Fe}^{3+}:(\text{HA})_3$ centers through reduction to the Fe^{2+} form in the cytosol²⁷. Along these lines, the treatment of BMC3 cages with a strong reductant (dithionite; $E_{\text{red}} < -500$ mV)³⁰ led to their disappearance and the emergence of monomeric species (**Figure 4.9e**). By contrast, a weaker reductant (ascorbate; $E_{\text{red}} > -100$ mV at pH 7)³¹ with a reduction potential higher than that of $\text{Fe}^{3+}:(\text{HA})_3$ had considerably less effect (**Figure 4.9e**), suggesting that the disassembly of BMC3 cages occurs through the reduction of the Fe centers. These observations establish BMC3 cages as a distinctive system among natural and artificial protein architectures the assembly and disassembly of which can be controlled through multiple stimuli: chemical, thermal or redox. BMC3 cages can passively encapsulate small fluorogenic molecules in either their lumen or interprotomer interfaces, retain them for several days and release them upon treatment with EDTA (**Figure 4.11**).

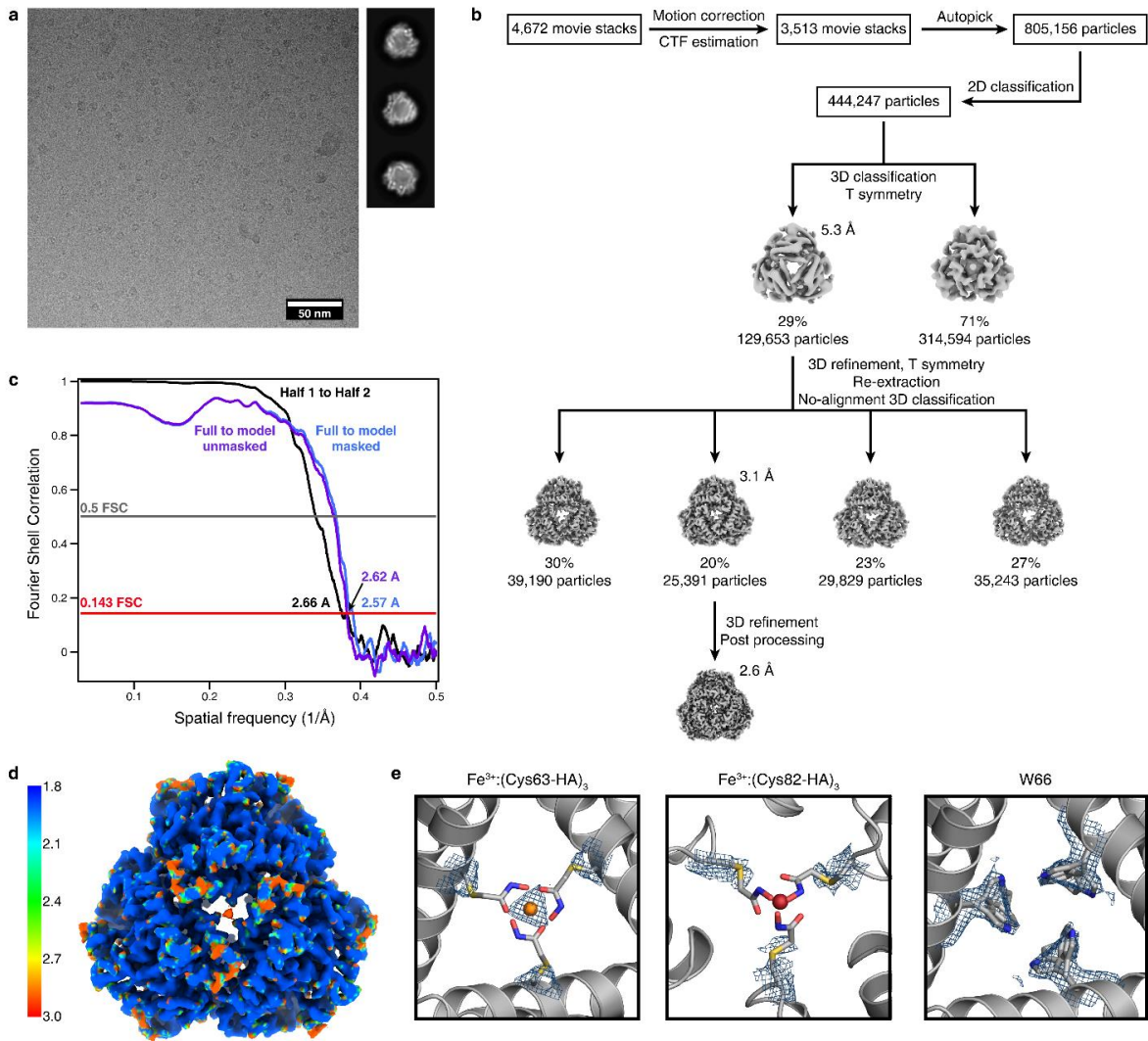


Figure 4.10 Cryo-EM analysis of BMC3 cages. **(a)** Representative cryo-EM micrograph and 2D class averages. **(b)** Flowchart detailing image processing from collected movie stacks to final map. Additional details can be found in the Methods. **(c)** FSC curves calculated between the half-maps (black line), atomic model to the unmasked full map (purple line) and atomic model to the masked full map (blue line). Resolution values are indicated at the gold-standard FSC 0.143 criterion. **(d)** Local resolution estimates of the final reconstruction calculated using ResMap. **(e)** Electron density shown at BMC3 C_3 interfaces highlighting poorly resolved density (reflecting high flexibility) at hydroxamate sites and multiple conformations of W66.

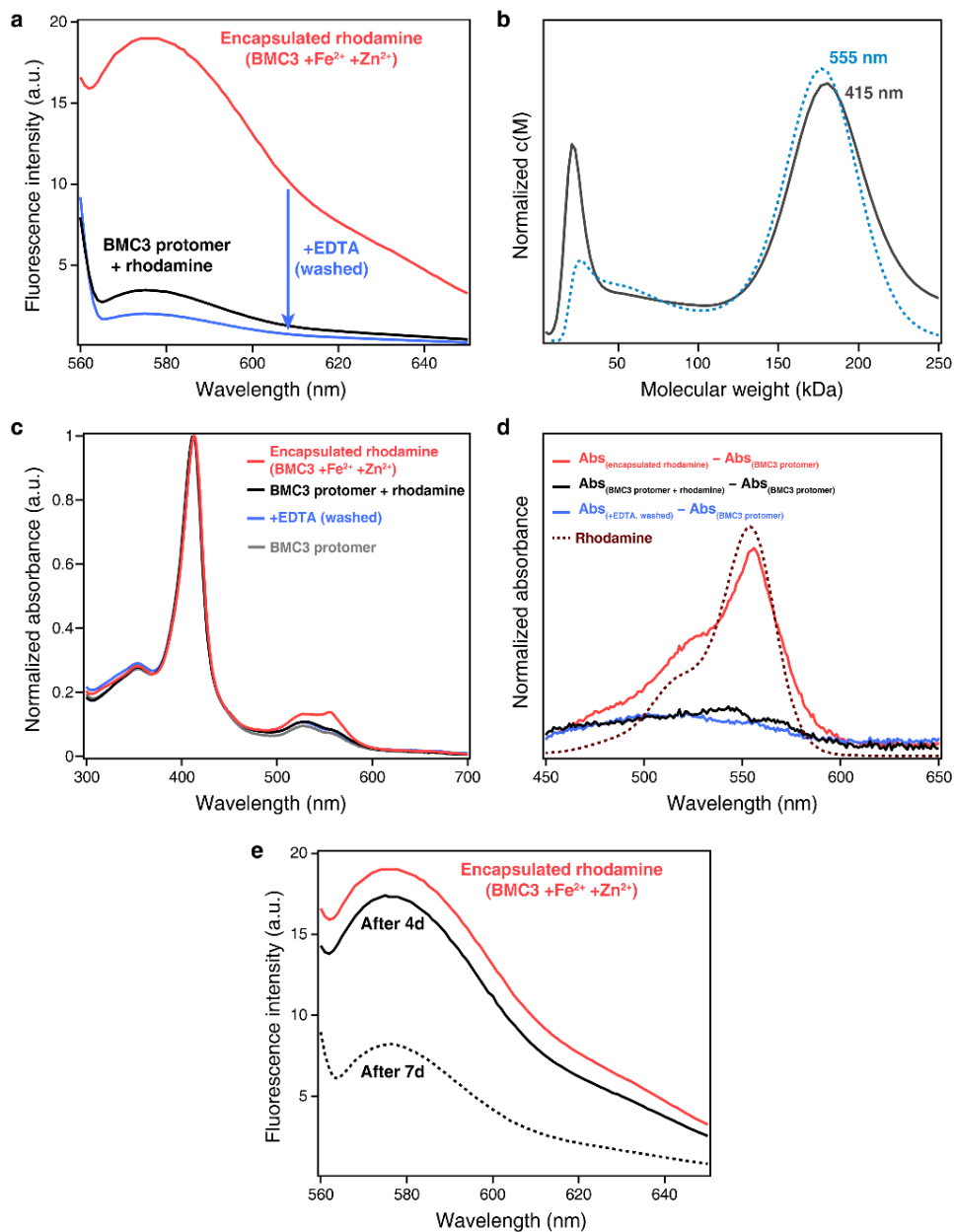


Figure 4.11 Encapsulation of rhodamine inside BMC3 cages. **(a)** Fluorescence characterization of BMC3 samples incubated with rhodamine. Cages encapsulating rhodamine were treated with EDTA and washed prior to measuring fluorescence intensity. **(b)** AUC profiles of cages encapsulating rhodamine monitored at the heme Soret absorption maximum ($\lambda_{\max} = 415$ nm) and rhodamine absorption maximum ($\lambda_{\max} = 555$ nm). **(c)** UV-vis characterization of BMC3 samples incubated with rhodamine. **(d)** Difference spectra of BMC3 samples and BMC3 protomer shown in c. Free rhodamine dissolved in buffer is shown as dark-red dashes. **(e)** Repeated fluorescence characterization of a solution containing BMC3 cages encapsulating rhodamine over several days. The sample was washed 3x prior to each fluorescence measurement.

4.3.3 Formation of a hexameric cage

Unexpectedly, AUC measurements indicated that the other second-generation variant, BMC4, self-assembled as a hexamer upon Fe and Zn coordination, with yields exceeding 70% (**Figure 4.12a**). The 1.50-Å resolution crystal structure of the BMC4 assembly revealed a D_3 symmetric, cage-like architecture with a composition of [2 Fe:9 Zn:6 protomers], outer dimensions of 75 Å × 50 Å and a cavity volume of more than 7,800 Å³ (Extended Data Fig. 4). The overall shape is a trigonal bipyramid (**Figure 4.12b**), which is the smallest polyhedral architecture with a sizeable interior cavity that can be constructed from an asymmetric building block. The apical vertex of each pyramidal half is formed by a Fe³⁺:(Cys82–HA)₃ motif shared by three protomers (**Figure 4.12b**). These C_3 symmetric vertices are further reinforced by Zn²⁺ ions that link pairs of protomers through Asp21, Glu25 and His77 coordination (**Figure 4.12c**). The pyramids are joined by three equatorial, C_2 symmetric vertices mediated by Zn centers coordinated to Glu5 and His8 (**Figure 4.12b, c**). A comparison to the BMC2 and BMC3 cages indicates that this unexpected coordination motif requires an approximately 9-Å slip of each protomer along the C_2 symmetric interfaces (**Figure 4.12d**). The shift markedly reduces the C_2 symmetric contact area between protomers, effectively transforming the edges in the tetrahedral BMC2 and BMC3 cages to vertices in the trigonal bipyramidal BMC4 cages. BMC4 cages exhibited similar thermal stability to BMC3 cages, with both species disassembling at below 70 °C. The thermal robustness of the BMC3 and BMC4 cages appear to be limited, at least in part, by the relative instability of the individual protomers (**Figure 4.9d**).

The large structural transformation is accompanied by a reduction in the apical angle formed at the Fe³⁺:(Cys82–HA)₃-mediated vertices from 101° in the BMC2 and BMC3 cages to 81° in the BMC4 cage (**Figure 4.12e**). This observation highlights the conformational adaptability

of the $\text{Fe}^{3+}:(\text{Cys82-HA})_3$ coordination motif, enabling it to accommodate different polyhedral geometries. Such behavior is reminiscent of the interfacial flexibility in some icosahedral virus capsids in which the same protomer can form hexamers on capsid faces and pentamers on capsid vertices⁷. It is worth noting that BMC4 contains all of the Zn-coordinating residues on helix 1 to form the C_2 symmetric interfaces observed in the dodecameric BMC3 cage, indicating that the self-assembly process selects an alternative interfacial arrangement of lower free energy, enabled by the reversibility of metal coordination interactions. In terms of protein design, a caveat of interfacial flexibility is that it may lead to nonspecific or unintended self-assembly products, although it can also allow error correction during self-assembly and increase tolerance to design imperfections.

4.4 Conclusions

The self-assembly and function of biomolecular systems are predicated upon their specificity, stability and adaptiveness, which, in turn, are enabled by extensive networks of non-covalent interactions. Here, we have shown that fundamental concepts in inorganic coordination chemistry can be applied to achieve all of these attributes in protein self-assembly and, specifically, to construct complex polyhedral protein architectures from a simple, asymmetric building block. Despite their minimal design footprints, these cage-like architectures are distinguished by their structural compactness and responsiveness — hallmarks of evolved systems such as viral capsids. Key to our construction strategy was the reimagination of a biological coordination motif, hydroxamic acid, within a new structural context: as a new amino acid side chain with the ability to chelate hard metal ions. This example expands the growing lexicon of post-translational modifications that broaden the chemical scope of proteins.

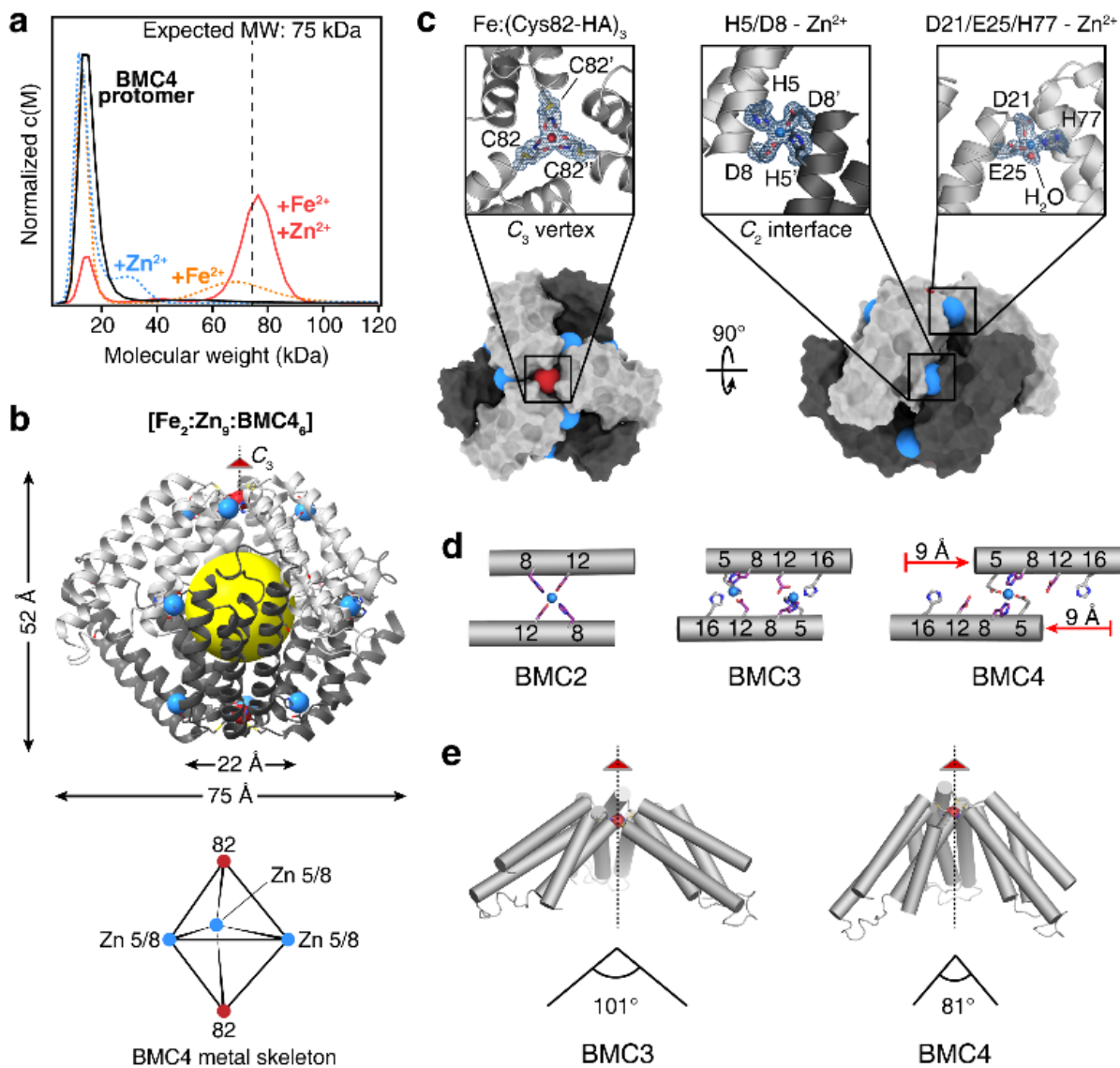


Figure 4.12 Characterization of BMC4 cages. **(a)** AUC characterization of BMC4 self-assembly. **(b)** Crystal structure of the BMC4 cage. Fe and Zn ions are represented as red and blue spheres, respectively. The central cavity is highlighted by a yellow sphere. The structural skeleton formed by Fe and Zn ions is shown below the structure. **(c)** Surface representations of the BMC4 cage, with metal ions shown as colored spheres. Atomic details of each metal coordination site are shown in insets, with the $mF_o - DF_c$ electron density omit maps (blue mesh) contoured at 3σ . **(d)** Comparison of the C_2 symmetric protein interfaces in different BMC constructs. The residues 8 and 12, common to all constructs, are colored in purple. The slippage of the 2-fold helix interface to accommodate the hexameric architecture of BMC4 is indicated with red arrows. **(e)** Comparison of the apical angle formed by the Fe:(Cys82-HA)₃-mediated vertices in BMC3 and BMC4 cages.

4.5 Materials and Methods

4.5.1 Synthesis of the IHA ligand

O-tritylhydroxylamine was synthesized as previously described³². Chloroacetyl chloride (0.58 mL, 7.3 mmol) was dissolved in 2 mL of CH₂Cl₂ and added dropwise to a suspension of *O*-tritylhydroxylamine (2.0 g, 7.3 mmol) and *N,N*-Diisopropylethylamine (2.5 mL, 14.5 mmol) in 15 mL of CH₂Cl₂ at 0 °C. The reaction mixture was gradually warmed to room temperature and stirred at room temperature for an hour. An additional 15 mL of CH₂Cl₂ was added and the reaction was extracted with H₂O (3 × 30 mL). The CH₂Cl₂ solution was collected and evaporated to dryness. A solution containing 15 mL of CH₂Cl₂ with 10% (v/v) trifluoroacetic acid was added and the solution was stirred for 30 min. The crude product was purified by silica gel chromatography using a gradient of 0–100% ethyl acetate in hexanes as the eluent (See **Appendix 2** for NMR spectra of pure product). The product was visualized using a FeCl₃ stain. Yield = 55%. Measured molecular weight (m/z): 108.37 (M – H⁺); Calculated: 107.99 (M – H⁺).

¹H NMR: (300 MHz, DMSO-d₆) δ 10.88 (s, 1H), δ 9.15 (s, 1H), δ 3.93 (s, 2H)

¹³C NMR: (500 MHz, DMSO-d₆) δ 162.88, δ 40.45

2-chloro-*N*-hydroxyacetamide (400 mg, 3.7 mmol) and NaI (2.7 g, 18.3 mmol) were refluxed in 30 mL of acetone for 1 h. The reaction mixture was purified by silica gel chromatography with 100% ethyl acetate as the eluent and dried *in vacuo* (See **Appendix 2** for NMR spectra of pure product). Yield >90%. Measured molecular weight (m/z): 223.85 (M + Na⁺); Calculated: 223.95 (M + Na⁺).

¹H NMR: (300 MHz, DMSO-d₆) δ 10.81 (s, 1H), δ 9.09 (s, 1H), δ 3.51 (s, 2H)

¹³C NMR: (500 MHz, DMSO-d₆) δ 164.83, δ –2.01

4.5.2 Protein expression and purification

All constructs (**Table 4.1**) were derived from the parent pET-20b(+) plasmid containing the CFMC1 gene via site-directed mutagenesis as previously described^{28,33,34}. The appropriate plasmids were transformed into BL21(DE3) Escherichia coli cells (New England Biolabs) housing a CCM (cytochrome C maturation) cassette containing a chloramphenicol-resistance marker and expressed as previously described³⁵ with minor adjustments. Multiple 2.8-l flasks containing 1.5 l of LB medium were shaken at 200 rpm for 12 h at 37 °C and then at 100 rpm for an additional period of around 7 h. Cells were collected by centrifugation (5,000 rpm for 10 min at 4 °C), resuspended in a buffered solution containing 5 mM sodium acetate (NaOAc) (pH 5.0) and 2 mM dithiothreitol (DTT) and lysed via sonication. The pH of the crude lysate was first raised to 10 using NaOH to precipitate cellular contaminants, then reduced to pH 4.5. After centrifugation (12,000 rpm for 20 min at 4 °C), the clarified supernatant was decanted and diluted 15-fold with additional buffer. This solution was applied to a CM Sepharose gravity column (GE Healthcare) pre-equilibrated with the aforementioned buffer and subjected to multiple buffer washes before elution using a stepwise-gradient of NaCl (0–0.5 M). Peak elution fractions were combined and concentrated using a 400-ml Amicon Stirred Cell (Millipore) and buffer-exchanged by overnight dialysis against a buffered solution containing 10 mM phosphate (pH 8.0) at 4 °C. Next, the protein was purified via a DuoFlow workstation station fitted with a MacroPrep High Q-cartridge column (BioRad) and eluted using a linear gradient over 0–0.5 M NaCl. Fractions that exhibited an RZ ratio (A_{421}/A_{280}) > 4.4 were pooled, treated with 2 mM EDTA for 1 h, concentrated, and buffer-exchanged into 20 mM tris(hydroxymethyl) aminomethane (Tris) (pH 7.5) pretreated with Chelex 100 resin (BioRad), via desalting column (Econo-Pac 10DG pre-packed columns, BioRad). Demetallated and purified proteins were concentrated to around 2 mM and stored at 4 °C.

4.5.3 Protein labelling and post-labelling purification

Purified protein solutions were treated with a 100-fold excess of DTT and placed in an anaerobic Coy chamber for approximately 2 h for slow degassing to remove dissolved oxygen. The fully reduced protein solution was buffer-exchanged into 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5) via desalting column to remove DTT, and the concentration of the resulting protein solution was determined spectroscopically (Agilent 8452 spectrophotometer) using the $\epsilon_{421}(\text{red}) = 162,000 \text{ M}^{-1}\text{cm}^{-1}$ (ref. 34). Solid iodohydroxamic acid (IHA) was dissolved in 100 μl degassed DMF to generate solutions containing a 15-fold excess IHA per protein monomer, which were then added to protein aliquots and incubated overnight. The HA-functionalized variants were removed from the Coy chamber and separated from unreacted or partially reacted protein via FPLC using a Q-column equilibrated with 10 mM N-cyclohexyl-2-aminoethanesulfonic acid (CHES) (pH 9.3) and 2 mM DTT and eluted using a linear gradient over 0–0.5 M NaCl. Protein functionalization was verified using electrospray ionization mass spectrometry (ESI-MS; Extended Data Fig. 1) and the resulting protein solutions were buffer-exchanged into demetallated 20 mM Tris (pH 7.5) via desalting column, concentrated to around 2 mM and stored at 4 °C for further use.

4.5.4 Redesign of CFMC1 interfaces

To render the CFMC1 protomer competent for the bimetallic design strategy, we first performed the following mutations to remove potential competitive interactions: C67E, H59S and H73N. A negative design strategy was then used to disrupt a noncovalent dimerization interface found in CFMC1, leading to the mutations A34Q and A38Q. We further identified the dearth of protein–protein interactions within the core and periphery of the three-fold axis engulfing the 82 position as a likely contributor to poor cage assembly and crystallization in general. Accordingly,

as a means to facilitate cage formation, we adopted Rosetta-prescribed mutations at the following positions: A24T, Q25(T/E), N80K and E81Q.

Table 4.8 Crystallization conditions for the different variants

Variant	Metal content	[BMC]^a	[Fe]	[Zn]	Mother liquor
BMC1	Fe, Zn	2.1 mM	1.05 mM ^b	2 mM	22.5% PEG 400, 0.1 M HEPES pH 7.5, 0.2 M NaCl
BMC2	Fe, Zn	2.2 mM	1.65 mM ^c	2 mM	30% PEG 400, 0.1 M Tris pH 8.5, 0.2 M NaCl
BMC3	Fe, Zn	2.2 mM	1.65 mM ^c	3 mM	30% PEG 400, 0.1 M HEPES pH 7.5, 0.2 M (NH ₄) ₂ SO ₄
BMC4	Fe, Zn	2.2 mM	1.05 mM ^b	3 mM	30% PEG 400, 0.1 M HEPES pH 7.5, 0.2 M MgCl ₂

^a[protein] relates to the concentration of the monomeric functionalized cytochrome subunit.

^bProtein solutions were premixed with metal ion salts 1 hour prior to crystallization. ^cMetal ion salts were mixed briefly prior to crystallization.

4.5.5 Crystallography

Screening and crystallization of all BMC variants were conducted via sitting drop vapor diffusion. Briefly, solutions containing 2.1-2.2 mM BMC protomer were mixed with mother liquor (1 μ l + 1 μ l) and equilibrated against 200- μ l reservoir volumes. **Table 4.8** details the experimental conditions for crystal growth. Protein solutions of BMC1 and BMC4 were first incubated with FeSO₄ for 1 hour prior to mixing with ZnCl₂. Solutions of BMC2 and BMC3 were mixed with FeSO₄ and ZnCl₂ stock solutions and were immediately combined with the mother liquor (to prevent rapid aggregation of the proteins functionalized with two HA units). Crystals for all mutants typically appeared within several hours and were harvested within a week of maturation. Crystals were cryoprotected by submersion into perfluoropolyether cryo oil (Hampton Research) for a few seconds and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K at either the Advanced Light Source (ALS) beamline BL 8.3.1 (using 1.12-Å radiation for BMC3 and 1.33-Å radiation for BMC4) or at the Stanford Synchrotron Radiation Lightsource (SSRL) beamlines 9-2 (using 0.98-Å radiation for BMC2) and 12-2 (using 0.98-Å radiation for BMC1).

Data integration was performed using the XDS Program Package, truncated at $CC_{1/2} > 0.5$ ³⁶. Datasets of the same structure recorded at different wavelengths were scaled to the highest resolution dataset with XSCALE^{37,38}. Phaser-MR³⁹ was employed to carry out molecular replacement with search models based on the CMFC1 monomer (PDB ID: 3M4B) containing the expected side chain mutations (generated in Pymol⁴⁰) but lacking HA. Rigid-body and structure refinement was performed using multiple rounds of Phenix.refine³⁹, interspersed with manual model rebuilding and metal/ligand placement with COOT⁴¹. Restraint files for the Cys-hydroxamic acid conjugates were generated using phenix.eLBOW to maintain the distances Cys-SG-HA-C1 ($1.816 \text{ \AA} \pm 0.02 \text{ \AA}$) and angles Cys-CB-Cys-SG-HA-C1 as well as Cys-SG-HA-C1-HA-C2 (both $109^\circ \pm 3^\circ$) during refinement. Where necessary, the metal binding geometry of the hydroxamic acids was restrained to the distances Fe-HA-O1 ($1.98 \text{ \AA} \pm 0.05 \text{ \AA}$) and Fe-HA-O2 ($2.057 \text{ \AA} \pm 0.05 \text{ \AA}$) as well as through a planarity constraint for the atoms Fe-HA-O1-HA-O2-HA-C1 following data from a high-resolution structure of Fe(III)-tris-benzhydroxamate trihydrate⁴². Simulated annealing omit maps (metal atoms and sidechain ligands) were generated for each metal binding site and model accuracy was assessed critically against these omit maps. Electron density maps were generated using Phenix and all molecular graphics images were produced with either Pymol or the UCSF ChimeraX package from the Computer Graphics Laboratory, University of California, San Francisco⁴³.

4.5.6 Crystallographic metal content analysis

Metal ions, with their relatively high-energy inner electrons, can absorb and resonate with soft X-rays; this leads, among other effects, to differences in the intensity of otherwise centrosymmetric Bragg diffraction peaks used for X-ray crystallography. Density maps calculated from these differences are routinely used to locate and identify metal ions in protein crystals. The

magnitude of this anomalous X-ray diffraction varies with the X-ray energy, with stark differences around the energies of the K- and L-shell electrons of the respective elements allowing one to discern between elements at a position in question, if diffraction datasets are measured at the appropriate wavelengths. For a visual analysis of the bound metals, the scaled datasets of different wavelengths were used separately as an input for a single phenix.refine run, each with the final model of the highest resolution dataset. Importantly, only the B-factor or occupancy were allowed to change during refinement, resulting in anomalous difference density maps for each wavelength. Using these maps, isomorphous difference maps from data at wavelengths above and below the respective element K-edges were generated (if applicable) with Phenix and were inspected manually (Extended Data Fig. 5). To gain a more quantitative understanding of the identity of the bound metals for each site, the anomalous difference signal of each dataset was used to generate CCP4 format maps with phenix.mtz2map. The generated maps were used subsequently as inputs to calculate the mean signal in a sphere of 1 Å radius centered on each metal atom with the program MAPMAN (Uppsala Software Factory). For each pair of datasets above and below a metal-absorption edge, the ratio of the anomalous signal above and below the edge for every metal atom was tabulated. The experimental ratio was compared to the theoretical ratio for both Fe and Zn (Extended Data Fig. 5) according to <http://skuld.bmsc.washington.edu/scatter>, as calculated using the Cromer and Liberman approximation. Theoretical ratios were also calculated for hypothetical mixed occupancy Fe/Zn metal sites and compared to experimentally observed values (**Tables 4.3-6**).

4.5.7 Protein cage sample preparation

Self-assembled cages

All samples were prepared in a low-O₂ atmosphere (Coy glovebox) to minimize undesired oxidation of Fe²⁺ ions before self-assembly. Protein solutions containing 20 μM BMC3 or 100 μM BMC4 in 20 mM Tris (pH 8.5) were incubated with either 20 μM FeSO₄ and 60 μM ZnCl₂ for BMC3 or 50 μM FeSO₄ and 200 μM ZnCl₂ for BMC4 for 2–3 h to yield the metallated cages. We note that the addition of FeSO₄ was followed by a small but observable change in the color of the solution from red to pink, attributed to a shift of the heme Soret band to longer wavelengths, which suggested reduction of the heme by the ferrous ions and generation of ferric ions in close proximity to HA group(s). The final BMC3 solutions were then concentrated sevenfold before overnight incubation to improve the total cage yield. After self-assembly, the resulting solutions were diluted back to their original concentrations with the self-assembly buffer before characterization.

Dissolved crystals

Fe:Zn:BMC1 and Fe:Zn:BMC2 crystals were dissolved using buffer containing 100 mM HEPES (pH 7.5), 200 mM MgCl₂ and 800 μM ZnCl₂. Mature crystals were removed from their pedestal droplet, briefly submerged in fresh buffer to remove uncrystallized protein and surface-bound precipitates, and transferred into a new sitting drop crystallization well containing 8 μl buffer solution. The crystals were physically crushed with a small metal scalpel and vigorously pipetted until a large portion of the crystals dissolved. Undissolved crystals were removed by centrifugation (10,000 rpm for 5 min at 25 °C), yielding a light-red supernatant and dark-red precipitate.

4.5.8 Negative-stain transmission electron microscopy

A 4- μ l droplet of BMC cages (either self-assembled or from dissolved crystals) was deposited onto formvar/carbon-coated Cu grids (Ted Pella) (pretreated by negative-mode glow discharge up to 15 min beforehand) and allowed to bind for 5 min. The grids were then washed with 50 μ l MilliQ water, blotted using Whatman filter paper and stained using 2% uranyl acetate solution in water and blotted again. Grids were imaged using a FEI Sphera transmission electron microscope operating at 200 keV, equipped with a LaB₆ filament and a Gatan 4K CCD camera. Micrographs were collected using objective-lens underfocus settings ranging from 250 nm to 2 μ m and analyzed using Fiji (<http://fiji.sc/Fiji>).

4.5.9 Oligomerization state determination using AUC

Sedimentation velocity measurements were performed at 41,000 rpm and 25 °C using an XL-1 analytical ultracentrifuge (Beckman Coulter) equipped with an AN-60 Ti rotor. Data processing was performed using Sedfit⁴⁴ with the following parameters as calculated using SEDNTERP: viscosity: 0.01000 poise, density: 0.9988 g/ml (self-assembled samples) or viscosity: 0.0113191 poise, density: 1.0196 g/ml (dissolved crystals), and a partial specific volume of 0.7313 ml/g for all samples. All reported results correspond to a confidence level of 0.95.

4.5.10 Preparation of samples involving crystal dissolution

Dissolved crystal samples (BMC1 and BMC2), prepared as described above at ambient conditions, were diluted to 350 μ l with 10 mM HEPES (pH 7.5), 200 mM MgCl₂ and 800 μ M ZnCl₂. The solution was clarified via brief centrifugation in order to remove crystal debris and the supernatant was placed inside the cells.

4.5.11 Calculation of BMC void volumes

Structures of complete cage assemblies for BMC2, BMC3 and BMC4 were generated via the application of crystallographic symmetry operations to the fully refined asymmetric unit of each construct. These coordinates were re-centered at the origin and stripped of waters, hydrogens, alternative conformations and crystallization reagents (PEG-400). Volumetric maps and volumes for the internal cavity of each cage were calculated using VOIDOO⁴⁵, and are reported as the solvent-accessible volume for a 1.4 Å rolling probe on a 0.25 Å grid spacing for all constructs. The cavity volumes using these parameters were determined to be approximately 32,700 Å³ (BMC2), 32,700 Å³ (BMC3) and 7,900 Å³ (BMC4).

4.5.12 Solution self-assembly, disassembly and thermal stability of BMC3 and BMC4

Assembled samples were prepared as described above and placed inside the AUC measurement cells anaerobically (20 µM BMC3 and 100 µM BMC4). Disassembly of the cages via metal-ion removal was performed by treating the protein cages with 2 mM EDTA for 1 h. Redox-controlled disassembly of the protein commenced by the addition of either 5 mM sodium dithionite or 5 mM sodium ascorbate to the cage solution anaerobically and subsequent incubation of the samples at around 22 °C for 16 h. Samples were then loaded into the AUC measurement cell.

For thermal stability measurements, samples were placed in a thermoregulated chamber preequilibrated at the appropriate temperature for 2 h, and subsequently removed from the chamber and equilibrated at room temperature for 30 min before AUC analysis. Circular dichroism (CD) spectra were measured using an Aviv 215 spectrometer. CD measurements were performed using 10 µM protein in a buffered solution containing 20 mM Tris (pH 8.5). Thermal melts were measured at 222 nm at a 1 nm slit width, scanning at 1-nm intervals with a 1-s integration time.

Measurements were taken from 25 °C to 85 °C at 2-degree intervals with a 2 min equilibration at each temperature. Unfolding data were fit to a two-state model with van't Hoff's enthalpy using the CalFitter web server⁴⁶.

4.5.13 Cryo-EM sample preparation

Self-assembled BMC3 cages were removed from the anaerobic Coy chamber immediately before grid preparation. A 3.5- μ l aliquot of self-assembled BMC3 cages was dropped onto holey carbon grids (Electron Microscopy Sciences, Quantifoil R1.2/1.3 holey carbon on 300 mesh copper) that had been freshly glow-discharged for 30 s. The initial application of the sample was side blotted manually with Whatman No. 1 filter paper immediately followed by a secondary application of a 3.5- μ l aliquot, blotted for 3.5 s and plunge-frozen in liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV (FEI).

4.5.14 Cryo-EM data acquisition and image processing

Samples were imaged on a Titan Krios G3 transmission electron microscope (FEI) operating at 300 kV equipped with a K2 Summit direct electron detector (Gatan) and a GIF Quantum energy filter. The slit-width of the energy filter was set to 10 eV. Movies were collected at a magnification of 165,000 \times in EFTEM mode giving a physical pixel size of 0.84 Å/pixel. In total, 4,672 movie stacks (50 frames/movie) were collected using a 10 s exposure at a dose rate of 1.2 e⁻ /Å² per frame for a total electron dose of 60 e⁻ /Å² per movie. Objective-lens underfocus settings varied between 0.6 μ m and 1.6 μ m. Data collection was performed using software EPU (FEI). All image processing was performed in the Relion-3.0 pipeline⁴⁷. Motion correction and dose weighting were performed using MotionCor2⁴⁸, and defocus values were estimated with Gctf⁴⁹ using a pixel size of 0.8 Å/pixel. A total of 3,513 movie stacks were selected following motion correction and CTF estimation, and 805,156 particles were auto-picked using RELION-

3.0. Particle images were extracted and binned by 2 (1.6 Å/pixel, 100 pixel box size) and subjected to two-dimensional (2D) classification. A total of 444,247 particles were selected corresponding to good 2D class averages and subjected to three-dimensional (3D) classification imposing T symmetry and using an initial model generated from a subset of the particles. A total of 129,653 particles were chosen from a 3D class showing strong secondary-structural elements and subjected to 3D auto-refinement with T symmetry. The particles were re-centered and re-extracted to their original pixel size of 0.8 Å/pixel. These particles were subjected to 3D auto-refinement with T symmetry and the yield map was then postprocessed towards 2.6 Å resolution based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion. The pixel size of the map was manually adjusted using Relion image handler to match the physical pixel size of the images. Local resolution was calculated in Relion 3.0 using ResMap⁵⁰.

4.5.15 Model building and refinement

The BMC3 crystal structure (PDB ID: 6OT7) stripped of hydrogens and waters was used as an initial model and manually docked into the cryo-EM density using UCSF Chimera⁵¹. The structural model was subject to real space refinement in Phenix against the cryo-EM map with geometry restraints for the Fe-binding sites and molecular coordinates for the Cys–HA ligand. The atomic model was manually improved using Coot. Tightly bound waters were identified based on clear density in the EM density map. Whereas the structural flexibility of the hydroxamate sites manifested in poor electron density, the twofold interface was much more rigid and unambiguous density was observed for Zn-binding. A tryptophan at the 66 position, which had shown high-temperature factors in the BMC3 crystal structure, was identified in multiple conformations in the EM density map. The final model was subjected to real space refinement using Phenix³⁹ and

evaluated using MolProbity⁵². All molecular graphics images were rendered in PyMol or UCSF ChimeraX.

4.5.16 Encapsulation of rhodamine in BMC3 cages

BMC3 cages were self-assembled in a low-O₂ atmosphere in the presence of rhodamine for the passive encapsulation of the dye. Solutions containing 20 μ M BMC3 were incubated with 20 μ M FeSO₄, 60 μ M ZnCl₂ and 2 mM rhodamine. A control sample was prepared in the absence of added metal ions (20 μ M BMC3 incubated with 2 mM rhodamine). Samples were incubated for 2–3 h and concentrated sevenfold before overnight incubation. Protein solutions were buffer exchanged on a PD-10 desalting column using a buffer containing 20 mM Tris (pH 8.5) (with 5 μ M FeSO₄ and 10 μ M ZnCl₂ supplemented for solutions already containing metal ions) to separate unassociated dye from protein. Cage solutions were split in two: one half was treated with 1 mM EDTA and incubated for 2 h before washing. All protein solutions were additionally washed three times using a centrifugal filter to completely remove any remaining free rhodamine. Fluorescence measurements were performed using 6 μ M protein solutions after the previously mentioned wash steps. For each sample, an excitation wavelength of 555 nm with a 2 nm slit width was used and emission was measured between 560 and 650 nm with a 2 nm slit width and 0.2 s integration time. For the time-course experiments, cages encapsulating rhodamine were washed three times after 4 days and after 7 days and diluted to 6 μ M before fluorescence measurements. AUC measurements were performed at the λ_{max} of the cytochrome (415 nm) and at the λ_{max} of rhodamine (555 nm) to assess whether there was a sufficiently large rhodamine signal associated with BMC3 cages. Ultraviolet–visible light (UV-vis) absorbance measurements were performed on each solution to measure the protein and rhodamine concentrations. Difference spectra were taken between each rhodamine-incubated sample and BMC3 protomer to eliminate any background signal.

4.5.17 Statistics and reproducibility

All reported samples represent technical replicates. The ns-TEM micrograph of BMC2 cages after 3D crystal dissolution (Fig. 2a) is representative of experiments repeated independently four times. AUC experiments for BMC2 (Fig. 2b) were performed in duplicate. Self-assembly of BMC3 cages and subsequent AUC characterization (Fig. 3a) were performed the following number of times: BMC3 protomer (n=2), +Fe²⁺ (n=4), +Zn²⁺ (n=4), +Fe²⁺, +Zn²⁺ (n=6). Self-assembly of BMC4 cages and subsequent AUC characterization (Fig. 4a) was performed the following number of times: BMC4 protomer (n=2), +Fe²⁺ (n=1), +Zn²⁺ (n=1), +Fe²⁺, +Zn²⁺ (n=5). Mass spectra (Extended Data Fig. 1c-f) were collected in duplicate for native and HA-labeled proteins; AUC experiments were performed in duplicate. TEM characterization of BMC constructs (Extended Data Fig. 3) were performed the following number of times: dissolved BMC1 crystals (n=1), dissolved BMC2 crystals (n=4), BMC2 +EDTA (n=2), self-assembled BMC3 cages (n=5), BMC3 +EDTA (n=4). AUC experiments following the incubation of BMC3 with first-row transition metals (Extended Data Fig. 6a) were performed in duplicate. Self-assembly of BMC3 in the presence of Fe(acetylacetonate)₃ (Extended Data Fig. 6b) was performed in duplicate. BMC3 cage disassembly in the presence of EDTA (Extended Data Fig. 6c) was performed in triplicate. AUC characterization of BMC variants after equilibration at different temperatures (Extended Data Fig. 6d) was performed the following number of times: BMC3 at 50 °C (n=2), BMC3 at 70 °C (n=2), BMC4 at 50 °C (n=3), BMC4 at 70 °C (n=3), BMC4 at 90 °C (n=3). Thermal unfolding of BMC variants as measured by CD spectroscopy (Extended Data Fig. 6d) was performed in duplicate. Treatment of BMC3 cages with chemical reductants (Extended Data Fig. 6e) was performed in duplicate. Cryo-EM characterization of BMC3 cages was performed after collecting 4,672 movie stacks. Shown in Extended Data Fig. 7a is a representative micrograph and 3

representative 2D class averages. Fluorescence characterization of BMC3 samples incubated with rhodamine were performed (Extended Data Fig. 8a) in triplicate. AUC characterization of BMC3 cages encapsulating rhodamine (Extended Data Fig. 8b) was performed in duplicate. UV-vis characterization of BMC3 samples incubated with rhodamine (Extended Data Fig. 8c, d) was performed in triplicate. Repeated fluorescence characterization of a solution containing BMC3 cages encapsulating rhodamine (Extended Data Fig. 8e) was performed in duplicate.

4.6 Acknowledgments

This work was supported by the US Department of Energy (Division of Materials Sciences, Office of Basic Energy Sciences; DE-SC0003844; for the design strategy, EM imaging and analysis, and biochemical analysis) and by the National Science Foundation (Division of Materials Research; DMR-1602537; for crystallographic analysis). E.G. acknowledges support by an EMBO Long-Term Postdoctoral Fellowship (ALTF 1336-2015). J.E. acknowledges support by a DFG Research Fellowship (DFG 393131496). R.H.S. was supported by the National Institute of Health Chemical Biology Interfaces Training Grant UC San Diego (T32GM112584). We acknowledge the use of the UCSD Cryo-EM Facility, which is supported by NIH grants to T.S.B. and a gift from the Agouron Institute to UCSD. Crystallographic data were collected either at Stanford Synchrotron Radiation Lightsource (SSRL) or at the Lawrence Berkeley Natural Laboratory on behalf of the Department of Energy.

Chapter 4 is reproduced, in part, with permission, from Golub, E.; Subramanian, R. H.; Esselborn, J.; Alberstein, R. G.; Bailey, J. B.; Chiong, J. A.; Yan, X.; Booth, T.; Baker, T. S.; Tezcan, F. A. Constructing Protein Polyhedra via Orthogonal Chemical Interactions, *Nature* **2020**, *578*, 172-176.

4.7 References

- (1) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551-575.
- (2) 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. USA* **2001**, *98* (5), 2217-2221.
- (3) Bai, Y.; Luo, Q.; Liu, J. Protein self-assembly via supramolecular strategies. *Chem. Soc. Rev.* **2016**, *45* (10), 2756-2767.
- (4) Hamley, I. W. Protein Assemblies: Nature-Inspired and Designed Nanostructures. *Biomacromolecules* **2019**.
- (5) Churchfield, L. A.; Tezcan, F. A. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **2019**, *52* (2), 345-355.
- (6) Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **2017**, *46* (1), 23-42.
- (7) Johnson, J. E.; Speir, J. A. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol.* **1997**, *269* (5), 665-675.
- (8) Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M. A.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G. et al. Solving the Structure of Human H-Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature* **1991**, *349* (6309), 541-544.
- (9) Lai, Y.-T.; Cascio, D.; Yeates, T. O. Structure of a 16-nm Cage Designed by Using Protein Oligomers. *Science* **2012**, *336* (6085), 1129.
- (10) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **2016**, *353* (6297), 389-394.
- (11) 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. USA* **2016**, *113* (31), 8681-8686.
- (12) 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. et al. Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* **2013**, *340* (6132), 595-599.
- (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. et al. An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **2019**, *569* (7756), 438-442.

- (14) Pluth, M. D.; Bergman, R. G.; Raymond, K. N. Acid catalysis in basic solution: A supramolecular host promotes orthoformate hydrolysis. *Science* **2007**, *316* (5821), 85-88.
- (15) Chakrabarty, R.; Mukherjee, P. S.; Stang, P. J. Supramolecular Coordination: Self-Assembly of Finite Two- and Three-Dimensional Ensembles. *Chem. Rev.* **2011**, *111* (11), 6810-6918.
- (16) Yoshizawa, M.; Klosterman, J. K.; Fujita, M. Functional Molecular Flasks: New Properties and Reactions within Discrete, Self-Assembled Hosts. *Angew. Chem. Int. Ed.* **2009**, *48* (19), 3418-3438.
- (17) Mal, P.; Breiner, B.; Rissanen, K.; Nitschke, J. R. White Phosphorus Is Air-Stable Within a Self-Assembled Tetrahedral Capsule. *Science* **2009**, *324* (5935), 1697-1699.
- (18) Liu, Y.; Hu, C.; Comotti, A.; Ward, M. D. Supramolecular Archimedean Cages Assembled with 72 Hydrogen Bonds. *Science* **2011**, *333* (6041), 436-440.
- (19) Mateu, M. G. Assembly, stability and dynamics of virus capsids. *Arch. Biochem. Biophys.* **2013**, *531* (1), 65-79.
- (20) Sun, X.; Johnson, D. W.; Caulder, D. L.; Raymond, K. N.; Wong, E. H. Rational Design and Assembly of $M_2M'_3L_6$ Supramolecular Clusters with C_3h Symmetry by Exploiting Incommensurate Symmetry Numbers. *J. Am. Chem. Soc.* **2001**, *123* (12), 2752-2763.
- (21) Smulders, M. M. J.; Jiménez, A.; Nitschke, J. R. Integrative Self-Sorting Synthesis of a $Fe_8Pt_6L_{24}$ Cubic Cage. *Angew. Chem. Int. Ed.* **2012**, *51* (27), 6681-6685.
- (22) Brodin, J. D.; Ambroggio, X. I.; Tang, C. Y.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **2012**, *4* (5), 375-382.
- (23) 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*, 369.
- (24) Radford, R. J.; Nguyen, P. C.; Tezcan, F. A. Modular and Versatile Hybrid Coordination Motifs on Alpha-Helical Protein Surfaces. *Inorg. Chem.* **2010**, *2010* (49), 7106-7115.
- (25) Pearson, R. G. Hard and Soft Acids and Bases. *J. Am. Chem. Soc.* **1963**, *85* (22), 3533-3539.
- (26) Wong, G. B.; Kappel, M. J.; Raymond, K. N.; Matzanke, B.; Winkelmann, G. Coordination chemistry of microbial iron transport compounds. 24. Characterization of coprogen and ferricrocin, two ferric hydroxamate siderophores. *J. Am. Chem. Soc.* **1983**, *105* (4), 810-815.

- (27) Crumbliss, A. L. Iron bioavailability and the coordination chemistry of hydroxamic acids. *Coord. Chem. Rev.* **1990**, *105*, 155-179.
- (28) Ni, T. W.; Tezcan, F. A. Structural Characterization of a Microperoxidase Inside a Metal-Directed Protein Cage. *Angew. Chem. Int. Ed.* **2010**, *49* (39), 7014-7018.
- (29) Failes, T. W.; Hambley, T. W. Crystal Structures of Tris(hydroxamato) Complexes of Iron(III). *Aust. J. Chem.* **2001**, *53* (10), 879-881.
- (30) Mayhew, S. G. The Redox Potential of Dithionite and SO_2^- from Equilibrium Reactions with Flavodoxins, Methyl Viologen and Hydrogen plus Hydrogenase. *Eur. J. Biochem.* **1978**, *85* (2), 535-547.
- (31) Borsook, H.; Keighley, G. Oxidation-Reduction Potential of Ascorbic Acid (Vitamin C). *Proc. Natl. Acad. Sci. USA* **1933**, *19* (9), 875-878.
- (32) Michalak, K.; Wicha, J.; Wójcik, J. Studies towards dynamic kinetic resolution of 4-hydroxy-2-methylcyclopent-2-en-1-one and its E-O-trityloxime. **2016**, *72* (32), 4813-4820.
- (33) 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.
- (34) Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R. Stability and folding kinetics of structurally characterized cytochrome *cb*₅₆₂. *Biochemistry* **2006**, *45* (35), 10504-10511.
- (35) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. In *Methods in Enzymology*; Pecoraro, V. L., Ed.; Academic Press, 2016; Vol. 580.
- (36) Karplus, P. A.; Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **2012**, *336* (6084), 1030-1033.
- (37) Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **2010**, *66* (Pt 2), 133-144.
- (38) Kabsch, W. XDS. *Acta Crystallogr. D* **2010**, *66* (Pt 2), 125-132.
- (39) Terwilliger, T. C.; Hung, L.-W.; Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I.; Echols, N.; Headd, J. J.; Grosse Kunstleve, R. W. et al. Phenix - a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, Medium: ED.
- (40) Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.3.* **2010**.
- (41) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **2010**, *66* (4), 486-501.

- (42) Lindner, H. J.; Gottlicher, S. Die Kristall- und Molekülstruktur des Eisen(III)-benzhydroxamat-Trihydrates. *Acta Crystallogr. B Struct. Cryst. Cryst. Chem.* **1969**, *25* (5), 832-842.
- (43) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **2018**, *27* (1), 14-25.
- (44) Schuck, P. A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. *Biophys. Chem.* **2004**, *108* (1-3), 187-200.
- (45) Kleywegt, G. J.; Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **1994**, *50* (2), 178-185.
- (46) Mazurenko, S.; Stourac, J.; Kunka, A.; Nedeljković, S.; Bednar, D.; Prokop, Z.; Damborsky, J. CalFitter: a web server for analysis of protein thermal denaturation data. **2018**, *46* (W1), W344-W349.
- (47) Zivanov, J.; Nakane, T.; Forsberg, B. O.; Kimanius, D.; Hagen, W. J. H.; Lindahl, E.; Scheres, S. H. W. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **2018**, *7*, e42166.
- (48) Zheng, S. Q.; Palovcak, E.; Armache, J.-P.; Verba, K. A.; Cheng, Y.; Agard, D. A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **2017**, *14*, 331.
- (49) Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **2016**, *193* (1), 1-12.
- (50) Kucukelbir, A.; Sigworth, F. J.; Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **2013**, *11*, 63.
- (51) Pettersen, E. F.; Goddard, T. D.; Huang, C.; Couch, G.; Greenblatt, D. M.; Meng, E. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25*.
- (52) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **2010**, *66* (1), 12-21.

Chapter 5: Conclusions

5.1 Introduction

The assembly of monomeric proteins into higher-order structures is responsible for the structural and functional protein complexity in a cell. Engineering proteins to predictably form sophisticated architectures has potential uses in sensing, catalysis, therapeutics and diagnostics.¹ Furthermore, creating hybrid assemblies that interface biological and materials components is a necessary step towards generating novel architectures with functions that rival, or surpass, what is observed in Nature. A variety of design strategies have been used to promote protein association and generate artificial 0-, 1-, 2- and 3D protein-based materials.² In this work, we demonstrated advancements both in using existing crystalline materials as functional platforms and creating novel protein architectures using biologically inspired interaction motifs.

5.2 Construction of functionalized crystalline lattices

We generated functionalized protein materials by repurposing native enzymatic processes to modify the large crystalline surfaces of 2D protein lattices. RIDC3, an engineered variant of cytochrome *cb₅₆₂*, self-assembles into 2D crystalline materials upon incubation with Zn²⁺.³ The functional peptide ybbR⁴, which is site-specifically labeled by PPTases using modified-CoA substrates, was chemically or genetically appended to RIDC3 such that the peptide is surface exposed in the crystalline lattice. Incubation of chemically modified RIDC3+ybbR arrays with Sfp PPTase and fluorescent-CoA analogs resulted in enzymatic labeling of surface-exposed ybbR peptides. However, genetic incorporation of ybbR onto RIDC3 rendered the construct unstable, forming a cleaved peptide fragment that was no longer recognized by Sfp as a target for enzymatic modification. Thus, we chose to use ^{C98}RhuA, a tetrameric protein that self-assembles into 2D

crystals by the oxidation of Cys residues at the four corners of the protein, positing that genetic incorporation of ybbR into a larger protein scaffold may limit protein stability issues. C^{98} RhuA-ybbR retain the formation of 2D lattices and enabled selective functionalization of ybbR on the lattice surface using fluorescent-CoA analogs.

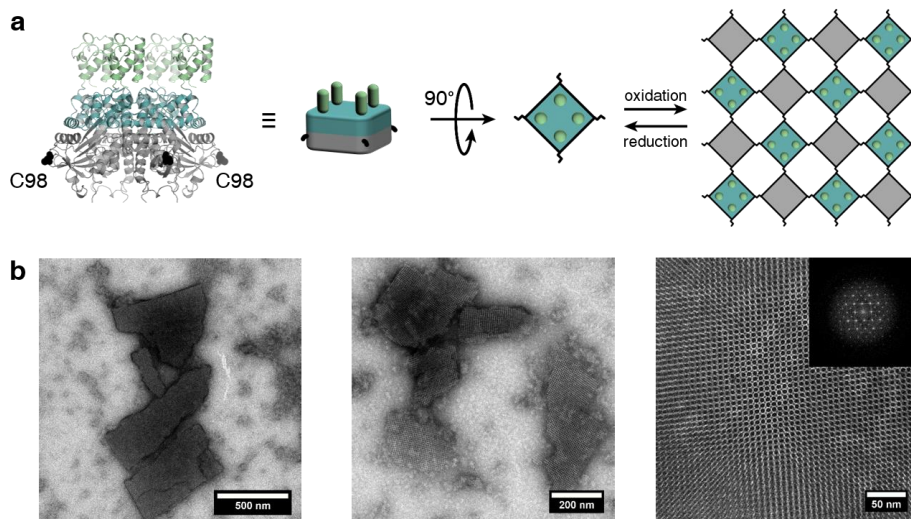


Figure 5.1 Formation of C^{98} RhuA-ACP arrays. **(a)** Schematic of C^{98} RhuA-ACP construct. **(b)** TEM images showing lattice formation of C^{98} RhuA-ACP.

Our proof-of-principle studies demonstrated that enzymatic modifications are a useful method for attaching different functional groups on the surfaces of crystalline lattices. While ybbR was used as a surrogate for the 77-residue ACP, the native biological substrate for PPTases⁵, the larger C^{98} RhuA can accommodate the genetic incorporation of ACP while retaining the ability to form 2D lattices. Preliminary experiments with a C^{98} RhuA-ACP construct have shown the formation of 2D crystalline lattices (**Figure 5.1**). Further investigation of enzymatic labeling is necessary to ensure these new lattices can be functionalized in a similar fashion to ybbR. In addition, previous studies have shown successful AcpH hydrolysis of the pantetheine label on ACP⁶, which could facilitate recyclable labeling of crystalline platforms with a RhuA-ACP construct on which fluorescent-CoA analogs can be tagged on using Sfp and removed using AcpH. Additional covalent conjugation strategies can be interfaced with RIDC3 or C^{98} RhuA to broaden

the scope of functional modifications of 2D lattices. SpyTag, a short peptide that forms an irreversible covalent bond with its partner SpyCatcher, can be genetically appended to either protein to coat SpyCatcher fusion partners on the surface.⁷ Sortases, a transpeptidase that attaches proteins to cell walls, can be adapted to modify LPXTG-bearing surface proteins.⁸ In this manner, a comparative study can be performed exploring different functionalization methods by varying the both the method of conjugation and the identity of the labeled molecule (*e.g.* small molecules, proteins, polymers). Overall, continuing to develop and improve functionalization of 2D materials will enable the hierarchical or layer-by-layer assembly of functional protein systems that can be implemented as diagnostic or sensing devices.

5.3 Design of protein-NA hybrid materials

We set out to create a synthetic nucleoprotein architecture whose self-assembly could be governed through the interplay of Watson-Crick base pairing, metal-protein interactions, and protein-DNA interactions. We covalently modified a single Cys residue on ²¹C-RIDC3 with ssDNA strands using a bifunctional linker. Site-specific control of complementary 10-bp DNA strands resulted in two hybrids, RIDC3-10a and RIDC3-10b. Incubation of both proteins with Zn²⁺ at pH 4.75 resulted in μm-sized crystals. Cryo-EM experiments and molecular dynamics simulations allowed us to determine the 3D RIDC3-10a/b architecture, which consists of dimeric units of RIDC3-10a and RIDC3-10b linked by dsDNA and a four-coordinate Zn-binding motif and connected to neighboring dimeric modules in an antiparallel fashion through a tridentate Zn coordination motif. This 2D protein-DNA lattice stacks in the 3rd dimension through weak protein-NA contacts between layers.

Given the specificity and modularity of DNA hybridization, it will serve as an important tool in continuing our efforts to create complex biological architectures that rely on both protein-

and NA-interactions in their assembly. Motivated by the success of our first-generation nucleoprotein assemblies, the first goal is to tune the self-assembly properties of the overall nucleoprotein architecture to: 1) improve the pH sensitivity of the overall scaffold such that it can assemble and remain stable in a greater range of conditions and 2) alter protein-DNA affinity to increase 3D growth of the crystals. While crystals form reliably under self-assembly conditions and remain stable for months, their integrity is dramatically reduced when exchanged into other solutions, even at moderate pH values (7-8). We hypothesize that this behavior is a direct result of the preponderance of acidic residues participating in Zn^{2+} coordination and the weak 3-coordinate Zn-binding mode between dimeric units. To probe the importance of these interactions, we propose the introduction of His residues in place of Asp/Glu to enable stronger Zn-binding at pH >5 and stabilize the assemblies (**Figure 5.2a**). The intra-dimer coordination of Glu27/Glu31 pairs will be replaced with His27/Glu31. Inter-dimer metal binding will be adjusted through Glu8His and Ile67His mutations to replace the previous 3-coordinate with a more robust 4-coordinate Zn motif. In addition, the previously installed His73/His77 residues will be reverted to Asp73 and Lys77 (the original residues at those locations).

In addition to the introduction of His residues to bolster metal binding at pH >5, we propose a second set of mutations to promote 3D stacking within the crystalline architecture. The structural model revealed that dsDNA components within the 2D layer wedge into the corrugations of the protein dimer above and below to loosely stack in 3D. Improving the noncovalent interactions between protein and DNA could promote 3D growth and enhance crystallinity in the 3rd dimension which may result in atomic resolution characterization of the architecture through single crystal X-ray diffraction. From MD simulations of the crystal bilayer, we identified sites of protonation of surface-exposed Glu residues at pH 4.75, increasing contact with the DNA backbone both

directly through H-bonding and indirectly by the release of basic residues (Lys42, Lys95) from intraprotein salt bridges. In order to improve protein-DNA contacts, we propose several charge-reversal mutations to increase salt bridge contacts between the protein and negatively charged DNA. We identified residues Glu2, Glu4, Asp5 and Asp39 positioned near dsDNA within the crystalline network as ideal sites to mutate into positively charged polar residues (Lys or Arg). A full set of the proposed mutations results in a highly positively charged surface as shown using electrostatic potential maps in **(Figure 5.2b)**. The charge-reversal mutations, in combination with the Zn-binding mutations previously outlined, would generate the most versatile version of the RIDC3-10a/b scaffold. Although these experiments focus solely on improvements to the RIDC3-10a/b scaffold, our observations on the effects of altered metal-binding and/or protein surface charge will inform the design of future protein-NA scaffolds. One can envision the use of symmetric protein building blocks that rely on metal association (*e.g.* C_3 -symmetry afforded by $Fe^{3+}:(HA)_3$ binding, metal free C_2 symmetric dimer of RIDC1) to predictably organize protein-

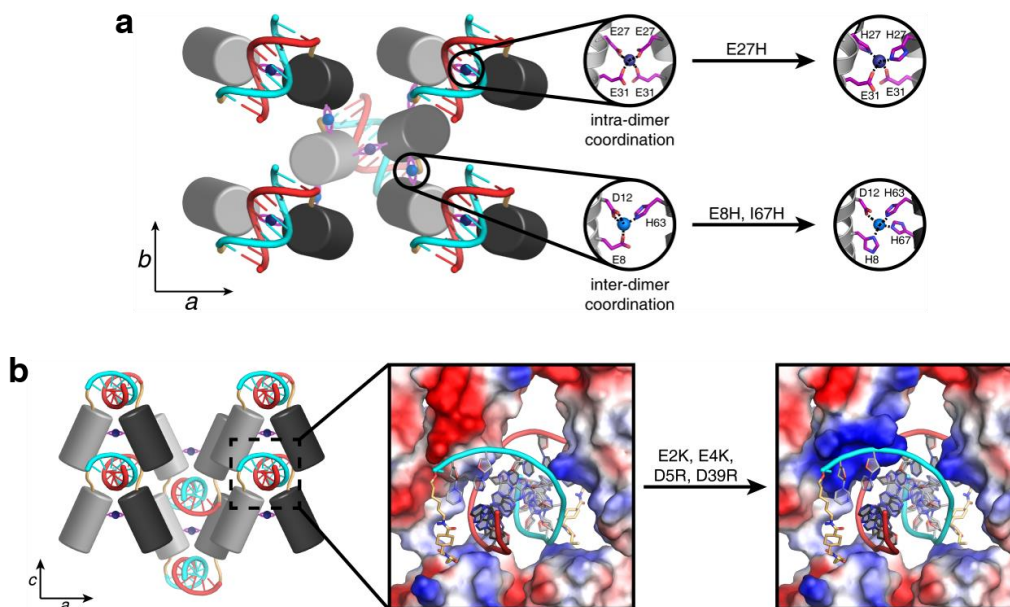


Figure 5.2 Proposed mutations for improving RIDC3-10a/b stability. **(a)** Model of His mutations introduced to improve pH stability of RIDC3-10a/b nucleoprotein architectures. **(b)** Model of electrostatic mutations to promote 3D stacking of 2D protein-DNA layers.

DNA chimeras to generate new structures. Further design efforts will help create design rules for the bottom-up construction of sophisticated protein-NA architectures using symmetric protein and NA nodes for directed assembly.

5.4 Design of protein cages with emergent functions

Protein-metal interactions are attractive as structural struts within a self-assembled protein complex due to their limited interfacial footprint and strong bonding.⁹ The majority of designed protein assemblies use native metal coordinating residues such as His, Glu, Asp to direct the assembly of discrete and extended oligomers. We used a combination of such native metal-binding residues with biologically inspired hydroxamate motifs to generate a bimetallic protein scaffold with selective coordination of Zn^{2+} and Fe^{3+} . Using the crystal packing of a previously characterized cyt *cb562* variant, positions 63 and 82 were identified at C_3 symmetric pores as suitable locations for installing Cys–HA motifs whereas the C_2 symmetric axes were mutated to strengthen Zn-protein interactions. Concurrent binding of both metal ions results in the formation of dodecameric BMC3 cages in which C_3 symmetric nodes at two vertices of a tetrahedron are formed by $\text{Fe}^{3+}:(\text{HA})_3$ and the edges formed through C_2 symmetric Zn-binding. Cryo-EM characterization of these cages confirmed the same architecture forms in solution as seen in the crystal lattice. Further engineering of the BMC scaffold resulted in the serendipitous formation of a hexamer when only a single HA motif at the 82 position was present on the protein. This highlighted the conformational flexibility at the HA-mediated vertices to accommodate both sets of protein geometries. These results present a new avenue for the design of flexible and reversible protein architectures.

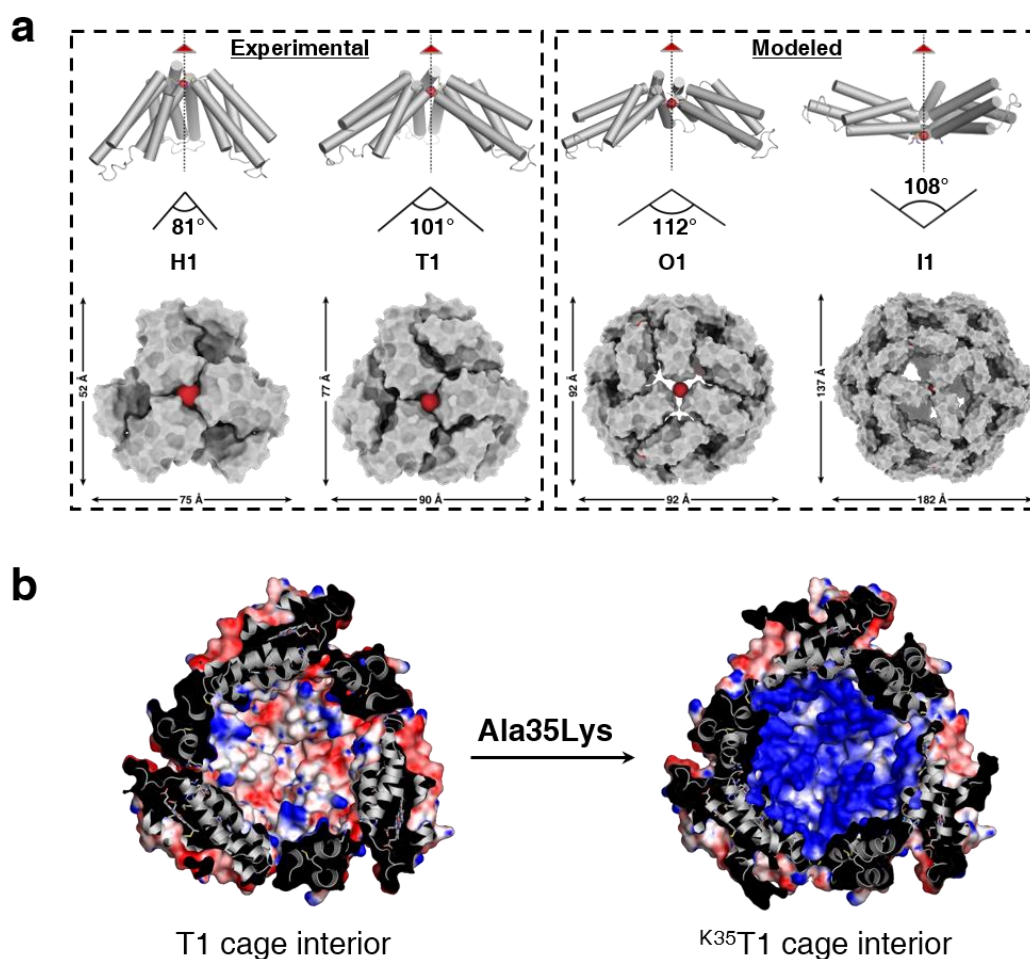


Figure 5.3 Generating new hydroxamate-based cages. **(a)** Experimental and modeled protein cages that can be formed due to the conformational adaptability of the hydroxamate-guided C_3 vertices **(b)** Model of electrostatic mutations to create a positive cage interior for the encapsulation of nucleic acids.

There are several future avenues to explore with both the BMC scaffolds and more broadly, the incorporation of HA-motifs to drive protein symmetrization. While the BMC3 dodecamers and BMC4 hexamers form repeatably in solution, the cage yields and stability over time could be improved. One strategy could be to support, or even eliminate, Zn-binding by using computational design strategies^{10,11} at the C_2 symmetric interfaces to both stabilize the final cage product as well as encourage protein association in the self-assembly process. In this case, cage self-assembly would be primarily directed by $Fe^{3+}:(HA)_3$ coordination to generate C_3 symmetry and additionally

guided by non-covalent interactions at the C_2 symmetric interfaces to afford the desired cage architecture. Furthermore, given the interfacial flexibility of the hydroxamate “hinge”, one can envision using the BMC scaffolds to create higher order symmetry cages (**Figure 5.3a**). This will likely require a combination of adjusting the position of Zn-binding residues and introducing computationally designed noncovalent patches to stabilize the desired polyhedral symmetry.

Finally, BMC protein cages also have potential uses as containers to bind to/transport cargo. Towards this end, one can engineer cages to bind and trap NAs in their interior. Encapsulation of proteins and NAs within engineered cages has been accomplished previously by the Baker¹² and Hilvert¹³ labs wherein designed cages are further engineered to create a more-positive cage interior and express *in vivo* containing RNA. Our cages are stimuli responsive to metal ions *in vitro* and in initial experiments, we have shown that BMC3 cages can encapsulate small fluorophores and retain them over several days (**Figure 4.11**). The affinity for guest molecules can be improved by making targeted mutations to the interior of BMC3 and BMC4 cages. Electrostatic potential maps (**Figure 5.3b**) of the native cages show a mixture of positive and negative residues in the cage interior and a single charge mutation (Ala35Lys) increases the positive charge of the interior and should promote cargo loading of negatively charged molecules. In addition to encapsulating dye molecules, the interior of BMC3 cages (~34 Å diameter) should permit ssDNA or ssRNA cargo to be captured and retained in their interior. If larger protein polyhedra can be created using the BMC scaffold, one can envision encapsulating larger and more biologically relevant double-stranded siRNA in octahedral or icosahedral cages.

5.5 References

- (1) Baneyx, F.; Matthaei, J. F. Self-assembled two-dimensional protein arrays in bionanotechnology: from S-layers to designed lattices. *Curr. Opin. Biotech.* **2014**, *28*, 39-45.
- (2) Bai, Y.; Luo, Q.; Liu, J. Protein self-assembly via supramolecular strategies. *Chem. Soc. Rev.* **2016**, *45* (10), 2756-2767.
- (3) Brodin, J. D.; Ambroggio, X. I.; Tang, C. Y.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **2012**, *4* (5), 375-382.
- (4) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (44), 15815-15820.
- (5) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep* **2014**, *31* (1), 61-108.
- (6) Kosa, N. M.; Pham, K. M.; Burkart, M. D. Chemoenzymatic exchange of phosphopantetheine on protein and peptide. *Chem. Sci.* **2014**, *5* (3), 1179-1186.
- (7) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (12), 690-697.
- (8) Pishesha, N.; Ingram, J. R.; Ploegh, H. L. Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol.* **2018**, *34*, 163-188.
- (9) Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A. Metals in Protein-Protein Interfaces. *Ann. Rev. Biophys.* **2014**, *43*, 409-431
- (10) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andrzejczyk, 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.
- (11) Liu, Y.; Kuhlman, B. RosettaDesign server for protein design. *Nucl. Acids. Res.* **2006**, *34*, W235-238.
- (12) 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. et al. Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **2017**, *552* (7685), 415-420.
- (13) Terasaka, N.; Azuma, Y.; Hilvert, D. Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (21), 5432-5437.

Appendix 1

Scripts used for generating a structural model of the RIDC3-10a/b nucleoprotein architecture

Script 1: Generating unique arrangements of protein and DNA, symmetrizing, and creating projection maps for structural modeling. Each pdb file contains two protein monomers derived from the RIDC3 structure (PDB ID: 3TOM) and duplex DNA generated using the Nucleic Acid Builder¹ via the make-na server. Each dimer generated is symmetrized in accordance with the *pgg* symmetry observed from SAXS and cryo-EM measurements. Symmetrization is performed using the Bsoft package.² Symmetrized PDB files are converted into MRC map files, projected into a 2D profile, and saved as a PNG image using the EMAN2 package.³ Output files were visualized in Pymol or UCSF Chimera.^{4,5}

Input used: ./makepdb_and_buildmap.py test_pdb/ -l 63,55 -d 3 -t 6 -r 5 -p 0.315

This resulted in 50, 625 arrangements of protein and DNA, each saved as a unique PDB file.

```
import os, sys

# Tell PyMOL to launch quiet (-q), fullscreen (-e) and without internal GUI (-i)
import __main__
__main__.pymol_argv = ['pymol', '-q']

import pymol
pymol.finish_launching()

try:
    os.symlink('/usr/share/pymol/data','data')
except:
    pass

try:
    from optparse import OptionParser
except:
    print "Module optparse not found: you need Python version 2.3 or later"
    sys.exit(-1)
version = '0.1'

def which(program):
    import os
    def is_exe(fpath):
        return os.path.isfile(fpath) and os.access(fpath, os.X_OK)
```



```

fpath, fname = os.path.split(program)
if fpath:
    if is_exe(program):
        return program
    else:
        for path in os.environ["PATH"].split(os.pathsep):
            path = path.strip("")
            exe_file = os.path.join(path, program)
            if is_exe(exe_file):
                return exe_file

return None

def main():

    (options,args) = parse_command_line()

    in_pfx = args[0]
    in_folder = in_pfx.split('/')[0]

    # locate all PDB files from input folder
    in_path = os.path.abspath(in_folder)

    dirs = []
    for file in os.listdir(in_path):
        if file.endswith('.pdb'):
            dirs.append(file)

    # path for creating output files
    out_path = in_path.split(in_folder)[0]
    out_path_pfx = out_path.split('/')[-2]

    # make map output folder in current directory
    try:
        os.makedirs('output_2D_maps')
    except OSError:
        print("\n Map directory already exists!")

    # geometry parameters
    (a,b) = [float(x) for x in options.latticeparams.split(',')]

    a_half = 0.5*a
    b_half = 0.5*b

```

```

# define input options
zoffset = options.zoffset
pix = options.pixelsize
res = options.resolution
deg = options.degrot

# copy number of PDBs to generate
copy_number = options.copynumber
copy_1 = -copy_number/2
copy_2 = copy_number/2

# take input number for generating PDBs and round to nearest integer
input_n = options.numrot
rot = options.degrot

for i in range(len(dirs)):

    # load pdb using pymol commands
    inpdb_path = in_path+'/'+dirs[i]
    inpdb_name = dirs[i].split('.')[0]

    pymol.cmd.load(inpdb_path,inpdb_name)

    # pdb folder output in current directory
    pdb_folder_name = inpdb_name+'_pdb'

    try:
        os.makedirs(pdb_folder_name)
    except OSError:
        pass

    pdb_out_pfx = pdb_folder_name+'/'+dirs[i].split('.')[0]

    # split protein into two chains, DNA chains into single object
    pymol.cmd.select('chainA', 'chain A')
    pymol.cmd.select('chainB', 'chain B')
    pymol.cmd.select('chainCandD', 'chain C or chain D')

    pymol.cmd.create('mon1', 'chainA')
    pymol.cmd.create('mon2', 'chainB')
    pymol.cmd.create('dna_duplex', 'chainCandD')

    pymol.cmd.reset()

    pymol.cmd.center('mon1')
    pymol.cmd.rotate('z', str(-input_n*rot), 'mon1')

```

```

pymol.cmd.reset()
pymol.cmd.align('mon2','mon1')
pymol.cmd.rotate('z','180','mon2')

pymol.cmd.reset()

pymol.cmd.center('dna_duplex')
pymol.cmd.rotate('z',str(-input_n*rot),'dna_duplex')
pymol.cmd.reset()
'''
pymol.cmd.rotate('z', str(-30), 'mon1')
pymol.cmd.rotate('z', str(-30), 'mon2')
pymol.cmd.rotate('z', str(-30), 'dna_duplex')

pymol.cmd.reset()
'''
count = 1

for i in range(0,input_n):

    # center camera to protein monomer and rotate in z

    pymol.cmd.center('mon1')
    pymol.cmd.rotate('z', str(rot), 'mon1')
    pymol.cmd.reset()
    pymol.cmd.align('mon2','mon1')
    pymol.cmd.rotate('z','180','mon2')

    pymol.cmd.reset()

    #print 'i = %i'%(i*rot)

    # create dummy objects for additional manipulation

    pymol.cmd.create('dummy1', 'mon1')
    pymol.cmd.create('dummy2', 'mon2')

    pymol.cmd.reset()

    pymol.cmd.center('dummy1')
    pymol.cmd.rotate('y', str(-input_n*rot), 'dummy1')
    pymol.cmd.reset()
    pymol.cmd.align('dummy2','dummy1')
    pymol.cmd.rotate('z','180','dummy2')

    for j in range(0,input_n):

```

```

# center camera to protein monomer and rotate protein in y

pymol.cmd.center('dummy1')
pymol.cmd.rotate('y', str(rot), 'dummy1')
pymol.cmd.reset()
pymol.cmd.align('dummy2','dummy1')
pymol.cmd.rotate('z','180','dummy2')

pymol.cmd.reset()

#print 'j = %i'%(j*rot)

# create dummy objects for additional manipulation

pymol.cmd.create('dummy3', 'dummy2')
pymol.cmd.create('dummy4', 'dummy1')

pymol.cmd.reset()

pymol.cmd.center('dummy3')
pymol.cmd.rotate('x', str(-input_n*rot), 'dummy3')
pymol.cmd.reset()
pymol.cmd.align('dummy4','dummy3')
pymol.cmd.rotate('z','180','dummy4')

for k in range(0,input_n):

    # center camera to protein monomer and rotate protein in x

    pymol.cmd.center('dummy3')
    pymol.cmd.rotate('x', str(rot), 'dummy3')
    pymol.cmd.reset()
    pymol.cmd.align('dummy4','dummy3')
    pymol.cmd.rotate('z','180','dummy4')

    pymol.cmd.reset()

    for l in range(0,input_n):

        pymol.cmd.create('dna_duplex2','dna_duplex')
        pymol.cmd.center('dna_duplex2')
        pymol.cmd.rotate('z', str(rot), 'dna_duplex2')

        pymol.cmd.reset()
'''

```

```

pymol.cmd.create('dummy5', 'dummy3')
pymol.cmd.create('dummy6', 'dummy4')
pymol.cmd.create('dna_duplex2', 'dna_duplex')
pymol.cmd.reset()

for m in range(0, input_n):

    pymol.cmd.rotate('z', str(60/input_n), 'dummy5')
    pymol.cmd.rotate('z', str(60/input_n), 'dummy6')
    pymol.cmd.rotate('z', str(60/input_n), 'dna_duplex2')
    ""
    # save new protein state
    pymol.cmd.save('%s_%i.pdb'%(pdb_out_pfx, count),
'dummy3 or dummy4 or dna_duplex2')

        count += 1
        #print 'k = %i'%(k*rot)

map_dirs = []
for file in os.listdir(os.path.abspath(pdb_folder_name)):
    if file.endswith('.pdb'):
        map_dirs.append(file)

num = 1

# cycle through the files in directory and apply pgg symmetry, generate 2D map and FRC
plot
for ii in map_dirs:

    # input PDB prefix is name of input folder + name of PDB file
    inpdb = pdb_folder_name+'/'+ii

    # make new directory for output files for each set of model PDBs

    model_folder_name = ii.split('.')[0]+'_model'

    try:
        os.makedirs(pdb_folder_name+'/'+model_folder_name)
    except OSError:
        pass

    try:
        os.makedirs('output_2D_maps/'+inpdb_name)
    except OSError:
        pass

```

```

try:
    os.makedirs('output_2D_mracs/'+inpdb_name)
except OSError:
    pass

# output prefix is input folder name + model folder name + input pdb name
out_pfx = pdb_folder_name+'/'+model_folder_name+'/'+ii.split('.')[0]
mrc_pfx = 'output_2D_mracs'++'/'+inpdb_name+'/'+ii.split('.')[0]
map_pfx = 'output_2D_maps'++'/'+inpdb_name+'/'+ii.split('.')[0]

# center model and generate copies covering a lattice unit

#print("\n Generating atomic model of lattice unit...")

pdb_num = 1
os.system("bmol -Translate %s %s_ref.pdb"%(inpdb,out_pfx))
os.system("bmol -rotate 1,0,0,180 %s_ref.pdb %s_flip.pdb"%(out_pfx,out_pfx))

os.system("bmol          -translate          0,0,%f          %s_ref.pdb
%s_ref.pdb"%(zoffset,out_pfx,out_pfx))
os.system("bmol          -translate          0,0,%f          %s_flip.pdb          %s_flip.pdb"%(-
zoffset,out_pfx,out_pfx))

for i in range(copy_1,copy_2+1):
    for j in range(copy_1,copy_2+1):
        os.system("bmol          -translate          %f,%f,0          %s_ref.pdb
%s_%i.pdb"%(i*a,j*b,out_pfx,out_pfx,pdb_num))
        pdb_num += 1

for i in range(copy_1,copy_2):
    for j in range(copy_1,copy_2):
        os.system("bmol -translate %f,%f,0 %s_flip.pdb %s_%i.pdb"%(i*a
+ a_half,j*b + b_half,out_pfx,out_pfx,pdb_num))
        pdb_num += 1

os.remove("%s_ref.pdb"%out_pfx)
os.remove("%s_flip.pdb"%out_pfx)
#print("\n Generating projected density from complete model...")

# combine PDB files into a single one

os.system("bmol %s_1.pdb %s_asu.pdb"%(out_pfx,out_pfx))

for i in range(2,pdb_num):

```

```

        os.system("bmoledit      -insert      %s_%i.pdb      %s_asu.pdb
%s_asu.pdb"%(out_pfx,i,out_pfx,out_pfx))

        # convert to 2D density and output all png files into single folder
        os.system("e2pdb2mrc.py --apix %f --res %f --box 400,400,1 %s_asu.pdb
%s_asu_model.mrc"%(pix,res,out_pfx,out_pfx))
        os.system("bfilter      -band      10000,%f,0.002      %s_asu_model.mrc
%s_asu_model.mrc"%(options.resolution,out_pfx,mrc_pfx))
        os.system("e2proc2d.py      %s_asu_model.mrc
%s_asu_model.png"%(mrc_pfx,map_pfx))

        print "%i "%(num)
        num += 1

    pymol.cmd.delete('all')

    pymol.cmd.quit()

print "Done!"

def parse_command_line():
    usage = "Usage: %prog [options] <input folder> \n\nReads an atomic model in PDB format and
generates a variety of rotations given by the user input. \n\nThe name of the output files will start
with the prefix given by the user."
    parser = OptionParser(usage=usage,version="%prog "+version)
    parser.add_option("-l","--
lattice",dest="latticeparams",metavar="<a,b>",type="string",help="Lattice parameters a and b,
separated by comma (default: none)")
    parser.add_option("-z","--zoffset",dest="zoffset",metavar="<z>",type="float",help="Shift
perpendicular to plane to apply to input model (angstroms, default: 0)",default=0)
    parser.add_option("-r","--resolution",dest="resolution",metavar="<image
resolution>",type="float",help="Resolution of simulated image (angstroms, default:
15)",default=15)
    parser.add_option("-p","--pixel",dest="pixelsize",metavar="<pixel
size>",type="float",help="Pixel size of simulated image (angstroms, default: 1)",default=1)
    parser.add_option("-n","--
number",dest="copynumber",metavar="<copy_number>",type="int",help="Copy number of
lattice unit, must be an even number (default: 2)",default=2)
    parser.add_option("-d","--deg",dest="degrot",metavar="<d>",type="float",help="Degrees of
rotation (default: 360)")
    #parser.add_option("-a","--
dna_deg",dest="dna_degrot",metavar="<a>",type="int",help="Degrees of rotation of DNA
models (default: 360)")
    parser.add_option("-t","--numrot",dest="numrot",metavar="<t>",type="int",help="Given
positive integer to generate n pdb files at d rotations in axis (default: 360)",default=360)

```

```

if len(sys.argv)<1:
    parser.print_help()
    sys.exit(-1)

(options, args)=parser.parse_args()

if len(args) < 1:
    parser.error("Input pdb and output prefix required!")

if not options.latticeparams:
    parser.error("Lattice parameters (option -l) required!")

bprograms_list = ['bmol', 'bimg', 'bmoledit', 'bfilter', 'bproject']
for bp in bprograms_list:
    if not which(bp):
        print("Can not find program %s from Bsoft package!"%bp)
        sys.exit(-1)
eprograms_list = ['e2pdb2mrc.py','e2proc2d.py']
for ep in eprograms_list:
    if not which(ep):
        print("Can not find program %s from EMAN2 package!"%ep)
        sys.exit(-1)

return (options,args)

if __name__ == "__main__":
    main()

```

Script 2: Generating projection maps using dimeric protein-DNA modules. Dimeric arrangements are symmetrized using the Bsoft package.² Symmetrized PDB files are converted into .mrc map files and projected into a 2D png image using the EMAN2 package.³ This script was primarily used during fine tuning the orientation of proteins or DNA for a particular dimeric model. Output files were visualized in Pymol or UCSF Chimera.^{4,5}

Example usage: `./bkr_gelat_large.py limiting-models/27-31.pdb limiting-models/27-31-interior/large/1 -l 63,55 -r 6 -n 20 -z 8.5`

```

import os, sys
try:
    from optparse import OptionParser
except:

```



```

        print "Module optparse not found: you need Python version 2.3 or later"
        sys.exit(-1)
version = '0.1'

def which(program):
    import os
    def is_exe(fpath):
        return os.path.isfile(fpath) and os.access(fpath, os.X_OK)

    fpath, fname = os.path.split(program)
    if fpath:
        if is_exe(program):
            return program
    else:
        for path in os.environ["PATH"].split(os.pathsep):
            path = path.strip("")
            exe_file = os.path.join(path, program)
            if is_exe(exe_file):
                return exe_file

    return None

def main():

    #try:
    #os.

    (options,args) = parse_command_line()

    inpdb = args[0]
    out_pfx = args[1]

    # geometry parameters
    (a,b) = [float(x) for x in options.latticeparams.split(',')]

    x = 0.5*a
    y = 0.5*b
    zoffset = options.zoffset
    pix = options.pixelsize
    res = options.resolution

    copy_number = options.copynumber
    c1 = -copy_number/2
    c2 = copy_number/2

```

```

# center model and generate copies covering a lattice unit
# generate copies based on given input 'X' from -X/2 to X/2
print("\n Generating atomic model of lattice unit...")
os.system("bmol -Translate %s %s_ref.pdb"%(inpdb,out_pfx))

os.system("bmol -translate 0,0,%f %s_ref.pdb %s_1.pdb"%(zoffset,out_pfx,out_pfx))
pdb_num = 2

for i in range(c1,c2+1):
    for j in range(c1,c2+1):
        os.system("bmol -translate %f,%f,0 %s_1.pdb
%s_%i.pdb"%(i*a,j*b,out_pfx,out_pfx,pdb_num))
        pdb_num += 1

os.system("bmol -rotate 1,0,0,180 %s_ref.pdb %s_flip.pdb"%(out_pfx,out_pfx))
os.system("bmol -translate 0,0,%f %s_flip.pdb %s_flip.pdb"%(-zoffset,out_pfx,out_pfx))

for i in range(c1,c2):
    for j in range(c1,c2):
        os.system("bmol -translate %f,%f,0 %s_flip.pdb
%s_%i.pdb"%(i*a+x,j*b+y,out_pfx,out_pfx,pdb_num))
        pdb_num += 1

os.remove("%s_ref.pdb"%out_pfx)
os.remove("%s_flip.pdb"%out_pfx)
print("\n Generating projected density from complete model...")

# combine pdb files into a single one
os.system("bmol %s_2.pdb %s_asu.pdb"%(out_pfx,out_pfx))
for i in range(3,pdb_num):
    os.system("bmoledit -insert %s_%i.pdb %s_asu.pdb
%s_asu.pdb"%(out_pfx,i,out_pfx,out_pfx))

# convert to 2D density and output all png files into single folder
os.system("e2pdb2mrc.py --apix %f --res %f --box 400,400,1 %s_asu.pdb
%s_asu_model.mrc"%(pix,res,out_pfx,out_pfx))
os.system("bfilter -band 10000,%f,0.002 %s_asu_model.mrc
%s_asu_model.mrc"%(options.resolution,out_pfx,out_pfx))
os.system("e2proc2d.py %s_asu_model.mrc %s_asu_model.png"%(out_pfx,out_pfx))

print "Done!"

def parse_command_line():

```

```
usage = "Usage: %prog [options] <input pdb model> <output prefix> \n\nReads an atomic model in PDB format and generates copies of it to fill the unit cell of a lattice with pgg symmetry. It also generates a simulated, projected electron density image to compare to the experimental one.\n\nThe name of the output files will start with the prefix given by the user."
```

```
parser = OptionParser(usage=usage,version="%prog "+version)
parser.add_option("-l", "--lattice", dest="latticeparams", metavar="<a,b>", type="string", help="Lattice parameters a and b, separated by comma (default: none)")
parser.add_option("-z", "--zoffset", dest="zoffset", metavar="<z>", type="float", help="Shift perpendicular to plane to apply to input model (angstroms, default: 0)", default=0)
parser.add_option("-r", "--resolution", dest="resolution", metavar="<image resolution>", type="float", help="Resolution of simulated image (angstroms, default: 15)", default=15)
parser.add_option("-p", "--pixel", dest="pixelsize", metavar="<pixel size>", type="float", help="Pixel size of simulated image (angstroms, default: 1)", default=1)
parser.add_option("-n", "--number", dest="copynumber", metavar="<copy_number>", type="int", help="Copy number of lattice unit, must be an even number (default: 2)", default=2)
```

```
if len(sys.argv)<2:
    parser.print_help()
    sys.exit(-1)
```

```
(options, args)=parser.parse_args()
```

```
if len(args) < 2:
    parser.error("Input pdb and output prefix required!")
```

```
if not options.latticeparams:
    parser.error("Lattice parameters (option -l) required!")
```

```
bprograms_list = ['bmol', 'bimg', 'bmoledit', 'bfilter', 'bproject']
for bp in bprograms_list:
    if not which(bp):
        print("Can not find program %s from Bsoft package!"%bp)
        sys.exit(-1)
```

```
eprograms_list = ['e2pdb2mrc.py']
for ep in eprograms_list:
    if not which(ep):
        print("Can not find program %s from EMAN2 package!"%ep)
        sys.exit(-1)
```

```
return (options,args)
```

```
if __name__ == "__main__":
    main()
```

References

- (1) Macke, T. J.; Case, D. A. In *Molecular Modeling of Nucleic Acids*; American Chemical Society, 1997; Vol. 682.
- (2) Heymann, J. B. Bsoft: Image and Molecular Processing in Electron Microscopy. **2001**, *133* (2), 156-169.
- (3) Tang, G.; Peng, L.; Baldwin, P. R.; Mann, D. S.; Jiang, W.; Rees, I.; Ludtke, S. J. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **2007**, *157* (1), 38-46.
- (4) Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.3*. **2010**.
- (5) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605-1612.

Appendix 2

NMR spectra

2-chloro-*N*-hydroxyacetamide

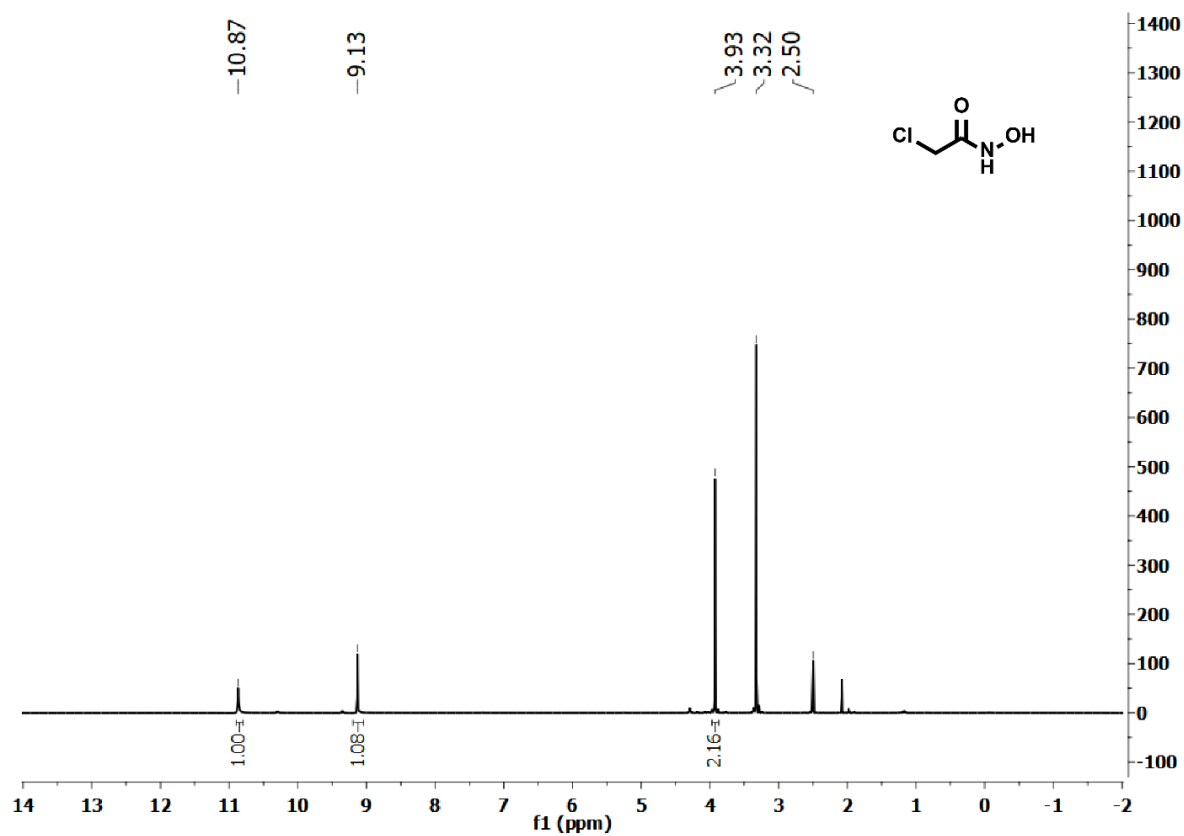


Figure A2.1 ¹H NMR spectrum of 2-chloro-*N*-hydroxyacetamide. (300 MHz, DMSO-*d*₆) δ 10.88 (s, 1H), δ 9.15 (s, 1H), δ 3.93 (s, 2H)

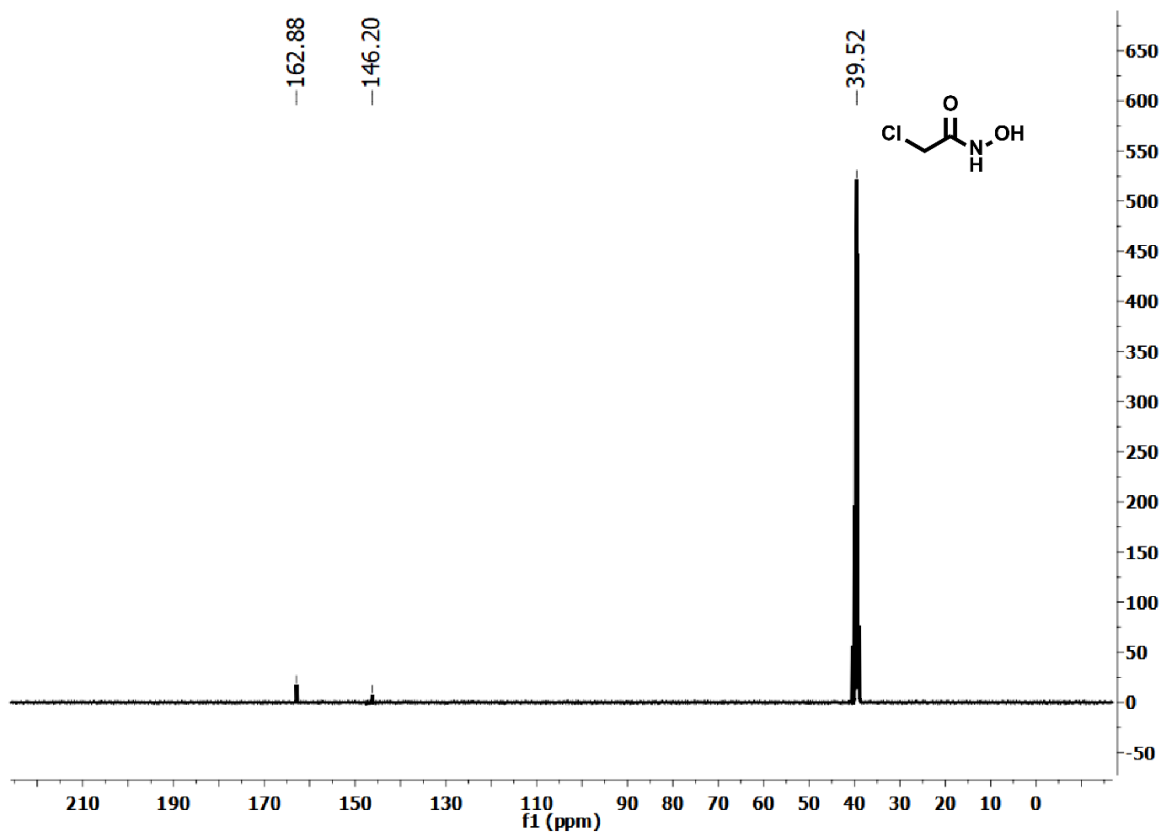


Figure A2.2 ^{13}C NMR spectrum of 2-chloro-*N*-hydroxyacetamide. (500 MHz, DMSO- d_6) δ 162.88, δ 40.45

2-iodo-*N*-hydroxyacetamide

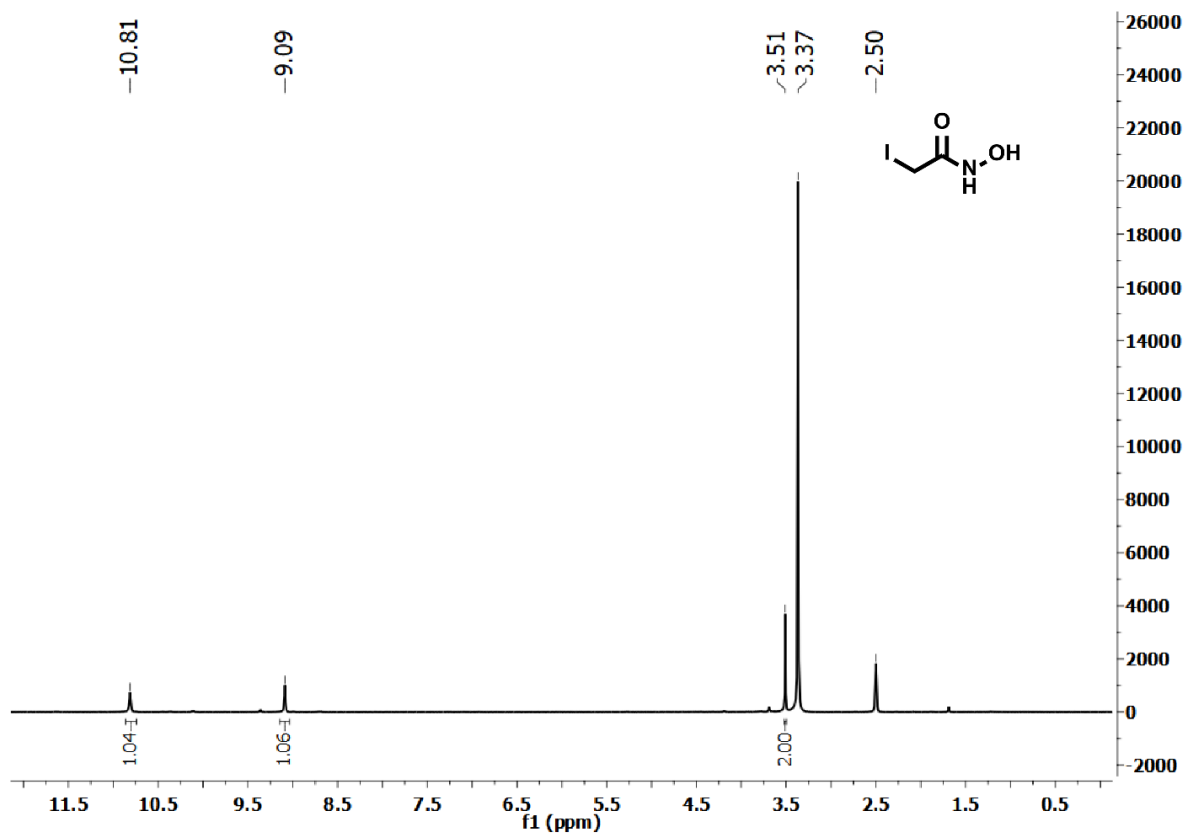


Figure A2.3 ^1H NMR spectrum of 2-iodo-*N*-hydroxyacetamide. (300 MHz, DMSO-d_6) δ 10.81 (s, 1H), δ 9.09 (s, 1H), δ 3.51 (s, 2H)

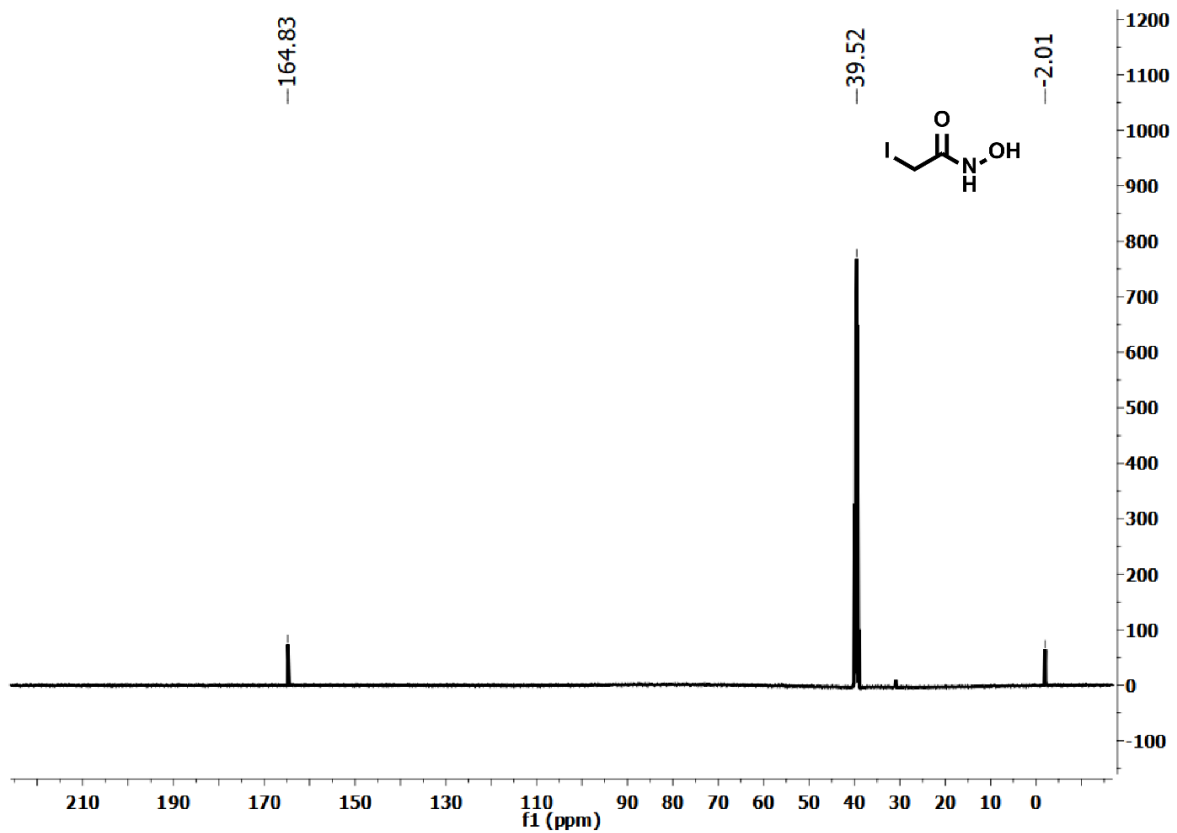


Figure A2.4 ^{13}C NMR spectrum of 2-iodo-*N*-hydroxyacetamide. (500 MHz, DMSO- d_6) δ 164.83, δ -2.01