UNIVERSITY OF CALIFORNIA, SAN DIEGO

Design and Characterization of an Allosteric Metalloprotein Assembly

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

Doctor of Philosophy

in

Chemistry

by

Lewis Anthony Churchfield

Committee in charge:

Professor F. Akif Tezcan, Chair
Professor Michael D. Burkart
Professor Neal K. Devaraj
Professor Tracy M. Handel
Professor Valerie A. Schmidt

2018

This Dissertation of Lewis Anthony Churchfield is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2018

EPIGRAPH

Sometimes science is more art than science.

Dr. Rick Sanchez

TABLE OF CONTENTS

Signature Page iii

Epigraph iv

Table of Contents v

List of Abbreviations vii

List of Figures ix

List of Tables xiii

Acknowledgements xv

Vita xviii

Abstract of the Dissertation xix

Chapter 1: A survey of designed protein assemblies 1

1.1 Foundations of protein design: drawing inspiration from Nature 1

1.2 Repurposing native interactions for engineered protein assemblies 9

1.3 Engineering protein assemblies using de novo designed interactions 16

1.4 Synergy of design strategies for native-like engineered protein assemblies 27

1.5 Challenges and opportunities: emerging design paradigms protein self-assembly 30

References 37

Chapter 2: *De novo* design of an allosteric metalloprotein. 45

2.1 Introduction 45

2.2 Results and discussion 49

2.3 Materials and methods 70

References 77

Chapter 3: Biochemical dissection of an engineered allosteric protein 80
3.1 Introduction 80
3.2 Results and Discussion 84
3.3 Materials and Methods 116
References 122
Chapter 4: Molecular dynaimcs simulations of $^{C38/C81/C96}R1_4$. 125
4.1 Introduction 125
4.2 Development of custom simulation parameters 127
4.3 Molecular dynamics simulations of C38-C38 disulfide breakage. 148
4.4 Molecular dynamics simulations of C81-C81 and C96-C96 breakage. 164
4.5 End-state thermodynamic analysis of C38-C38 disulfide breakage. 168
4.5 Methods. 174
References 176
Chapter 5: Conclusion and future directions. 179
5.1 Concluding remarks 179
5.2 Toward a deeper understanding of $^{C38/C81/C96}R1_4$ allostery. 182
5.3 Exploring new elements of design complexity in the R1 scaffold. 184
5.4 Materials and Methods 191
References 194
Appendix 195

LIST OF ABBREVIATIONS

ATP	adenosine triphosphate
AUC	analytical ultracentrifugation
Csa	cysteine sulfenate
Cso	cysteine sulfenic acid
cyt b_{562}	cytochrome b_{562} (contains b-type heme)
cyt cb_{562}	cytochrome cb_{562} (contains a c-type heme)
DHFR	dihydrofolate reductase
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
GaMD	Gaussian accelerated molecular dynamics
ITC	isothermal titration calorimetry
K_d	dissociation constant
LB	lysogeny broth
M1	Metal Binding Protein Complex 1; MBPC1
Mb	Myoglobin
MD	Molecular dynamics
MDPSA	Metal Directed Protein Self-Assembly
MeTIR	Metal Templated Interface Redesign
MOPS	3-(N-morpholino)propanesulfonic acid
MTX	methotrexate

R1	Rosetta Interface Designed Construct 1; RIDC1
RhuA	rhamnulose-1-phosphate aldolase
Sav	streptavidin
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TEM	transmission electron microscopy
T_M	apparent melting temperature
TMV	tobacco mosaic virus
Tris	tris(hydroxymethyl)aminomethane

LIST OF FIGURES

Figure 1.1 Notable examples of naturally occurring protein assemblies. 3

Figure 1.2 Cartoon depiction of the roles of flexibility and rigidity in an artificial biological assembly. 6

Figure 1.3 Survey of strategies for engineering protein assemblies. 8

Figure 1.4 Design of multivalent ligands for inducing protein assembly. 10

Figure 1.5 Engineering a homodimeric protein via domain swapping. 13

Figure 1.6 Design of a self-assembling protein cage by genetic fusion of oligomers. 15

Figure 1.7 Design of two-component protein-DNA lattices. 18

Figure 1.8 Designed protein scaffolds for metal-templated assembly. 20

Figure 1.9 Design of disulfide-linked protein nanorods. 23

Figure 1.10 Designing protein-protein contacts for engineered supramolecular assemblies. 25

Figure 1.11 Metal tempalted interface redesign (MeTIR): a versatile strategy for protein engineering. 28

Figure 1.12 Emerging trends in designing protein assemblies. 32

Figure 2.1 Successive engineering of cyt cb_{562} to form disulfide-linked oligomeric architectures. 48

Figure 2.2 Structural rearrangements upon Zn^{II} removal from Zn-$^{C81/C96}R1_4$. 50

Figure 2.3 SDS-PAGE gel of the products of $^{C38/C81/C96}R1$ and $^{C38/C81/C96}M1$ self-assembly reactions in the presence or absence of Zn^{II} 52

Figure 2.4 Non-reducing SDS-PAGE of $^{C38/C81/C96}R1_4$ in the absence and presence of DTT. 54

Figure 2.5 Sedimentation velocity profiles of $^{C38/C81/C96}R1_4$. 55

Figure 2.6 Binding isotherm of Zn^{II} to Fura-2 in the presence of $^{C38/C81/C96}R1_4$. 56

Figure 2.7 ITC thermograms of Zn^{II} titrated into disulfide-linked R1 tetramers..... 58

Figure 2.8 Overview of Zn-induced structural changes in $^{C38/C81/C96}R1_4$..... 61

Figure 2.9 Structural alignments of metal-bound disulfide-linked R1 tetramers. 62

Figure 2.10 Electron density maps of key features in $^{C38/C81/C96}R1_4$ structures. 63

Figure 2.11 Structural alignment of Zn-bound and metal-free $^{C38/C81/C96}R1_4$. 65

Figure 2.12 Electron density maps of alternative models for CSO38 in apo-$^{C38/C81/C96}R1_4$. 67

Figure 2.13 Fluorescence scans of Zn^{II} titrated into a solution of Fura-2 and $^{C38/C81/C96}R1_4$. 74

Figure 3.1 Interfaces and disulfide crosslinks in the allosteric protein $^{C38/C81/C96}R1_4$..... 82

Figure 3.2 Hypothetical speciation of C38-linked R1 variants following Zn-removal. ... 87

Figure 3.3 Representative templating reactions of $^{C38/C96}$R1 and $^{C38/C96}$M1. 88

Figure 3.4 SDS-PAGE of isolated $^{C38/C96}R1_4$ in the presence and absence of reductant.. 90

Figure 3.5 SDS-PAGE of ^{C38}R1 protein under varied templating conditions. 91

Figure 3.6 Sedimentation velocity profiles of templated ^{C38}R1 and $^{C38/C96}$R1..... 94

Figure 3.7 Binding isotherm of Zn^{II} to Fura-2 in the presence of templated C38-bearing R1 variants. 96

Figure 3.8 Zn^{II} binding curves of PAR mixed with Zn-loaded $R1_4$ variants..... 102

Figure 3.9 Thermal denaturation of disulfide-linked R1 tetramers. 104

Figure 3.10 Representative thermogram of Zn^{II} binding to $^{C38/C96}R1_4$. 107

Figure 3.11 Assessment of disulfide breakage in $^{C38}R1_2$ by SDS-PAGE..... 109

Figure 3.12 Crystallization of metal-free disulfide-linked $^{C38/C96}$R1. 111

Figure 3.13 Crystal structure of Zn-$^{C38/C96}R1_4$. 115

Figure 4.1 Chemical structures of sulfenic acid sidechains and paramterization models. 128

Figure 4.2 Geometry optimization of sulfenic acid parameterization models. 130

Figure 4.3 Water placement or partial charge parameterization of Cso and Csa. 132

Figure 4.4 Torsion angle scan energy calculations and fits for Cso and Csa sidechains.136

Figure 4.5 Schematic representation of the cationic dummy atom (CADA) Zn^{II} model. 139

Figure 4.6 Solvating the internal of cavities in the $^{C38/C81/C96}R1_4$ structures. 146

Figure 4.7 Equilibration of internal water number in Zn-$^{C38/C81/C96}R1_4$ MD simulations. 147

Figure 4.8 Selecting internalized waters in fully-solvated $^{C38/C81/C96}R1_4$. 149

Figure 4.9 Metal-free states of $^{C38/C81/C96}R1_4$ simulated by all-atom molecular dynamics. 150

Figure 4.10 Topology and naming convention of key $^{C38/C81/C96}R1_4$ structural parameters. 152

Figure 4.11 Changes in $^{C38/C81/C96}R1_4$ structural parameters following simulated bond breakage. 155

Figure 4.12 Structural changes in $^{C38/C81/C96}R1_4$ during GaMD following a preparatory simulation. 159

Figure 4.13 Structural changes in $^{C38/C81/C96}R1_4$ during GaMD with preparatory phase omitted. 160

Figure 4.14 Structural changes in $^{C38/C81/C96}R1_4$ from constant-velocity pulling of a single helix pair. 162

Figure 4.15 Structural changes in $^{C38/C81/C96}R1_4$ from multi-parameter constant-velocity pulling. 163

Figure 4.16 Thermodynamic analysis of $^{C38/C81/C96}R1_4$ opening following C81-C81 breakage. 165

Figure 4.17 Thermodynamic analysis of $^{C38/C81/C96}R1_4$ opening following C96-C96 breakage. 167

Figure 4.18 Convergence of structural potental energy for $^{C38/C81/C96}R1_4$ simulations. .. 171

Figure 5.1 Non-reducing SDS-PAGE gels of $^{C74/C81/C96}$R1 templating reactions. 186

Figure 5.2 Crystal structure of $^{C74/C81/C96}R1_4$ bound to Zn^{II}. 187

Figure 5.3 Homology model of an R1 fusion construct..... 188

Figure 5.4 SV-AUC profile of RF_{15} in the presence and absence of Zn^{II}..... 189

LIST OF TABLES

Table 2.1 Thermodynamic parameters for Zn^{II} binding to disulfide-linked $R1_4$ tetramers. 53

Table 2.2 X-ray crystallography data collection and refinement statistics for $^{C38/C81/C96}R1_4$ structures. 59

Table 2.3 Dihedral strain energies of disulfide bonds in $^{C38/C81/C96}R1_4$ structures. 64

Table 2.4 Relative accessible surface areas of cysteine residues in $^{C38/C81/C96}R1_4$ 69

Table 3.1 C-α separations of residue 38 pairs in available crystal structures of R1 variants. 85

Table 3. 2 Speciation of templating reactions of $^{C38/C96}$R1 and $^{C38/C96}$M1. 89

Table 3.3 Relative accessibility of residue 38 in metal-free R1 and ^{C96}R1 structures and models. 93

Table 3.4 Dissociation constants of Zn^{II} binding to dislufide-linked ^{C38}R1 and $^{C38/C96}$R1. 98

Table 3.5 Average free energies of Zn^{II} binding to disulfide-linked R1 variants. 99

Table 3. 6 Pseudo first-order Zn^{II} release rates from disulfide-linked $R1_4$ tetramers. 101

Table 3.7 Apparent melting temperatures of disulfide-linked R1 tetramers. 105

Table 3.8 Binding enthalpies of disulfide-linked R1 tetramers 108

Table 3.9 X-ray crystallography data collection and refinement statistics for $^{C38/\ C96}$R1 structures. 112

Table 3.10 Dihedral strains of disulfide bonds in crystal structures of disulfide-linked R1 variants. 114

Table 4.1 CHARMM-type partial charges of sulfenic acid and chemically related residues. 133

Table 4. 2 Bond paramters for the sidechain atoms of the Cso and Csa residues. 134

Table 4. 3 Angle parameters for the sidechain atoms of the Cso and Csa residues..... 135

Table 4.4 Dihedral parameters for the sidechain atoms of the Cso and Csa residues. ... 137

Table 4.5 Lennard-Jones parameter sets for Zn^{II} ions. 141

Table 4.6 Averaged Zn-ligand distances from an unconstrained Zn-$^{C38/C81/C96}R1_4$ simulation..... 142

Table 4. 7 Harmonic constraints for Zn^{II} binding to $^{C38/C81/C96}R1_4$. 144

Table 4.8 Structural parameters of metal-free $^{C38/C81/C96}R1_4$..... 153

Table 4.9 GaMD boost parameters for simulations of $^{C38/C81/C96}R1_4$ opening. 158

Table 4.10 Average disulfide strain in simulations of metal-free $^{C38/C81/C96}R1_4$. 170

Table 4.11 Potential energy differences of key residues in metal-free $^{C38/C81/C96}R1_4$ opening..... 172

Table 5. 1 X-ray crystallography data collection and refinement statistics for $^{C74/C81/C96}R1_4$..... 190

Table 5.2 Mutageneic primers for constructing R1 fusion genes. 193

Table 5.3 Linking petides encoded in R1 fusion genes. 193

ACKNOWLEDGMENTS

First, I would like to thank my thesis advisor, Akif Tezcan, for agreeing to take me on in his lab and for his guidance throughout my thesis research. His unwavering passion for science and his boundless optimism in the face of difficulties at the bench have been truly inspirational. I owe my growth as a scientist, and as a person, in large part to his encouragement, his limitless capacity for new ideas, and his urging me to always be at my best.

In the Tezcan lab, I had the considerable privilege to learn from and work alongside an extraordinarily talented group of researchers. In particular, I owe a huge debt of gratitude to Annette Medina and Woon Ju Song, from whom I constantly sought advice when I was starting in the lab. More recently, I have learned a great deal from Rob Alberstein, without whom I could not have surmounted the many hurdles to getting started with computational chemistry. These and many other members of the lab, past and present, have made my graduate career feel less like a 9-to-5 job, and more like being part of a family. In particular, I thank Jake Bailey, whom I could count on to help blow off steam, Faith Katz, whom I could count on to discuss all the big things in life, and Rohit Subramanian, whom I could count on to indulge me in our shared obsessions.

I owe a great deal to my friends who helped me through the tough times in graduate school, and who shared in the good times. I had the great fortune of coming to UCSD at the same time as a very special group of friends, without whom being here would have been far less memorable. I thank Cameron, Cory, Melissa, Mike, Sarah, and

Vivi for picking me up when I fell, and for making San Diego feel like home. I must also thank Lauryn for her many long-distance phone calls, and for always knowing exactly what I needed to hear if I was feeling low.

Last, I thank my family, my parents Lisa and Jeff, and my brother Clark, for their love and support as I prepared to move across the country, and for their encouragement throughout my time in graduate school. They are a constant source of strength for me, and a constant reminder of the things that matter in life.

This dissertation was funded in part by UCSD's Molecular Biophysics Training Grant (NIH).

Chapter 1 was reproduced in part, with permission, from a manuscript currently being prepared for submission for publication: Churchfield, L.A.; Alberstein, R.G.; Tezcan, F.A. Methods for the design of protein complexes and biomaterials.

The dissertation author is primary author on all reprinted materials.

Chapter 2 was reproduced in part, with permission, from Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A. "*De novo* design of an allosteric metalloprotein." *J. Am. Chem. Soc.* **2016**, *138*, 13163–13166. Copyright 2016, American Chemical Society.

The dissertation author is primary author on all reprinted materials.

Chapters 3 and 4 were reproduced in part, with permission, from a manuscript currently being prepared for submission for publication: Churchfield, L.A.; Williamson, L.A.; Alberstein, R.G; Tezcan, F.A. Investigating the structural and mechanistic basis of allostery in a designed protein scaffold.

The dissertation author is primary author on all reprinted materials.

VITA

Education

2018 Doctor of Philosophy, University of California, San Diego

2014 Master of Science, University of California, San Diego

2011 Bachelor of Science, University of Minnesota, Twin Cities

Publications

Churchfield, L.A.; George, A.; Tezcan, F.A. "Repurposing proteins for new bioinorganic functions. *Essays Biochem.* **2017**, *61*, 245-258.

Bailey, J.B.; Subramanian, R.H.; Churchfield, L.A.; Tezcan, F.A. "Metal-directed design of supramolecular protein assemblies." *Methods Enzymol.* **2016,** *580,* 223-250**.**

Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A. "*De novo* design of an allosteric metalloprotein." *J. Am. Chem. Soc.* **2016**, *138*, 13163–13166.

Chao, F.A.; Morelli, A.; Haugner III, J.C.; Churchfield, L.; Hagmann, L.N.; Shi, L.; Masterson, L.R.; Sarangi, R.; Veglia, G.; Seelig, B. "Structure and dynamics of a primordial catalytic fold generated by *in vitro* evolution." *Nature Chemical Biology*, **2013**, *9*, 81-13.

Honors and Awards

2017 Distinguished Graduate Student Fellowship

2013-2015 Molecular Biophysics Training Grant (NIH)

2013 NSF Graduate Research Fellowship (Honorable Mention)

Fields of Study

Major Field: Chemistry

Studies in Biochemistry, Bioinorganic Chemistry, and Biophysics
Professor F. Akif Tezcan

ABSTRACT OF THE DISSERTATION

Design and Characterization of an Allosteric Metalloprotein Assembly

by

Lewis Anthony Churchfield

Doctor of Philosophy in Chemistry

University of California, San Diego, 2018

Professor F. Akif Tezcan, Chair

Proteins are one of the main building blocks of life. Among their numerous functions are roles as biocatalysts for essential chemical reactions, signaling agents that coordinate communication down to the sub-cellular level, and as modules for constructing cellular superstructures. In many cases, these functions rely on the ability of proteins to form stable and specific associative interactions with other proteins, as well as with metal ions. Drawing inspiration from nature, we deploy these functional features as design elements that aid in the construction of artificial metalloprotein assemblies. Using an iterative design approach, we have engineered the monomeric protein cytochrome cb_{562} by application orthogonal strategies: Metal Designed Protein Self-Assembly

(MDPSA), Metal-Templated Interface Redesign (MeTIR), and the installation of intermolecular disulfide bond crosslinks. This synergistic design approach afforded a self-assembling protein variant, $^{C38/C81/C96}$R1, which bears metal chelating groups, a designed dimerization interface (termed *i1*), and surface-exposed cysteine residues. In the presence of Zn^{II} this engineered variant self-assembled into a tetramer, Zn-$^{C38/C81/C96}$R1$_4$. Removal of Zn^{II} from the assembled tetramer resulted in hydrolysis of a single C38-C38 disulfide bond, leaving the five remaining crosslinks intact. Thus, $^{C38/C81/C96}$R1$_4$ is an allosteric protein, capable of remotely coupling Zn^{II} binding at is central binding sites to the breakage of a peripheral disulfide bond. In this work, we discuss the design of the $^{C38/C81/C96}$R1$_4$ protein as well as the demonstration of this allosteric behavior. We further carry out biochemical and biophysical characterization of this protein and related variants to determine the structural and energetic basis of this Zn-disulfide allostery. The C38-C38 disulfide bonds are embedded in the *i1*interface which, crucially, forms malleable protein-protein contacts. The flanking disulfide crosslinks of $^{C38/C81/C96}$R1$_4$ serve as structural conduits for coupling the pair of *i1* interfaces, which is critical to develop the requisite driving force to effect disulfide bond hydrolysis. In this designed system, it is the underlying malleability of the structure that gives rise to the allosteric behavior. We consider this as a successful application of synergistic design to give a coordinated protein function, and anticipate that adoption of similar design paradigms will greatly benefit ongoing protein engineering efforts in the community at large.

Chapter 1: A survey of designed protein assemblies

1.1 Foundations of protein design: drawing inspiration from Nature

Proteins are ubiquitous members of the cellular toolkit that fulfill a stunningly diverse array of functions: dynamic structural scaffolds which give cells their form and serve as intracellular highways, prodigious catalytic complexes that carry out the chemical reactions integral to life, and faithful messengers within signaling networks that orchestrate biological function, to name a few. A pool of 20 amino acids serve as the building blocks for the myriad three-dimensional folded structures that have been observed in Nature, structures which in turn enable this functional prowess. One critical aspect of this structure-function paradigm is the ability of many proteins, including as

many as 80% of *Escherichia coli* proteins,[1] to form complexes in carrying out their specific roles. A self-assembling protein presents practical benefits to the cell, such as lower energy cost and error rates in gene synthesis.[1] However, self-assembling proteins also possess functional advantages relative to their monomeric subunits, including allosteric modulation and cooperative function, enabled by emergent properties such as symmetry and long-range order.[1, 2] Notable examples of these principles at work can be found in protein assemblies that catalyze challenging multi-step reactions (*e.g.*, nitrogen fixation by nitrogenase), molecular machines (*e.g.*, gradient-coupled formation of adenoseine triphosphate, or ATP, by ATP-synthase) and responsive biomaterials (*e.g.*, dynamically assembling microtubule filaments) (**Figure 1.1**).[3, 4] Despite the impressive capabilities of these and other natural protein assemblies, all are limited by the constraints of evolutionary pressures attendant to operating as part of a biological system. One can, however, readily imagine removing such constraints to generate artificial protein assemblies with useful functional properties. This presents the grand challenge of protein design: the creation of novel functional protein complexes with properties that rival or surpass the efficiency, complexity, and fidelity of those found in Nature.

It is helpful to understand the opportunities and challenges facing protein design by comparing them with efforts to construct self-assembling DNA architectures. These efforts have given rise to the field of DNA nanotechnology, which generates exceedingly complex structures composed of a long single-stranded scaffold polynucleotide that adopts the desired configuration by base-pairing with small, so-called staple DNA strands.[5] The success of these designs rests on the simplicity and programmability of the Watson-Crick base pairing interactions that mediate robust sequences-specific assembly.

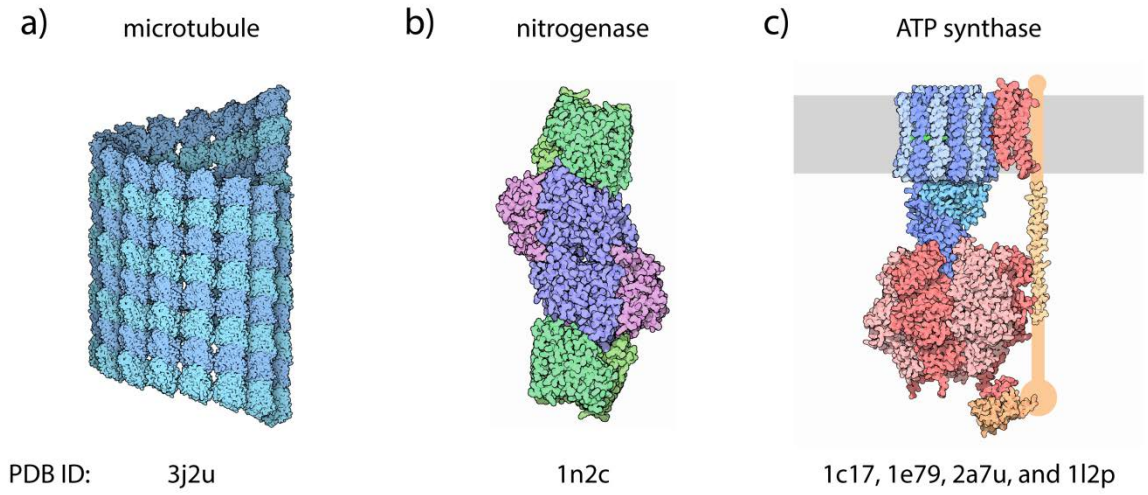

Figure 1.1 Notable examples of naturally occurring protein assemblies.
(a) Segment of a microtubule assembled from the dimeric protein tubulin. Microtubules are cytoskeletal filaments that dynamically assemble and disassemble to form key cellular structures, such as the mitotic spindle during cell division. (b) The octameric complex of nitrogenase consists of two Fe protein homodimers (depicted in green) and the heterotetrameric MoFe protein (depicted in blue and magenta). Fe protein supplies electrons to the MoFe protein, which contains the catalytic center that fixes N_2 into NH_3, making this essential element of life available for biological use. (c) The ATP synthase complex is an assembly of 12 polypeptide chains of seven distinct types. This molecular machine is responsible for converting ADP into ATP, a ubiquitious biological energy source. ATP synthase is embedded in a cellular membrane, and couples movement of H^+ down its concentration gradient to ATP formation by rotary catalysis. Images are reproduced with permission from PDB Molecule of the Month entries (microtubules - doi:10.2210/rcsb_pdb/mom_2014_7, nitrogenase - doi:10.2210/rcsb_pdb/ mom_2002_2, ATP Synthase – doi:10.2210/rcsb_pdb/mom_2005_12). Images are based on the indicated entries in the Protein Databank (rcsb.org)

The use of DNA origami to generate exquisite structures and molecular devices is a testament to the utility of these simple interacting rules.[5] However, the simplicity of the nucleotide alphabet limits the functional diversity all-DNA assemblies, and therefore are sometimes deployed as scaffolds for protein.[5] Proteins, by contrast, consist of building blocks with greater chemical diversity. However, this complexity has largely precluded the elucidation of robust general rules that govern protein assembly. Herein lays the promise and the challenge of protein design.

This dual nature of the chemical complexity underlying protein-protein interactions is true of their dynamics as well. Flexible proteins often adopt an ensemble of conformational states in carrying out their native function. This is now a widely recognized characteristic of proteins with implications that have shaped our understanding of allosteric protein function. Simple, structure-focused models originally proposed for allostery (Monod-Wyman-Changeaux model) and catalysis (lock-and-key model) have been supplanted by views that recognize protein dynamics as a crucial aspect of these and other functional properties.[6-8] Adoption of a particular structural arrangement is a key aspect of protein function, but this alone fails to account for the prowess and diversity displayed by natural proteins.

Clearly, the importance of structural dynamics as a driver of protein function naturally-occurring protein complexes must be true of their engineered counterparts as well. Let us briefly consider efforts to emulate the biological complexity through the lens of designing an artificial humanoid robot composed of multiple building blocks (head, torso, arms, etc.). One crucial aspect of successful design is link the building blocks with the desired connectivity. While necessary for constructing our artificial system, this is not

a sufficient condition for successful design if we consider the possibility of loose connections that limit any coordinated action of the linked components (**Figure 1.2**). Compensating for the shortcomings of this design, one may consider the use of tight well-fastened junctions to link the building blocks. While this approach furnishes an assembly that holds the desired shape and which is suitable to serve as a container or scaffold, the absence of any dynamic behavior limits the functional capabilities of such monolithic artificial systems (**Figure 1.2**). Instead, judicious deployment of both rigid and flexible interfaces acting in concert is required to enable sophisticated functions, such as coordinated movement. Additionally, the ideal interactions between building blocks allow for modular replacement, as well as for tailoring and functionalization of the artificial assembly, rather than having a high degree of interdependence that prohibits such modification (**Figure 1.2**). This simple analogy illustrates the importance of dynamics in natural biological systems, while also showing how systems with the correct structural properties may fall short of the functional breadth displayed by natural biological systems.

As discussed below, strategies for protein design must either confront the energetic idiosyncrasies of protein-protein interactions, or devise chemical and biochemical design strategies to circumvent this challenge. This has inspired protein engineers to develop a toolkit for creating novel self-assembling proteins beyond the boundaries of Nature's proteome. Broadly speaking, efforts to impart self-assembly can be viewed as either repurposing an existing protein interaction modality, or as efforts to

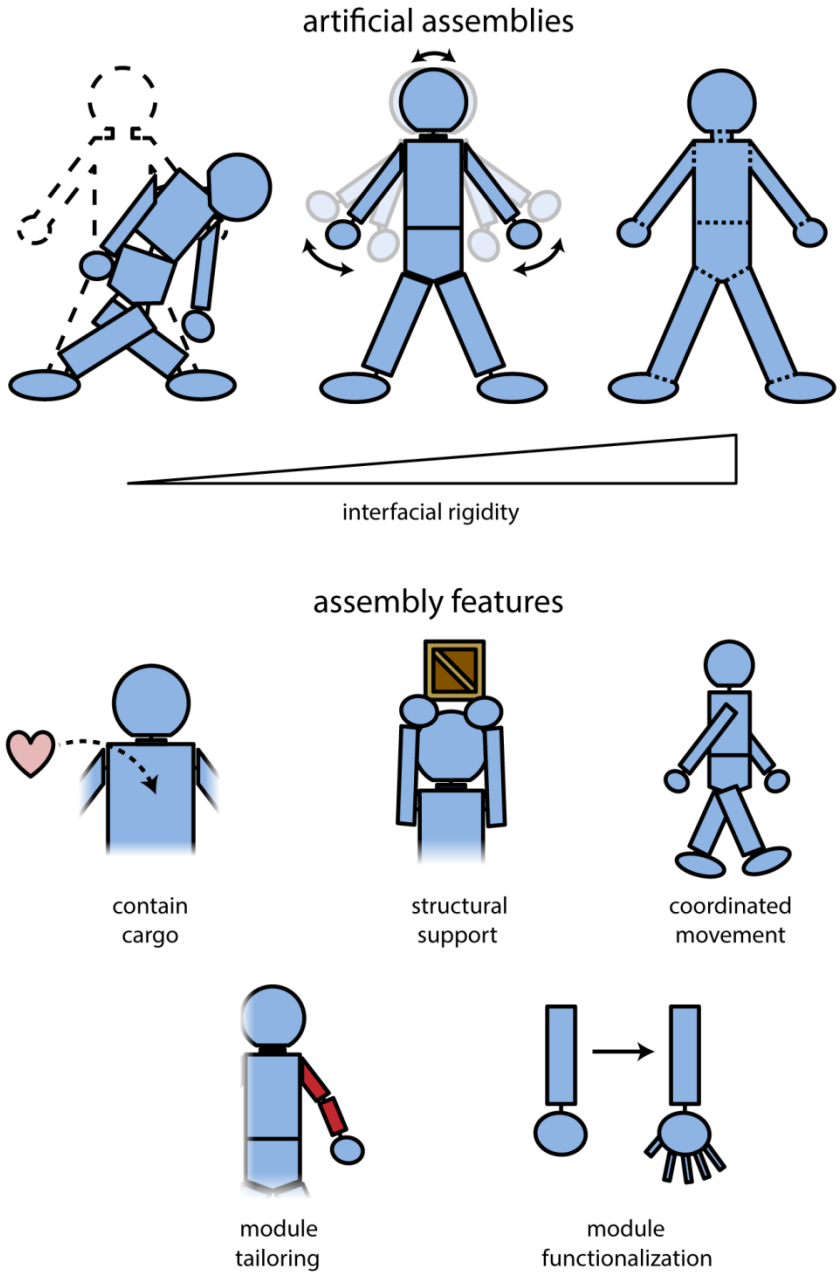

Figure 1.2 Cartoon depiction of the roles of flexibility and rigidity in an artificial biological assembly. Building block assembled in the desired configuration have different properties and capabilities depending on the nature of their interfaces. Overly loose connections and overly tight connections preclude many cooperative functions. Failure to achieve the desired structural configuration prohibits most, if not all, coordinated function. Adopting the desired structure allows for a structure to serve as a container or as a scaffold, but prohibits dynamic function and complicates design alteration. Structures with judicious placement of flexible interfaces with modest interdependence allow for coordinated movement, module tailoring, and module functionalization.

to impart self-assembly by installing new interaction motifs. Efforts to repurpose an existing interaction exploit established protein-ligand interactions, intramolecular protein contacts, or existing intermolecular protein-protein contacts (**Figure 1.3a-c**).[9-32] Installing new interaction motifs can be accomplished by appending a templating molecule, installing a metal chelating motif, using disulfide bond crosslinkers, or through the *de novo* design of a self-assembling protein surface (**Figure 1.3d-f**).[33-73]

An important predecessor to the field as it exists today was the demonstration that biologically-relevant natural protein complexes could be self-assembled *in vitro*. One of the earliest successes was the reconstitution of active Tobacco mosaic virus (TMV) from its purified coat protein and viral genome, laboriously isolated from natural sources.[74] The success of this experiment hinged on the inherent self-assembling properties of the capsid proteins to generate infection-competent viral particles *ex vivo*, an impressive feat considering the millennia of evolutionary optimization for *in vivo* assembly.

Efforts to engineer artificial supramolecular protein complexes also owe a great deal to work involving artificial peptides. A pioneering study by Ghadiri and co-workers used solid-phase peptide synthesis to introduce an abiologicial 2,2'-bipyridine metal-binding motif into a 15 residue peptide.[75] In the presence of Ni^{II}, the metal-binding motif templated the formation of a trimeric species with α-helical structure that formed a hydrophobic core, providing an early demonstration of using chemical first-principles design to generate biomolecular assemblies. Self-assembling peptide systems have continued to receive considerable attention as techniques for *de novo* design have advanced, recently culminating in the design of a self-assembling transmembrane Zn^{II}/H^{+}

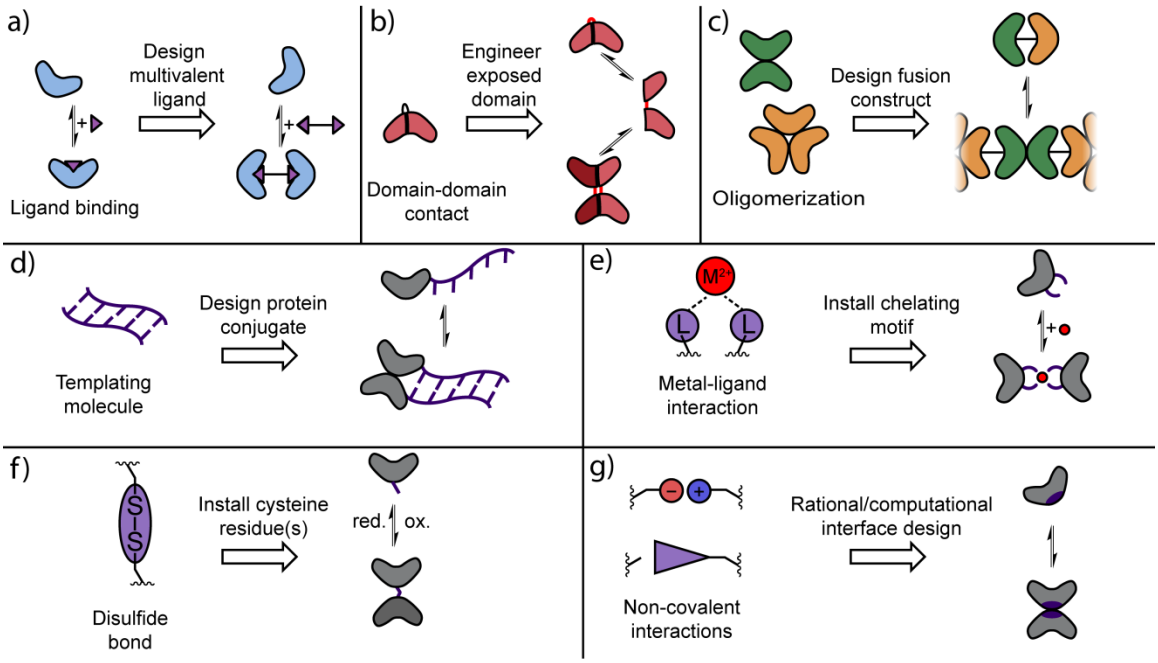

Figure 1.3 Survey of strategies for engineering protein assemblies.
(a) Protein-ligand interactions can be used to generate large protein assemblies through the design of multivalent ligands. (b) Engineering the loop region of a protein can alter the topology to redirect intramolecular domain-domain contacts to form intermolecular interactions via domain swapping. (c) Naturally-occurring oligomers fused together can form higher-order assemblies, including protein cages and fibers. (d) Appending a templating molecule, such as DNA, to a protein allows for the design of self-assembling protein conjugates. (e) Installing a chelating motif on a protein surface can be used to generate metal-templated self-assembling proteins. (f) Surface-exposed cysteine residues under oxidizing conditions form covalent disulfide bond crosslinks, which can be used for reversible oxidative self-assembly of proteins. (g) Rational or computational design strategies can alter protein surfacs to form favorable non-covalent interactions, such as Van der Waals contacts or electrostatic interactions, with steric complementarity to generate a self-assembling protein.

antiporter four-helix bundle[76] as well as the self-assembly of cages[77] and 2D arrays[78] from simple peptide building blocks. These and other advances in the field of designed peptide assemblies have recently been recently reviewed recently.[79, 80] The focus of the remainder of this chapter is strategies for the design of self-assembling proteins, or polypeptides of sufficient length to adopt a defined tertiary structure, as well as the emerging trends that are rising to meet the remaining challenges in protein engineering.

1.2 Repurposing native interactions for engineered protein assemblies

Natural protein-ligand complexes. Many proteins bind small molecule ligands as part of their biological function, and interaction that has been exploited to drive protein self-assembly. The synthesis of appropriate multivalent ligands that bear two or more binding moieties gives molecular bridges that bind multiple copies of their protein partner (**Figure 1.3a**).[9] One example comes from Freeman and coworkers, who used concanavalin A, a tetrameric protein that bears four sugar binding sites arranged in a tetrahedral fashion.[10] It was hypothesized that binding of each side of a C_2 symmetric bismannopyranoside would bridge the binding sites between tetramers to form 3D protein arrays, a result that was confirmed by transmission electron microscopy (**Figure 1.4a**).[10] This has proven to be a generalizable strategy, and has been exploited by several groups to generate protein assemblies. A commonly utilized motif is the high-affinity binding between streptavidin-, or Sav-tagged molecules, and derivatized biotin, including bisbiotin,[11, 12] biotinylated DNA molecules,[13] and biotinylated proteins.[14] The Wagner group has explored the use of drug molecules as the basis for generating assemblies of dihydrofolate reductase (DHFR) by synthesizing divalent versions a known inhibitor,

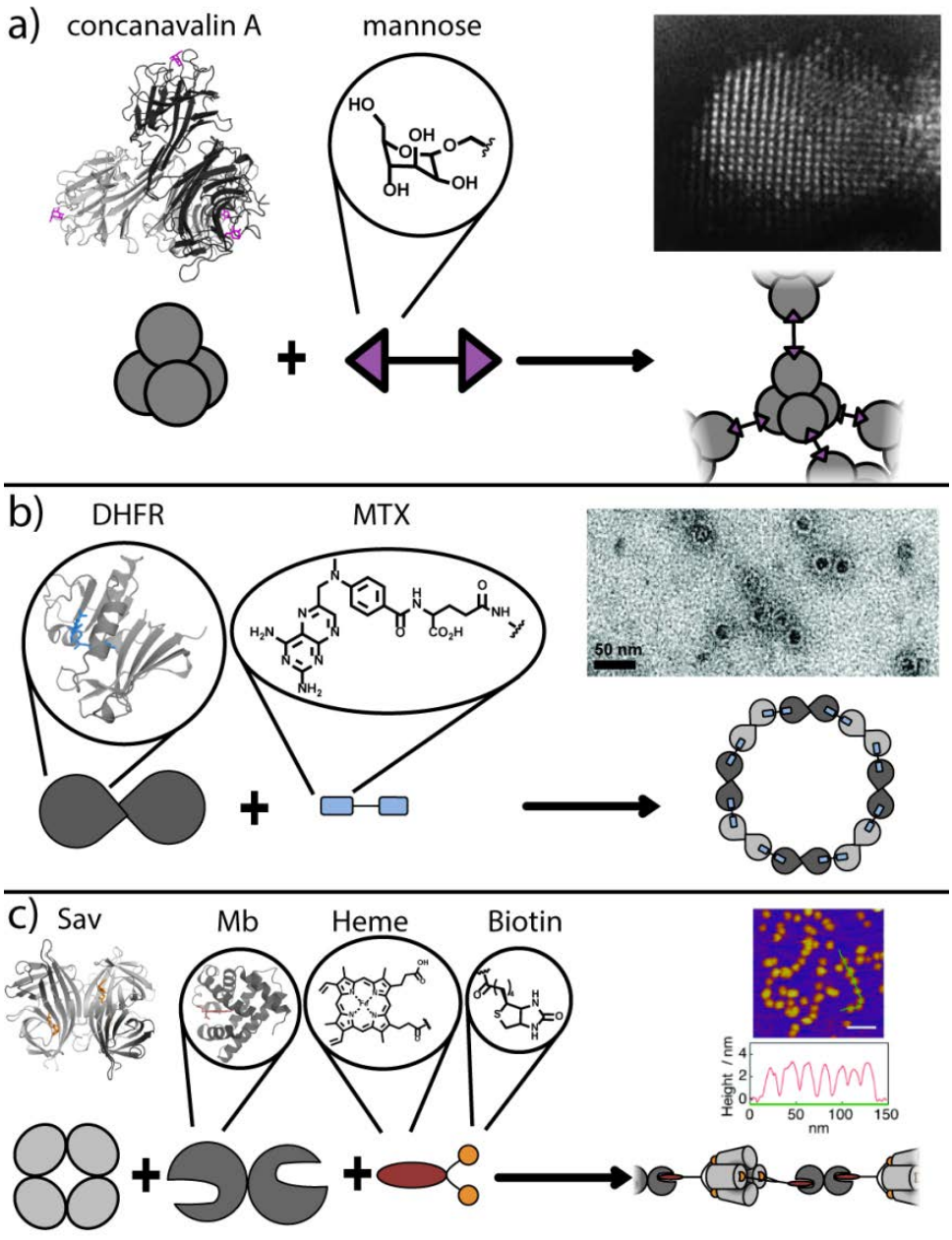

Figure 1.4 Design of multivalent ligands for inducing protein assembly.
(a) The homotetrameric protein concanavalin A (PDB: 5CNA) bears four binding sites for the monosaccharide mannose. Bis-mannose molecules can serve as a bridge between pairs of concanavalin A tetramers, which self-assemble into arrays observable by electron microscopy. Reproduced in part, with permission, from reference 10. (b) The natively-monomeric protein DHFR can be engineered to form dimers. Mixing this variant with a dimeric form of the drug MTX affords protein nanorings visible by electron micsroscopy. Reproduced in part, with permission, from reference 14. (c) The homotetrameric protein Sav, and an engineered dimer of Mb have been used to form two-component protein wires. A chimeric ligand with a single heme moiety and a pair of biotin moieties bridges members of these protein wires, as confrmed by atomic force microscopy. Reproduced in part, with permission, from reference 17.

methotrexate (MTX), to bridge protein dimers[15] and nanorings[16] (**Figure 1.4b**). Hayashi and coworkers have utilized triadic heme molecules as nucleation sites for cytochrome b_{562} (cyt b_{562}) assembly,[17] and synthesized chimeric heme-biotin linkers to generate two-component protein assemblies of alternating Sav and myoglobin (Mb) building blocks[18] (**Figure 1.4c**). As the above examples demonstrate, the underlying protein-ligand interaction can be used to form larger assemblies via synthesis of an appropriate molecular bridge, and this approach can be deployed without a need to modify the involved protein. However, this strategy is limited to proteins with known pre-existing interaction partners, and the interaction strength between proteins is wholly dictated by the binding constant of the ligand. This design approach may yield kinetically trapped assemblies if association is too tight, or limited assembly if association is too weak.

Domain swapping. As a consequence of the iterative process of molecular evolution, many proteins contain multiple "domains", which may fold and function independently from the rest of the polypeptide chain, providing modularity to the global protein structure. The relative arrangement of these domains is subject to changes in the protein environment and amino acid sequence, which may favor the formation of new intermolecular contacts by a process known as domain swapping. In Nature, this provides one mechanism for modulating protein function,[81, 82] as well as a route for the formation[83-85] and evolution[86] of oligomeric proteins. It has been suggested that almost any protein, under appropriate conditions, is capable of undergoing domain-swapping if the termini of the protein are sufficiently mobile.[87]

Drawing inspiration from these observations, protein engineers have used domain swapping to designing self-assembling protein systems. Such design efforts typically

focus on modifying or deleting part of a loop region to disfavor the intramolecular association of the flanking structural elements (**Figure 1.3b**). For example, deleting a six-residue loop from staphylococcal nuclease reoriented its C-terminal helix to favor the formation of a 2900 Å homodimeric interface (**Figure 1.5a**).[19] This approach was been applied to generate dimers and fibrils of loop-deleted 3-helix bundle proteins (**Figure 1.5b**),[20] as well as to generate a β-strand swapped homodimer.[21] Loh and coworkers employed a strategy, termed mutually exclusive folding, where a hinge protein is inserted into the loop region of another protein.[22] Hinge protein insertion disrupts the intra-molecular associations of the monomeric species, favoring the formation of domain-swapping (**Figure 1.5c**). The use of deactivated protein domains with this strategy provides a means of switching enzymatic activity on and off.[23] Domain-swapping can also occur in laboratory-evolved proteins, as observed by Hecht and co-workers when a patterned library of 4-helix bundle proteins unexpectedly yielded a domain-swapped dimer.[24] While this approach can readily be applied to virtually any protein, the resulting assemblies often form heterogenous assemblies and relatively weak contacts, limiting the utility of this method for achieving specific structures of interest.

Fusion of oligomeric proteins. As mentioned above, a significant fraction of proteins exist as oligomers in their functional state, and monomeric proteins which comprise these complexes (homo-oligomers in particular) have typically evolved to have symmetrically identical interfaces to achieve self-complementarity. Consequently, symmetry has become one of the most powerful tools available to both Nature and to protein engineers for generating protein assemblies. Just as C_2 symmetric small molecule ligands can link two protein binding partners, fusing a pair of protein monomers with

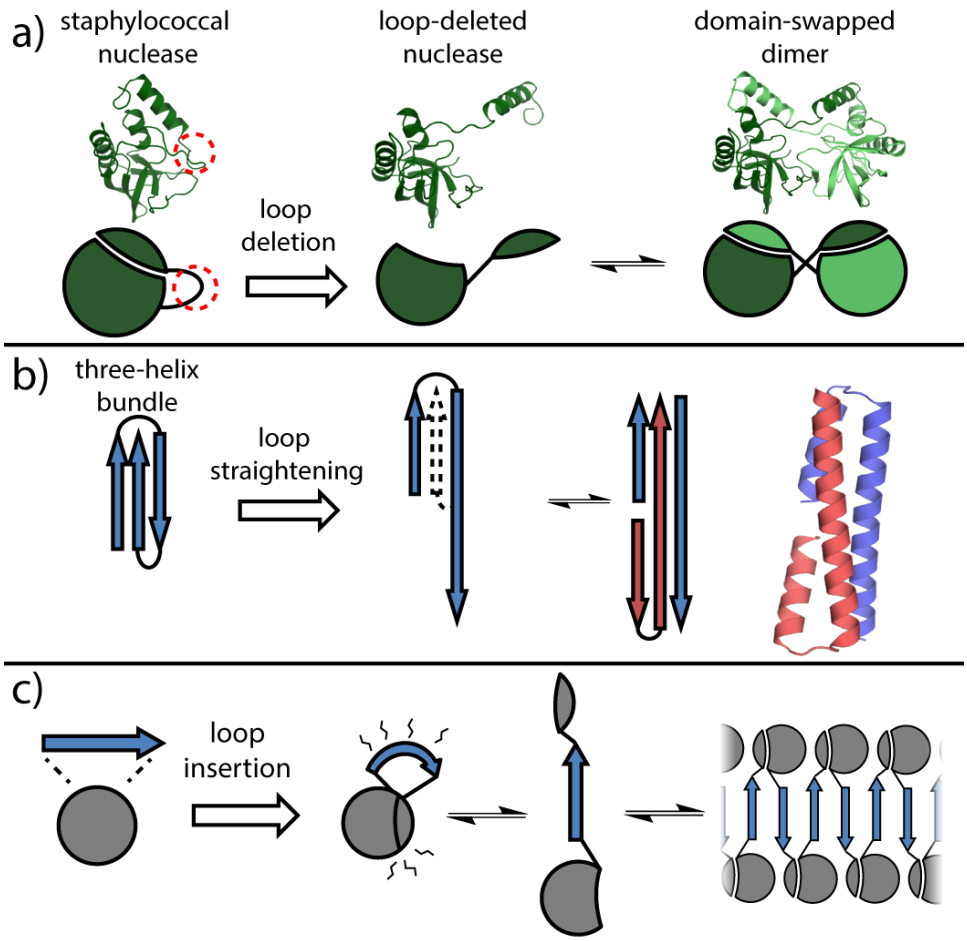

Figure 1.5 Engineering a homodimeric protein via domain swapping.
(a)The natively monomeric nuclease derived from *Staphylococcus aureas* possesses a C-terminal helix downstream of a short loop segment. Deletion of part of this loop affords a variant that undergoes dimerization via domain-swapping, with the C-terminal helix giving a monomer-like topology to the complementary domain of the binding partner. Reproduced in part, with permission, from reference 19. (b) Straightening the loop of an updown/up three-helix bundle protein alters the topology, resulting in a self-dimerizing variant. Reproduced in part, with permission, from reference 20. (c) Insertion of a protein in the loop region of a protein can strain the folding preferences of both proteins. The ensuing domain swapping was termed mutually exclusive folding. Reproduced in part, with permission, from reference 22.

compatible respective symmetries can give rise to extended crystalline arrays or discrete protein cages (**Figure 1.3c**). This is the strategy employed by Sinclair *et al.*, in which a fusion construct of D_2 and C_4 symmetric proteins with a shared rotational axis was observed to yield binary protein arrays.[25] The Yeates group has also used this approach to generate nanometer-scale cages by fusing dimer- and trimer-forming protomers in the appropriate relative orientation.[26, 27] The observed 15 nm spheres were consistent with the formation of the desired homododecameric cages, with trimers as the four vertices of a tetrahedron and dimers creating each of the six edges (**Figure 1.6**).[26] Strikingly, when two residues flanking the helical linker were mutated to mitigate potential unfavorable contacts (Q24V and K118A), monodisperse assembly was observed, allowing for the determination of the cage structure by x-ray crystallography.[27] While the general geometric principals behind the design of such constructs are well-understood,[28, 29] the success of these designs depends heavily on protein-specific interfaces that cannot, by their very nature, be generalized, as well as on local sequence effects that can be difficult to account for *a priori*.

In contrast with the above examples, supramolecular assemblies can also be generated through the fusion protein building blocks using flexible linkages. In one example, the Marsh group fused each member of a heterodimeric coiled-coil peptide to a homotrimeric protein via a soluble linker. Trimers bearing each binding partner, upon mixing, self-assemble into cylinders formed from two trimers, tetrahedral cages containing four trimers, and octahedral cages containing eight trimers.[30] In a similar study, the *de novo* designed WA20 peptide dimer was linked to the fibritin trimerization

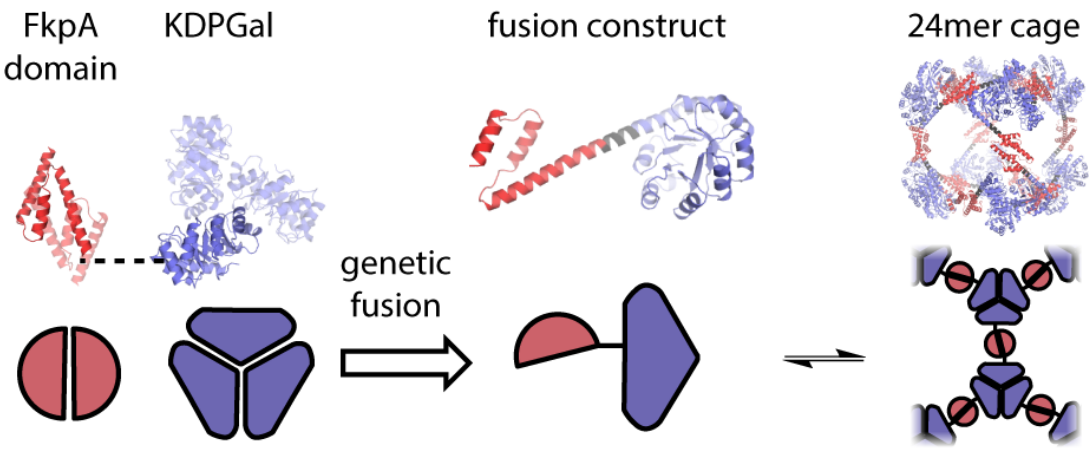

Figure 1.6 Design of a self-assembling protein cage by genetic fusion of oligomers.
The FkpA protein contains a dimerizing domain, which was genetically fused to the trimeric KDPGal protein. Use of a rigid helical linker allowed for control over the relative orientation of the two domains within the fusion construct. The fusion satisfies the oligomerization preferences of both domains by self-assembling into a 24mer protein cage. Adapted from reference 27.

domain, foldon.[31] Here, Arai and coworkers observed the formation of discrete assemblies with masses and solution-phase behavior consistent with a compact barrel and a tetrahedral cage, in addition to larger cage-like structures.[24] The flexible-design approach was further extended with higher-symmetry elements comprising a tetrahedral cage: a C_3-symmetric trimer at each face and a C_4- symmetric coiled-coil tetramer at each vertex [32]. In the design of such cages, the authors noted that the geometry of their fusion was relatively unconstrained, and focused on optimizing the length of the flexible peptide linker. Given the modularity of the approach and the availability of stimulus-responsive coiled-coil peptides, a proposed extension for this design strategy is the generation of cages whose assembly/disassembly is coupled the biochemical environment.

1.3 Engineering protein assemblies using de novo designed interactions

Protein assembly via molecular templating. There are numerous examples of non-protein species that form supramolecular complexes, ranging from pairwise interactions to larger aggregates, which one can envision as templates for driving protein assembly. Protein surfaces are heterogeneous and decorated with many functional groups of similar chemical properties, which present a challenge to protein engineers seeking to design self-assembling proteins. Conjugating a protein to a templating molecule can endow the protein with assembly properties, such as binding specificity or stimulus responsiveness, that are challenging properties to design in from scratch (**Figure 1.3d**).

Many small-molecule systems form stoichiometric interactions, which can be envisioned for use in directing the assembly of proteins with properties of interest. Host-guest binding partners that form specific binary supramolecular complexes, such as the β-

cyclodextrin/lithocholic acid pair, can be grafted to proteins and drive dimerization, including inside a living cell.[33, 34] This strategy has also been applied to methylviologen and naphthalene protein conjugates, which form a ternary complex with cucurbit[8]uril.[35] The cucurbit[8]uril host can also form a ternary complex with a pair of FGG tripeptides, which when appended to the N-terminus of a protein, can be used to template the formation of oligomers,[36, 37] nanowires of dimeric protein,[38] and contractible nano-springs.[39] DNA has also been employed as a molecular template, due to the sequence-specificity of base pairing that can be used to program assembly. The Mirkin group demonstrated the construction of hybrid materials by densely functionalizing the surfaces of catalase with ~12-15 DNA molecules (**Figure 1.7**).[40] These DNA molecules were designed to hybridize with a complementary sequence and leave an unpaired sticky end.[40] Generating complementary sticky ends drives the formation of self-assembly of two-component lattices containing a single enzyme, a pair of enzymes, or proteins and nanoparticles, and which contained active enzyme (**Figure 1.7**).[40] The choice of conjugation chemistry provided a means of changing the surface distribution of the associated DNA molecules, which could in turn alter the lattice parameters of the assembly.[88] Due to the universality of DNA base-pairing rules, this approach can be extended to other protein systems to generate lattices with a chemical and structural diversity beyond the scope of DNA-based materials.

Metal-mediated self-assembly. Metal ions are ubiquitous in protein biochemistry, and are some of the most common cofactors within natural protein assemblies. In fact, roughly 5% of structurally-characterized oligomers contain a metal ion or metallocofactor that serves a structural or functional role at an interfacial site.[71] Numerous metal coordin-

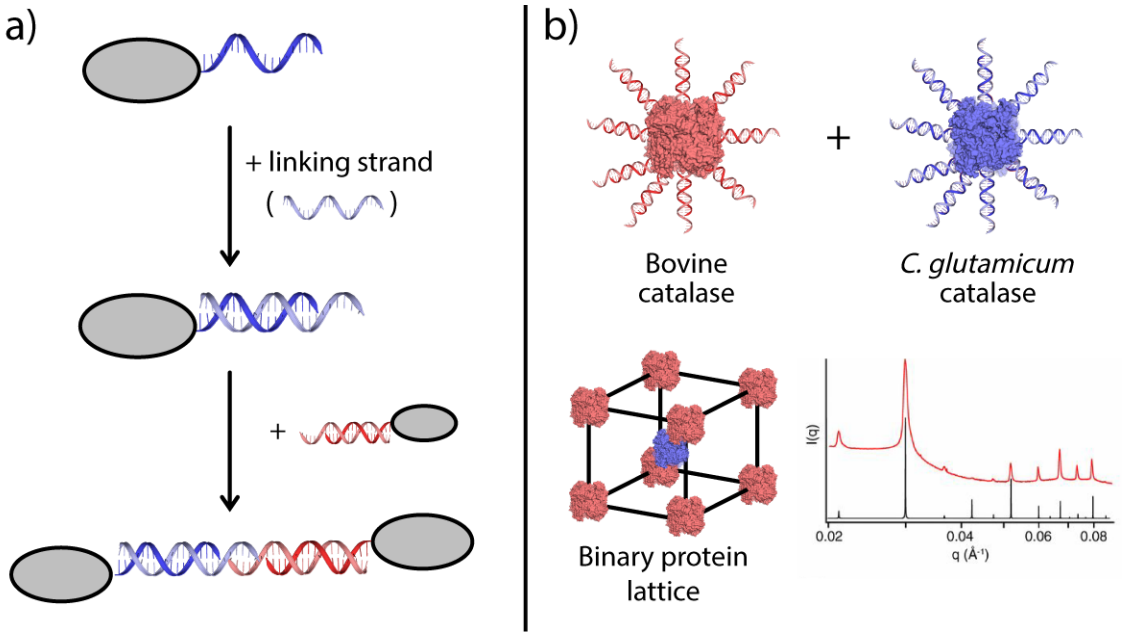

Figure 1.7 Design of two-component protein-DNA lattices.
(a) Single-stranded DNA molecules can be covalently appended to a protein surface, typically 12-15 (one shown for clarity). These oligonucleotides base pair with an added linking strand to form a stiff DNA duplex with a short sticky end. A second component bearing the complementary sticky end can undergo base pairing to link the proteins. (b) Two types of DNA-functionalized catalase were used to generate a two-component protein lattice, as confirmed by small angle x-ray scattering. Reproduced in part, with permission, from reference 40.

ation motifs have been characterized in proteins, typically involving chelating sites with multiple carboxylate- or sulfur-containing residues with histidine sidechains. If one considers the entire protein as a large ligand for metal coordination chemistry, it becomes clear that the strength, reversibility, and intrinsic directionality[89, 90] of these interactions are highly advantageous for self-assembly strategies. Nature has exploited this fact, as there is phylogenetic evidence that ancestral metal binding proteins provide crucial footholds in the evolution of some extant oligomeric metalloproteins.[91-97]

These principles have been successfully applied to protein design through a strategy termed Metal Designed Protein Self-Assembly (MDPSA).[41, 42] The application of MDPSA to a soluble, monomeric protein of choice is as follows: sets of surface-exposed chelating residues are appropriately installed such that binding of a free metal ion promotes the association of two or more monomers in order to satisfy its preferred coordination number and geometry (**Figure 1.3e**).[89, 90, 98] As a proof of principle, MDPSA was applied to an engineered version of cyt b_{562} which contains c-type heme linkages (cyt cb_{562}).[99] A pair of bis-His motifs were installed on helix 3 of the protein to yield a protein variant, Metal Binding Protein Complex One (M1), that undergoes Zn^{II}-induced assembly into a stable homotetramer,[41] and which could be readily tuned by modifying the coordination environment[43] or by using metal ions with different coordination preferences (**Figure 1.8a**).[44] This design logic can be applied in reverse to transform a self-assembling natural protein oligomer into an inducible assembly that requires the addition of metal.[45] Several native interfacial contacts of the 24-mer cage protein maxiferritin were removed, and metal coordinating residues were installed in their stead,

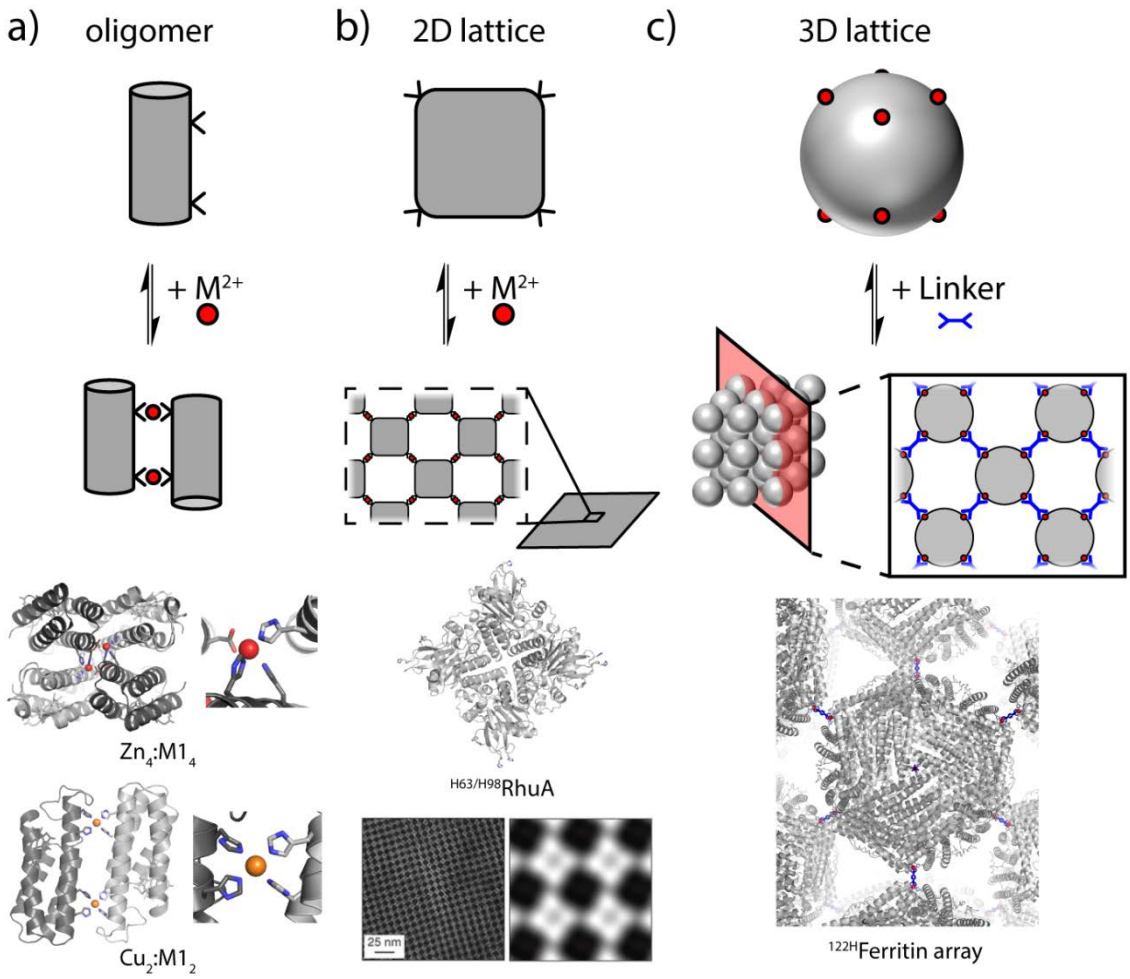

Figure 1.8 Designed protein scaffolds for metal-templated assembly.
(a) A natively-monomeric protein bearing a pair of chelating motifs on a single surface of the protein can bind metal ions. Multiple proteins can associate to satisfy the coordination preferences of the metal, giving a metal-templated oligomeric protein. In the case of the cyt cb_{562} variant M1, addition of Zn^{II} affords a tetramer, while addition of Cu^{II} forms a dimer. Adapted from reference 44. (b) The homotetrameric protein RhuAforms two-diemsional sheets upon binding Ni^{II} or Zn^{II}. Adapted from reference 48. (c) Spherical ferritin cages bearing tris-His metal-binding sites coordinate metals with open sites to bind linking mollecules and form three-dimensional arrays. Adapted from reference 49.

which yielded a soluble monomer in solution. This construct could then be reassembled into the native cage structure upon the addition of copper, which templated the reformation of the native interface via metal coordination.[45] The general chemical principles underlying MDPSA are not limited to discrete assemblies, and have been applied to the design of pseudo-infinite two-dimensional protein arrays.[46-49, 72] Installing metal-binding sites onto symmetric oligomeric protein scaffolds will yield multiple interaction sites capable of array formation. The C_4-symmetric homotetramer L-rhamnulose-1-phosphate aldolase (RhuA) was mutated to have bis-His clamps at its four corners, giving a variant ($^{H63/H98}$RhuA) that forms tessellated crystalline 2D lattices upon binding Ni^{II} or Zn^{II} ions (**Figure 1.8b**).[48] The *O*-symmetric homododecameric maxiferritin can, with a single mutation, afford a variant (^{H122}ferritin) bearing tris-His coordination sites at all eight of its three-fold symmetry axes, roughly corresponding to the eight corners of a cube (**Figure 1.8c**).[49] ^{H122}Ferritin cages preloaded with metal ions at each of these sites serve as giant metal clusters, from which crystalline 3D protein metal-organic-frameworks (MOFs) may be synthesized by addition of ditopic molecule linkers to form bridges between pairs of protein-associated metal ions (**Figure 1.8c**).[49] Similar to traditional MOFs, this system is highly modular, which allowed a library of ferritin-MOFs with cubic or tetragonal symmetry to be constructed by varying the identity of the metal ion or bridging ligand.[50]

Protein assembly with disulfide bond crosslinks. Pairs of thiol-containing cysteine residues can react to form disulfide bonds, which serve structurally important roles as inter- and intra-molecular covalent tethers,[100] including the linkage of antibody domains,[101] and mechanical stabilization of the protein domains found in collagen

fibers.[102] From a synthetic standpoint, disulfide bonds are particularly attractive for their strength, reversibility, and facile installation into a protein scaffold via single point mutation to include as cysteine residue. Disulfide bonds can, therefore, be exploited to crosslink proteins via oxidative self-assembly with minimal protein modification (**Figure 1.3f**).

As with other strong directional interactions, the structure of the final disulfide-mediated self-assembled product is largely dictated by the relationship between cysteine position and protein geometry. One demonstration of this comes from Mougous and co-workers, who introduced a pair of cysteine residues onto a hexameric protein such that each face of the ring-shaped complex displayed six crosslinkable residues (**Figure 1.9**).[51] The location of the thiols restricted assembly to 1D stacked nanorods, which were meant to mimic the arrangement observed in the crystal structure of wild-type Hcp1. Assembly was terminated by the introduction of capping variants with a single cysteine-containing face, and the resulting nanorods were observable by transmission electron microscopy (TEM). The success of this design relied on the fortuitous preferential stacking of this protein *in crystallo*. Templated formation of disulfide bonds in the absence of this bias presents a challenge to the use of disulfide bonds in other protein systems. Additionally, the use of multiple cysteine crosslinks in a monomeric protein remains challenging, but has been achieved in combination with other engineering approaches.[52] Indeed, given their small footprint, disulfide bonds are particularly well-suited as an element of synergistic design.

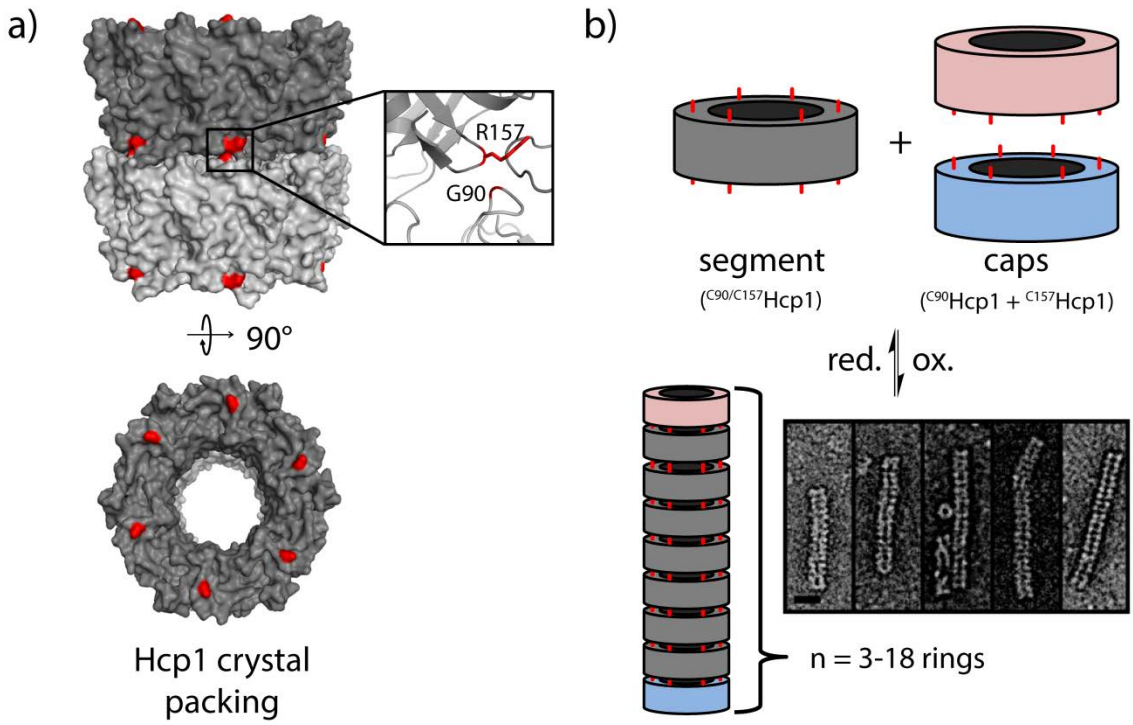

Figure 1.9 Design of disulfide-linked protein nanorods.
(a) The homohexameric protein Hcp1 exhibits a ring-like structure. The available crystal structure (PDB ID: 12Y12) shows a stacking interaction of the rings that positions residues Arg157 and Gly90 in close proximity. (b) Cysteine residues installed on both faces of the ring give a segment variant ($^{C90/C157}$Hcp1), while mutations on a single face give two capping variants (^{C90}Hcp1 and ^{C157}Hcp1). Shown are electron micrographs of nanotubes of varying length assembled by oxidative self-assembly *in vitro*. Reproduced in part, with permission, from reference 51.

Rational and computational design of protein-protein contacts. Protein-protein interfaces in natural complexes are typically composed of numerous weak interactions dispersed across relatively large surface areas.[103-105] Consequently, the design of a new protein-protein interaction often requires the simultaneous consideration of numerous participating residues, which presents a significant challenge to the protein engineer (**Figure 1.3g**). One strategy to mitigate this complication is to use symmetrical proteins, such that a single mutation is multiply propagated in a well-defined manner.[53, 54] Schulz and co-workers applied chemical intuition and structural insight in a design effort that incorporated multiple oligomeric proteins with cyclic symmetries (**Figure 1.10a**).[53] Installing hydrophobic residues at a single oligomeric surface afforded larger assemblies, with symmetric proteins generally requiring fewer mutations to form larger oligomers (**Figure 1.10a**). Recently, this principle has been applied to generate fibrillar protein assemblies *in vivo* from building blocks with dihedral symmetries.[54] Through analyzing the structures of nearly 2000 oligomeric complexes, they identified a trend for surface residues in positions that were the most at risk to permit self-association to contain residues with lower interaction propensities, suggesting that many protein complexes are poised to undergo assembly in this fashion if these protective residues are removed.[54] Additionally, protein engineers have formed patterned assemblies using electrostatic interactions, including between pairs of different proteins with different charge envelopes[55] and variants of a single protein engineered to have opposite charges,[56] illustrating that other weak intermolecular forces can drive self-assembly.

Computationally-guided remodeling and de novo design of protein interfaces. Recent advances in the computationally-guided modeling of protein structures now

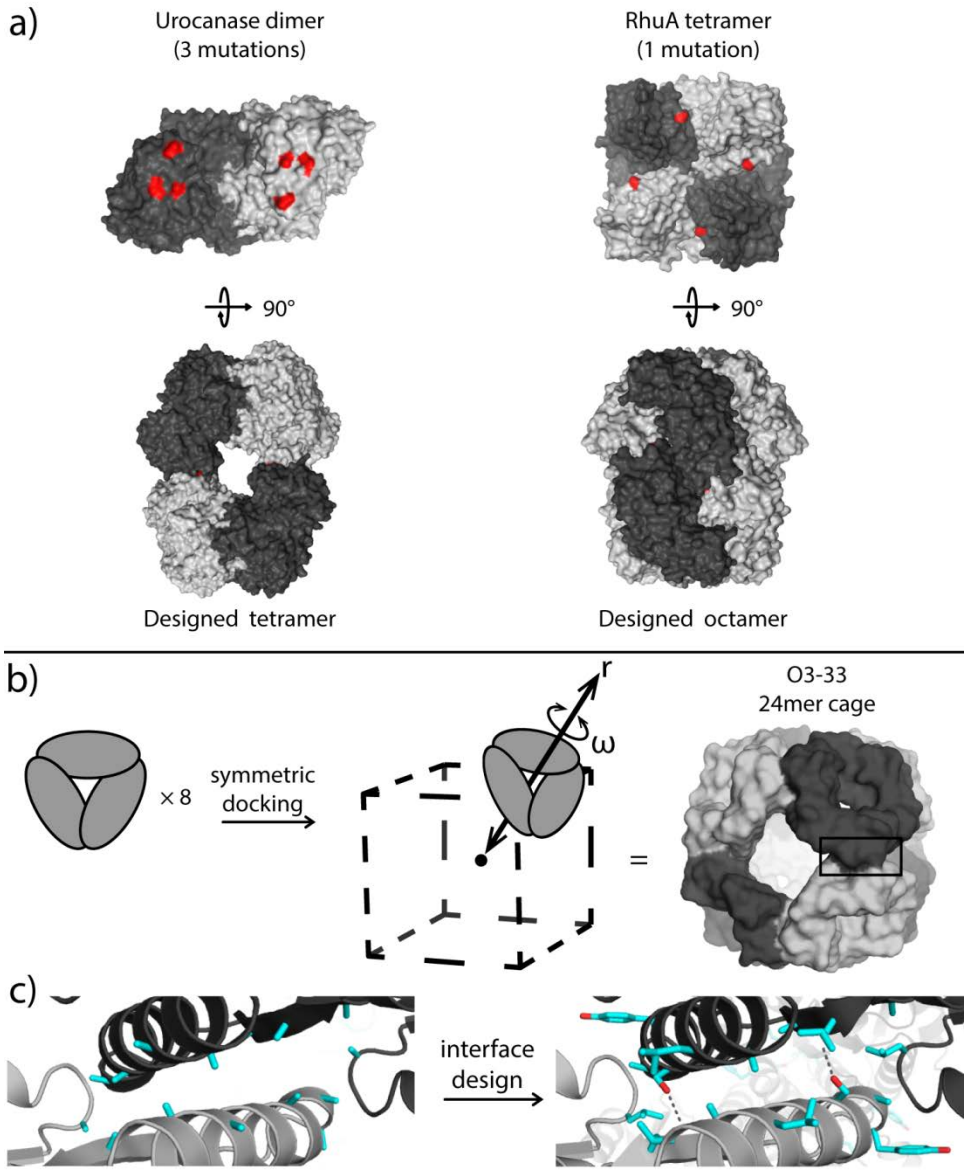

Figure 1.10 Designing protein-protein contacts for engineered supramolecular assemblies.
(a) Urocanase forms a C_2-symmetric dimer, while RhuA forms a C_4-symmetric tetramer. Hydrophobic mutations lie on a single surface (with respect to the rotational axis of symmetry) act in concert to form a larger binding surface. In the case of urocanase, three mutattions were sufficient to mediate formation of a tetramer. For the higher-symmetry RhuA, a single point mutations mediated formation of an octameric species. Adapted from reference 53. (b) Computational design efforts can exploit protein symmetry through the use of symmetry-nforced docking protocols. Docking of a eight copies of a homotrimer with respect to distance a distance (r) and an angle (ω) about a shared central point furnished a dodecameric cage, O3-33. (c) Candidate mutations are grafted onto the backbone of docked constructs to install favorable hydrophobic contacts, as well as hydrogen bond interactions (shown as black dashed lines). Adapted from reference 64.

enable the *de novo* design of new protein assemblies. A typical computational design workflow will start by screening databases of structurally characterized proteins for an appropriate starting construct. The scaffold will typically be chosen to contain structural features that sterically permit the formation of the interaction of interest, but lack favorable contacts that impart a propensity to self-assemble. From here, the desired interaction region is defined, and an ensemble of plausible mutations is generated for testing. For each member of this ensemble, candidate scaffolds are docked, and interfaces are energetically scored by scanning possible rotamers of the candidate mutations,[57, 58] giving rise to designs that bear optimized interfacial contacts. A number of software packages exist to aid in computational protein design efforts.[106] As with any modelling process, there is necessarily a tradeoff between accuracy and efficiency of the modeled systems, and the results of such design efforts must therefore be verified experimentally. While the algorithms involved in *de novo* protein design have improved dramatically in the past decade and have achieved success,[59-61] computational design remains challenging, and sometimes requires supplementary optimization with directed evolution techniques.[62, 63]

In addition to improved force fields for the energy evaluation, computational design efforts have benefitted from design principles employed elsewhere in rational protein design efforts. For example, the use of symmetric building blocks has been incorporated into the computational toolbox in the form of a symmetric protein docking protocol in the Rosetta software (**Figure 1.10b-c**).[64] This symmetric docking approach has led to the successful computational redesign of oligomeric proteins to yield crystalline 2D protein arrays that form *in vivo* [65], and the design discrete 3D cages over 10 nm in diameter.[107] As described below, this approach has been extended to the design

of large two-component cages.[66, 67] Computational design can be expected to improve as increasingly sophisticated and accurate design algorithms are available, and as the computing power continues to increase. An outstanding challenge in computational protein design is the formation of interfaces which are not only energetically favorable, but also flexible enough to permit dynamic protein motions central to the functions of natural proteins. As outlined in **Figure 1.2**, achieving a particular structural arrangement is not the sole prerequisite for obtaining a functional protein scaffold.

1.4 Synergy of design strategies for native-like engineered protein assemblies

Keeping in mind that the ultimate goal of protein engineering is the design of novel proteins with properties that rival their natural counterparts, it is prudent to recognize that all of the protein self-assembly strategies reviewed here have their own specific advantages and disadvantages. Installing interaction motifs (templating molecules, chelating groups, disulfide bonds) is effective, chemically intuitive, and often present a small design footprint. However, these approaches fail to capture the intricacies of the protein-protein interactions that enable the function of the vast majority of cellular proteins while remaining enigmatic. Computational *de novo* design is growing ever closer to the reliable prediction of extended noncovalent interfaces, but the successes of finding the global energy minimum may create rigid scaffolds, rather than dynamic machines. Consequently, progress in the field will undoubtedly benefit from the simultaneous application of both rational and computational design strategies.

MDPSA provides a means for the rapid design of metalloprotein assemblies via well-defined interactions, but does not prescribe further engineering beyond the introduc-

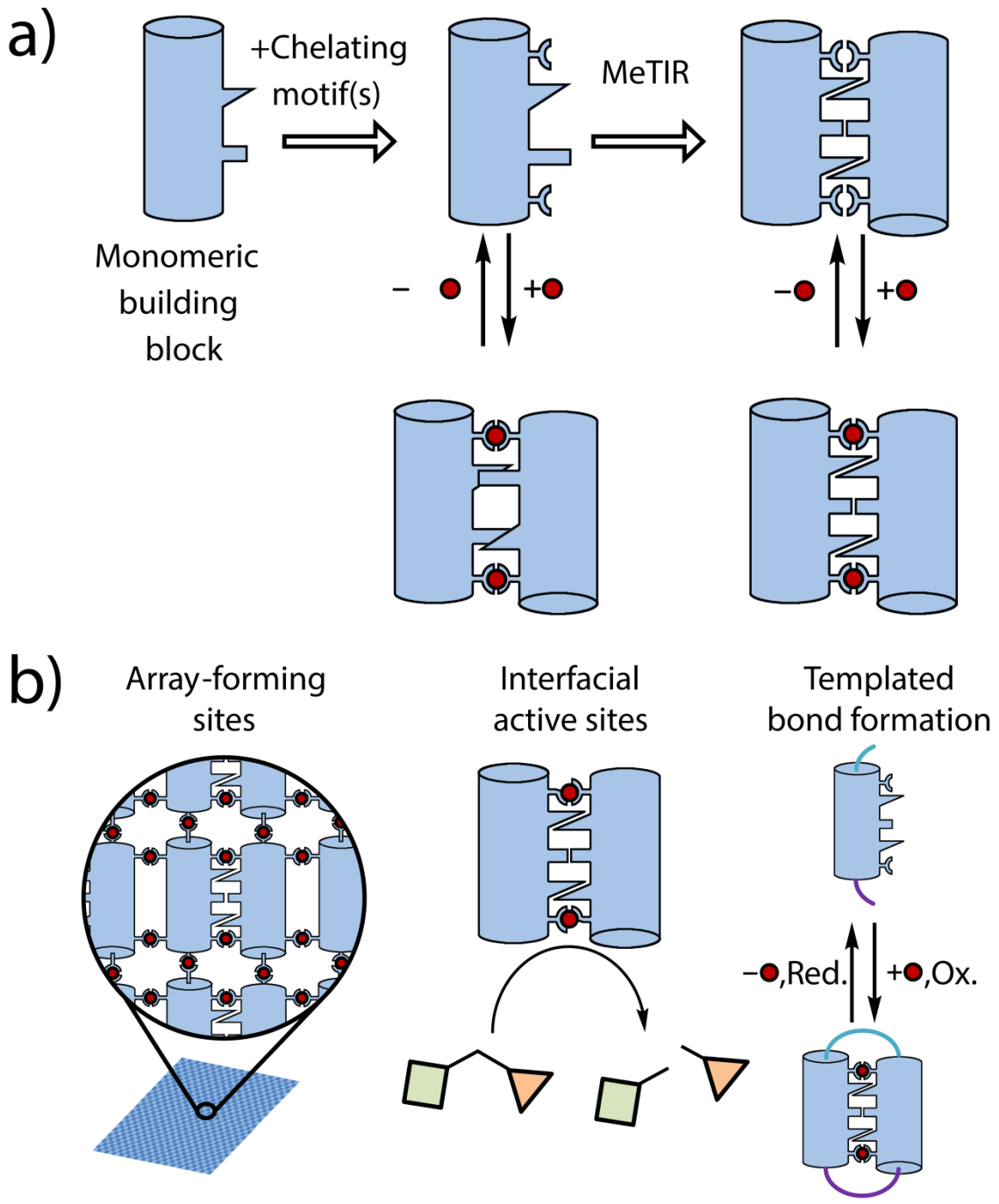

Figure 1.11 Metal tempalted interface redesign (MeTIR): a versatile strategy for protein engineering. (a) The exterior of a monomeric protein can be modified by installing chelating motifs that mediate metal-directed self-assembly, a process known as MDPSA. A second engineering step, the Metal-Templated Interface Redesign (MeTIR) strategy, generates self-assembling protein surfaces. This is done by imparting the nascent protein-protein interface with favorable amino acid contacts. (b) MeTIR can be used to generate infinite protein arrays, discrete assemblies with enzymatic activity, or to generate scaffolds that can selectively template bond formation. Such assemblies can exhibit emergetn structural and functional properites. Reproduced in part, with permission, from reference 68.

tion of metal coordination motifs. Many engineering efforts focus on installing or repurposing a metal binding site within a well-folded protein domain.[68] While the primary coordination sphere of a bound metal ion is itself a useful structural feature, it can also serve as an initial foothold for the installation of engineered protein-protein contacts at the nascent interfaces of the metalloprotein protein assembly (**Figure 1.11a**).[68] Using a metal-templated complex to aid interface design, rather than starting wholly *de novo*, tremendously reduces the complexity of the designed system, and can furnish a protein that robustly self-assembles in the absence of metal. This two-step design approach, termed Metal-Templated Interface Redesign (MeTIR), has been implemented by applying computational design methods to an engineered metalloprotein complex.[69] Applying MeTIR to the metal-binding cyt cb_{562} variant M1 generated a self-dimerizing variant, Rosetta Interfaced Designed Complex 1 (R1), that maintains the underlying ability to bind Zn^{II} to form a tetramer.[70] Metalloprotein assemblies designed in this manner provide scaffolds which can be utilized in further engineering efforts, including the use of open metal coordination sites to form protein arrays,[46, 47] the installation of interfacial active sites to give metalloenzyme complexes,[73, 108] and to furnish robust assembly that can be used to template the selective formation of distinct sets of disulfide bonds[52] from a single, simple parent building block (**Figure 1.11b**). The resultant engineered assemblies can possess emergent functional properties beyond the scope of their monomeric precursors, such as materials that withstand harsh conditions or possess high tensile strength,[46, 72] as well as enzymes with catalytic sites located at a protein-protein interface.[108] These and other emergent functional properties displayed by systems arrived at through MeTIR are discussed below, and in subsequent chapters.

1.5 Challenges and opportunities: emerging design paradigms protein self-assembly

The feats accomplished by natural systems are a source of inspiration to protein engineers. Natural protein assemblies frequently rely on the coordinated action of multiple protein components through the formation of heteromeric assemblies, which can adopt distinct functional conformations, in order to carry out their important biological functions. These principles are beautifully illustrated by the molecular machine F_OF_1 ATP-synthase (**Figure 1.1**). The membrane-embedded portion of ATP-synthase contains a homododecameric module through which protons are shuttled, and forms a complex with 12 additional subunits of seven distinct types. Proton shuttling drives the rotation of other portions of the enzymatic complex, and this rotary action provides the requisite driving force for the all-important reaction of ATP generation.[109] Moreover, the assembly of this protein complex is accomplished in a crowded cellular environment surrounded by many other proteins, and in a very specific cellular location. The amazing and as-yet unrivaled functional prowess also demonstrates the challenges facing protein engineers, as well as the opportunities for further advancement in the field. With this in mind, we briefly touch on several emerging topics in the area of designed protein assemblies, and the opportunities for further innovations that still remain.

Designed heterogeneous contacts in multi-component assemblies

Many natural protein complexes exist as heteromeric assemblies that require specific associations of distinct classes of protein components. These contacts have evolved to associate specifically with the interaction partner of interest, and must do so with high

fidelity and the appropriate level of affinity, in order to avoid undesirable self-associative or non-specific interactions. In engineered systems, the Baker and Yeates groups have successfully applied computational design to generate two-component protein cages.[66, 67] A designed interface between two different trimeric building blocks, or between a tetramer and a dimer, giving designed self-assembling protein pairs that form hetero-dodecameric cages.[66] This has been further extended to the formation of megadalton-sized cages, where designed pairwise contacts among pentameric, trimeric and dimeric building blocks yield cages with 120 protomers and I53, I52, and I32 symmetries (**Figure 1.12a**).[67] The ten successfully-realized cage designs range from 1.8-2.8 MDa, and could be assembled *in vitro* to encapsulate GFP cargo molecules. Recently, one of these cages was further altered to bear a positively-charged RNA-binding peptide on the cage lumen, and the resulting construct was shown to exhibit the virus-like property of encapsulating its own RNA genome *in vivo*, and to protect the encapsulated cargo from degradation when isolated from the cell.[110] Recognizing the power of a synergistic design, they used this genotype-phenotype linkage to generate a library of cage variants and carry out directed evolution for improved cargo loading and longer *in vivo* circulation times in mouse models, showing their promise as drug delivery vehicles or as vaccines.[110]

Even so, there are additional hurdles to overcome before designed protein complexes approach the sophistication of natural assemblies. For example, assemblies involving three or more distinct components can be found among natural proteins, but not engineered systems. It is also noteworthy that the two-component engineered systems reported thus far employ the design of a single A-B interface, and exploit existing A-A and B-B oligomerization interfaces. Therefore, the simultaneous design of two or more

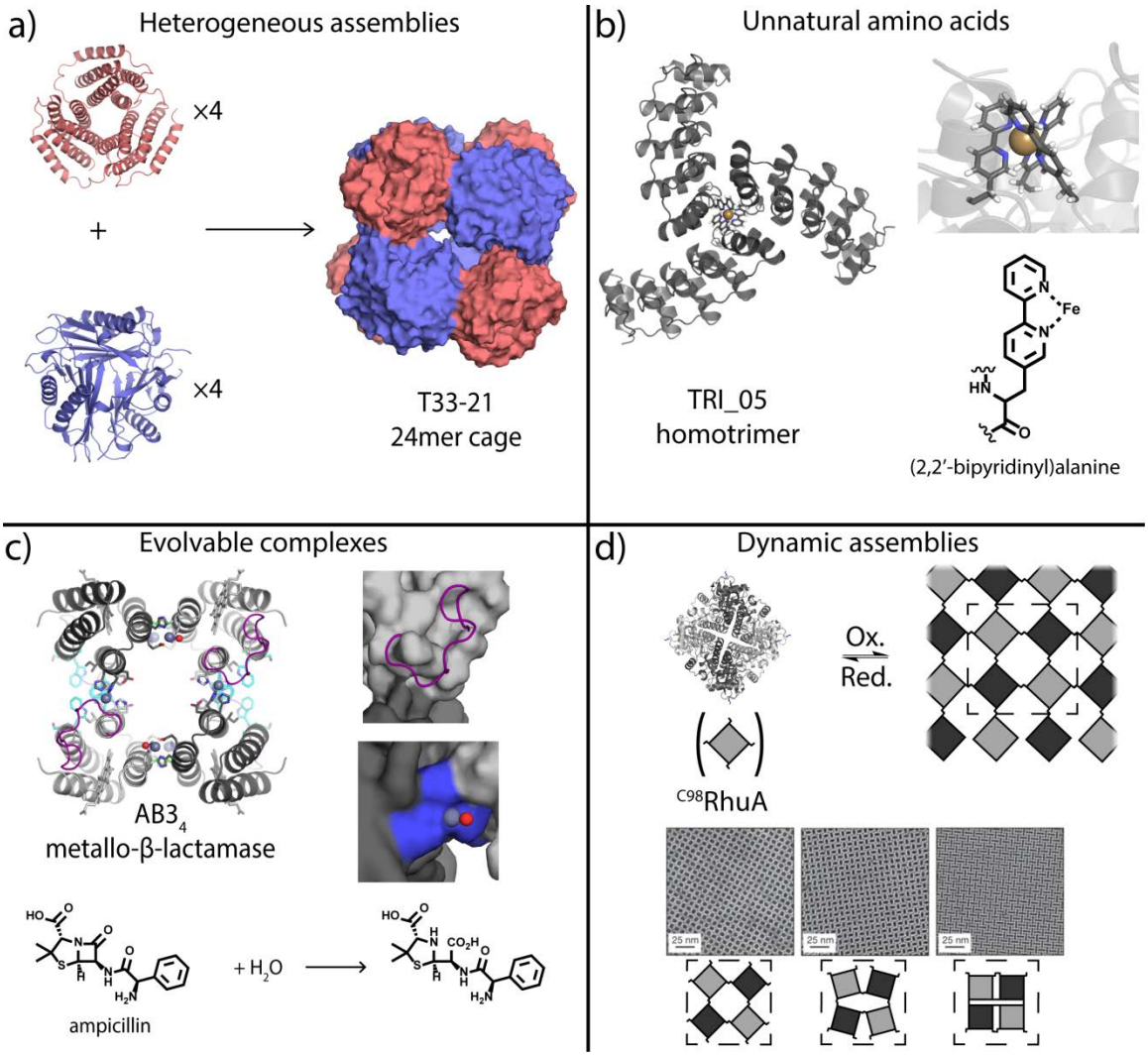

Figure 1.12 Emerging trends in designing protein assemblies.
(a) Computational design efforts have been applied to generate two-component cages, containing four copies of two types of protein trimers by the design of single interface between the pair of components. Adapted from reference 67. (b) The chelating unnatural amino acid (2,2'-bipyridinyl)alanine was installed to form a trimerizing metal binding motif. Adapted from reference 116. (c) A tetrameric metalloprotein bearing interfacial Zn^{II} active sites (gray spheres and blue surface) was subjected to directed evolution, which selected for a mutation that mobilized a surface-exposed loop (shown in purple) that improved ampicillin-degrading activity. Adapted from reference 108. (d) Installing cysteines at the four corners of the tetrameric protein RhuA afforded a variant that oxidatively self-assembles into 2D arrays which, due to the hinge-like properties of the disulfide bonds, are able to undergo concerted opening and closing. Reproduced in part, with permission, from reference 48.

distinct interfaces within a protein complex remains another major milestone yet to be achieved.

Engineered protein-protein interactions incorporating unnatural amino acids

Unnatural amino acids can be incorporated into engineered proteins to provide properties beyond the scope of their natural counterparts. The 20-residue amino acid code evolved under biological constraints, such as metabolic cost [111], that synthetic biology is not bounded by. Unnatural amino acids incorporated by chemical methods have been employed to enhance the stability of leucine zipper proteins,[112] or to generate artificial chelating motifs that permit metal-induced oligomerization.[113, 114] Additionally, fluorinated amino acids can be pan-specifically incorporated *in vivo* which, when supplemented by directed evolution techniques, can also enhance scaffold stability.[115] Advances in synthetic biology tools now permit the site-specific incorporation of unnatural amino acids *in vivo* by repurposing the amber stop codon to encode a 21st amino acid. The Baker group has incorporated the unnatural amino acid (2,2'-bipyridinyl)alanine into a computational design protocol that yielded a homotrimeric metalloprotein with a tris-bipyridyl motif at its core (**Figure 1.12b**).[116] As *in vivo* installation of unnatural amino acids becomes possible for a broader range of residues residues,[117] designed protein interfaces comprising unnatural amino acids will surely be investigated further. For example, interfaces bearing perfluorinated residues have been proposed as one route to achieve specific biorthogonal dimerization,[118] which would undoubtedly aid in generating specific protein assemblies in a cellular environment.

Evolvable protein assemblies

Proteins, as biomolecules, are subject to the forces of natural selection, and their functions are evolvable. This fact has been exploited by protein engineers in the laboratory through directed evolution experiments.[119-121] These efforts allow for mutations that confer improvements in function to be discovered without the need for presuppositions of what those mutations will be. Liu, Baker and co-workers used directed evolution to increase the binding affinity of the computationally designed Prb and Pdar pair.[63] Surprisingly, the evolved proteins dimer deviated significantly from the parent design, with one member flipped by 180, highlighting the power of unanticipated results when coupled with rational design efforts.[63] Directed evolution of cargo loading and stability can be improved and tailored in protein cages,[122], and was shown by the Baker group for their self-assembling protein cages.[110]

Our group has previously used *in vivo* directed evolution to select for improved catalytic activity in a supramolecular metalloenzyme.[108] Here, interfacial and coordinatively unsaturated Zn^{II} sites showed basal levels of β-lactamase activity in ampicillin degradation assays. Saturation mutagenesis of residues flanking the active site was used to generate a library for an *in vivo* survival assay, and the most active variant contained an unanticipated E57G mutation which mobilized a surface-exposed loop flanking the active site (**Figure 1.12c**).[108] *In vitro* enzymatic assays revealed that this mutation conferred a wider substrate scope, and docking studies suggested that the improved ampicillin binding may in part arise from favorable interactions with the newly-mobilized loop.[108] Again underscoring the importance of design plasticity, parallel directed evolution experiments on a tetrameric scaffold with a buried catalytic site did not afford variants with appreciably improved catalytic properties until scaffold rigidity was

slackened by removal of a flanking disulfide bond.[73] However, these efforts involve directed evolution on a protein complex. Selections or screens that do not require existing assembly behavior, and which can select for self-assembly, would constitute a significant advancement in generating new protein assemblies.

Dynamic and responsive protein systems.

Proteins do not exist as static atomic slabs, but rather as ensembles of conformations moving individually and collectively on varying timescales. The study of protein phenomena such as allostery,[6, 123] enzymatic catalysis,[7, 124] and binding interactions[125, 126] have all expanded from simple structure-driven models to views that acknowledge structural plasticity as central to protein functions. Clearly the biophysical principles that govern natural proteins apply to their engineered counterparts as well. Moreover, approaches to protein engineering that focus solely on achieving a particular atomic structure will remain barred from many of the capabilities of proteins.

Disulfide bonds can serve as hinge-like linkers that create strong associations while still allowing for structural plasticity. A single cysteine mutation at position 98 in the C_4-symmetric homotetramer RhuA provides crosslinkable residues at each of the protein's four corners, which allows for oxidative self-assembly into two-dimensional arrays.[48] Strikingly, mechanical perturbation of these arrays leads to an open arrangement, followed by spontaneous and coherent closure (**Figure 1.12d**).[48] Compression along one axis is accompanied by an extension in the transverse axis, making the materials auxetic. This dynamic property arises from the flexibility of the disulfide crosslinkers. Additionally, the small footprint of this feature allows for the

installation of additional engineered elements to tune the dynamic behavior to build in stimulus responsiveness.

Concluding remarks.

Rational design of self-assembling proteins remains a highly active area of research; great strides have been made in exploiting established interactions to generate engineered assemblies, or to install new interaction modalities that can readily furnish a thermodynamically favored assembly of interest. The rapid development of computational methods, synthetic biology tools for installing unnatural amino acids, and techniques for directed evolution are expected to provide even more tools for scientists to create artificial assemblies well beyond the scope of Nature's proteome. Just as the study of natural proteins has matured from a structure-focused approach to paradigms that recognize the importance of structural dynamics, so too should protein engineers focus on the design of multicomponent, flexible assemblies in furtherance of the quest for designed proteins which sophistication rivaling those of Natural protein complexes. The subsequent chapters of this work describe the application of synergistic design to engineer an allosteric protein complex, as well as the subsequent efforts to characterize this designed complex. It is hoped that his model system effectively illustrates the design principles discusses here, and informs future efforts to design functional protein complexes.

Chapter 1 was reproduced in part, with permission, from a manuscript currently being prepared for submission for publication: Churchfield, L.A.; Alberstein, R.G.; Tezcan, F.A. Methods for the design of protein complexes and biomaterials.

The dissertation author is primary author on all reprinted materials.

References

(1) Goodsell, D. S.; Olson, A. J. *Annual Review of Biophysics and Biomolecular Structure* **2000**, *29*, 105–153.

(2) Ali, M. H.; Imperiali, B. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 5013.

(3) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.

(4) Goodsell, D. S.; Dutta, S.; Zardecki, C.; Voigt, M.; Berman, H. M.; Burley, S. K. *PLOS Biology* **2015**, *13*, e1002140.

(5) Hong, F.; Zhang, F.; Liu, Y.; Yan, H. *Chemical Reviews* **2017**,

(6) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014**, *508*, 331.

(7) Kohen, A. *Accounts of Chemical Research* **2015**, *48*, 466.

(8) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433.

(9) Fegan, A.; White, B.; Carlson, J. C. T.; Wagner, C. R. *Chemical Reviews* **2010**, *110*, 3315.

(10) Dotan, N.; Arad, D.; Frolow, F.; Freeman, A. *Angewandte Chemie International Edition* **1999**, *38*, 2363.

(11) Morgan, H.; Taylor, D. M.; D'Silva, C.; Fukushima, H. *Thin Solid Films* **1992**, *210*, 773.

(12) Morgan, H.; Fukushima, H.; Taylor, D. M. *Journal of Polymer Science Part A: Polymer Chemistry* **1994**, *32*, 1331.

(13) Niemeyer, C. M.; Sano, T.; Smith, C. L.; Cantor, C. R. *Nucleic Acids Research* **1994**, *22*, 5530.

(14) Ringler, P.; Schulz, G. E. *Science* **2003**, *302*, 106–109.

(15) Carlson, J. C. T.; Kanter, A.; Thuduppathy, G. R.; Cody, V.; Pineda, P. E.; McIvor, R. S.; Wagner, C. R. *Journal of the American Chemical Society* **2003**, *125*, 1501.

(16) Carlson, J. C. T.; Jena, S. S.; Flenniken, M.; Chou, T.-f.; Siegel, R. A.; Wagner, C. R. *Journal of the American Chemical Society* **2006**, *128*, 7630.

(17) Kitagishi, H.; Kakikura, Y.; Yamaguchi, H.; Oohora, K.; Harada, A.; Hayashi, T. *Angewandte Chemie International Edition* **2009**, *48*, 1271.

(18) Oohora, K.; Burazerovic, S.; Onoda, A.; Wilson, Y. M.; Ward, T. R.; Hayashi, T. *Angewandte Chemie International Edition* **2012**, *124*, 3884.

(19) Green, S. M.; Gittis, A. G.; Meeker, A. K.; Lattman, E. E. *Nature Structural and Moleuclar Biology* **1995**, *2*, 746.

(20) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. *Proceedings of the National Academy of Sciences U.S.A.* **2001**, *98*, 1404.

(21) Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y. J.; Baker, D. *Proceedings of the National Academy of Sciences U.S.A.* **2001**, *98*, 10687.

(22) Ha, J.-H.; Karchin, J. M.; Walker-Kopp, N.; Huang, L.-S.; Berry, E. A.; Loh, S. N. *Journal of Molecular Biology* **2012**, *416*, 495.

(23) Ha, J. H.; Karchin, J. M.; Walker-Kopp, N.; Castaneda, C. A.; Loh, S. N. *Chemistry & Biology* **2015**, *22*, 1384.

(24) Arai, R.; Kobayashi, N.; Kimura, A.; Sato, T.; Matsuo, K.; Wang, A. F.; Platt, J. M.; Bradley, L. H.; Hecht, M. H. *The Journal of Physical Chemistry B* **2012**, *116*, 6789.

(25) Sinclair, J. C.; Davies, K. M.; Venien-Bryan, C.; Noble, M. E. M. *Nature Nanotechnology* **2011**, *6*, 558–562.

(26) Padilla, J. E.; Colovos, C.; Yeates, T. O. *Proceedings of the National Academy of Sciences U.S.A.* **2001**, *98*, 2217–2221.

(27) Lai, Y.-T.; Cascio, D.; Yeates, T. O. *Science* **2012**, *336*, 1129.

(28) Yeates, T. O. *Annual Review of Biophysics* **2017**, *46*, 23.

(29) Yeates, T. O.; Padilla, J. E. *Current Opinion in Structural Biology* **2002**, *12*, 464–470.

(30) Patterson, D. P.; Su, M.; Franzmann, T. M.; Sciore, A.; Skiniotis, G.; Marsh, E. N. G. *Protein Science* **2014**, *23*, 190.

(31) Kobayashi, N.; Yanase, K.; Sato, T.; Unzai, S.; Hecht, M. H.; Arai, R. *Journal of the American Chemical Society* **2015**, *137*, 11285.

(32) 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. *Proceedings of the National Academy of Sciences U.S.A.* **2016**, *113*, 8681.

(33) Uhlenheuer, D. A.; Wasserberg, D.; Nguyen, H.; Zhang, L.; Blum, C.; Subramaniam, V.; Brunsveld, L. *Chemistry – A European Journal* **2009**, *15*, 8779.

(34) Zhang, L.; Wu, Y.; Brunsveld, L. *Angewandte Chemie International Edition* **2007**, *46*, 1798.

(35) Uhlenheuer, D. A.; Young, J. F.; Nguyen, H. D.; Scheepstra, M.; Brunsveld, L. *Chemical Communications* **2011**, *47*, 6798.

(36) Nguyen, H. D.; Dang, D. T.; van Dongen, J. L. J.; Brunsveld, L. *Angewandte Chemie International Edition* **2010**, *49*, 895.

(37) Dang, D. T.; Schill, J.; Brunsveld, L. *Chemical Science* **2012**, *3*, 2679.

(38) Hou, C.; Li, J.; Zhao, L.; Zhang, W.; Luo, Q.; Dong, Z.; Xu, J.; Liu, J. *Angewandte Chemie International Edition* **2013**, *52*, 5590.

(39) Si, C.; Li, J.; Luo, Q.; Hou, C.; Pan, T.; Li, H.; Liu, J. *Chemical Communications* **2016**, *52*, 2924.

(40) Brodin, J. D.; Auyeung, E.; Mirkin, C. A. *Proceedings of the National Academy of Sciences U.S.A.* **2015**, *112*, 4564.

(41) Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A. *Journal of the American Chemical Society* **2007**, *129*, 13374–13375.

(42) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. *Methods in Enzymology* **2016**, *580*, 223.

(43) Salgado, E. N.; Lewis, R. A.; Faraone-Mennella, J.; Tezcan, F. A. *Journal of the American Chemical Society* **2008**, *130*, 6082.

(44) Salgado, E. N.; Lewis, R. A.; Mossin, S.; Rheingold, A. L.; Tezcan, F. A. *Inorganic Chemistry* **2009**, *48*, 2726–2728.

(45) Huard, D. J. E.; Kane, K. M.; Tezcan, F. A. *Nature Chemical Biology* **2013**, *9*, 169.

(46) Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A. *Journal of the American Chemical Society* **2015**, *137*, 10468–10471.

(47) Brodin, J. D.; Ambroggio, X. I.; Tang, C. Y.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. *Nature Chemistry* **2012**, *4*, 375–382.

(48) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A. *Nature* **2016**, *533*, 369–373.

(49) Sontz, P. A.; Bailey, J. B.; Ahn, S.; Tezcan, F. A. *Journal of the American Chemical Society* **2015**, *137*, 11598–11601.

(50) Bailey, J. B.; Zhang, L.; Chiong, J. A.; Ahn, S.; Tezcan, F. A. *Journal of the American Chemical Society* **2017**, *139*, 8160.

(51) Ballister, E. R.; Lai, A. H.; Zuckermann, R. N.; Cheng, Y.; Mougous, J. D. *Proceedings of the National Academy of Sciences U.S.A.* **2008**, *105*, 3733.

(52) Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A. *Journal of the American Chemical Society* **2013**, *135*, 12013.

(53) Grueninger, D.; Treiber, N.; Ziegler, M. O. P.; Koetter, J. W. A.; Schulze, M. S.; Schulz, G. E. *Science* **2008**, *319*, 206.

(54) Garcia-Seisdedos, H.; Empereur-Mot, C.; Elad, N.; Levy, E. D. *Nature* **2017**, *548*, 244.

(55) Liljeström, V.; Mikkilä, J.; Kostiainen, M. A. *Nature Communications* **2014**, *5*, *doi:10.1038/ncomms5445*,

(56) Künzle, M.; Eckert, T.; Beck, T. *Journal of the American Chemical Society* **2016**, *138*, 12731.

(57) Moal, I. H.; Moretti, R.; Baker, D.; Fernández-Recio, J. *Current Opinion in Structural Biology* **2013**, *23*, 862.

(58) Vajda, S.; Kozakov, D. *Current Opinion in Structural Biology* **2009**, *19*, 164.

(59) Mou, Y.; Yu, J.-Y.; Wannier, T. M.; Guo, C.-L.; Mayo, S. L. *Nature* **2015**, *525*, 230.

(60) Huang, P.-S.; Love, J. J.; Mayo, S. L. *Protein Science* **2007**, *16*, 2770.

(61) Stranges, P. B.; Machius, M.; Miley, M. J.; Tripathy, A.; Kuhlman, B. *Proceedings of the National Academy of Sciences U.S.A.* **2011**, *108*, 20562.

(62) Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E. M.; Wilson, I. A.; Baker, D. *Science* **2011**, *332*, 816.

(63) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. A.; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D. *Molecular Cell* **2011**, *42*, 250.

(64) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D. *Science* **2012**, *336*, 1171.

(65) Gonen, S.; DiMaio, F.; Gonen, T.; Baker, D. *Science* **2015**, *348*, 1365.

(66) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. *Nature* **2014**, *510*, 103.

(67) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. *Science* **2016**, *353*, 389.

(68) Churchfield, Lewis A.; George, A.; Tezcan, F. A. *Essays In Biochemistry* **2017**, *61*, 245.

(69) Der, B. S.; Machius, M.; Miley, M. J.; Mills, J. L.; Szyperski, T.; Kuhlman, B. *Journal of the American Chemical Society* **2012**, *134*, 375.

(70) Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A. *Proceedings of the National Academy of Sciences U.S.A* **2010**, *107*, 1827–1832.

(71) Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A. *Annual Review of Biophysics* **2014**, *43*, 409.

(72) Brodin, J. D.; Carr, J. R.; Sontz, P. A.; Tezcan, F. A. *Proceedings of the National Academy of Sciences U.S.A.* **2014**, *111*, 2897–2902.

(73) Song, W. J.; Yu, J.; Tezcan, F. A. *Journal of the American Chemical Society* **2017**,

(74) Fraenkel-Conrat, H.; Williams, R. C. *Proceedings of the National Academy of Sciences U.S.A.* **1955**, *41*, 690.

(75) Ghadiri, M. R.; Soares, C.; Choi, C. *Journal of the American Chemical Society* **1992**, *114*, 825.

(76) Joh, N. H.; Wang, T.; Bhate, M. P.; Acharya, R.; Wu, Y.; Grabe, M.; Hong, M.; Grigoryan, G.; DeGrado, W. F. *Science* **2014**, *346*, 1520.

(77) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. *Science* **2013**, *340*, 595.

(78) Lanci, C. J.; MacDermaid, C. M.; Kang, S.-g.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. *Proceedings of the National Academy of Sciences U.S.A.* **2012**, *109*, 7304.

(79) Lupas, A. N.; Bassler, J. *Trends in Biochemical Sciences 42*, 130.

(80) Woolfson, D. N. *Coiled-coil design: updated and upgraded*; Springer International Publishing: Cham, Switzerland, 2017.

(81) Reinert, D. J.; Carpusca, I.; Aktories, K.; Schulz, G. E. *Journal of Molecular Biology* **2006**, *357*, 1226.

(82) Zegers, I.; Deswarte, J.; Wyns, L. *Proceedings of the National Academy of Sciences U.S.A.* **1999**, *96*, 818.

(83) Bennett, M. J.; Schlunegger, M. P.; Eisenberg, D. *Protein Science* **1995**, *4*, 2455.

(84) Bennett, M. J.; Choe, S.; Eisenberg, D. *Proceedings of the National Academy of Sciences U.S.A.* **1994**, *91*, 3127.

(85) Gronenborn, A. M. *Current Opinion in Structural Biology* **2009**, *19*, 39.

(86) Ostermeier, M.; Benkovic, S. J. *Advances in Protein Chemistry* **2000**, *55*, 29.

(87) Liu, Y.; Eisenberg, D. *Protein Science* **2002**, *11*, 1285.

(88) McMillan, J. R.; Brodin, J. D.; Millan, J. A.; Lee, B.; Olvera de la Cruz, M.; Mirkin, C. A. *Journal of the American Chemical Society* **2017**, *139*, 1754.

(89) Harding, M. M.; Nowicki, M. W.; Walkinshaw, M. D. *Crystallography Reviews* **2010**, *16*, 247–302.

(90) Rulíšek, L.; Vondrášek, J. *Journal of Inorganic Biochemistry* **1998**, *71*, 115–127.

(91) Liu, C. L.; Xu, H. B. *Journal of Inorganic Biochemistry* **2002**, *88*, 77.

(92) Armstrong, R. N. *Biochemistry* **2000**, *39*, 13625.

(93) Bergdoll, M.; Eltis, L. D.; Cameron, A. D.; Dumas, P.; Bolin, J. T. *Protein Science* **1998**, *7*, 1661.

(94) Nakamura, K.; Go, N. *Cellular and Molecular Life Sciences* **2005**, *62*, 2050.

(95) Vasin, A.; Klotchenko, S.; Puchkova, L. *PLoS Currents* **2013**, *5*,

(96) Volbeda, A.; Hol, W. G. J. *Journal of Molecular Biology* **1989**, *206*, 531.

(97) Decker, H.; Terwilliger, N. *Journal of Experimental Biology* **2000**, *203*, 1777.

(98) Harding, M. M. *Acta Crystallographica Section D: Structural Biology* **2004**, *60*, 849.

(99) Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R. *Biochemistry* **2006**, *45*, 10504.

(100) Fass, D. *Annual Review of Biophysics* **2012**, *41*, 63.

(101) Schroeder, H. W., Jr.; Cavacini, L. *Journal of Allergy and Clinical Immunology* **2010**, *125*, S41.

(102) Shoulders, M. D.; Raines, R. T. *Annual Review of Biochemistry* **2009**, *78*, 929.

(103) Lo Conte, L.; Chothia, C.; Janin, J. *Journal Of Molecular Biology* **1999**, *285*, 2177.

(104) Janin, J.; Chothia, C. *Journal of Biological Chemistry* **1990**, *265*, 16027.

(105) Chen, J.; Sawyer, N.; Regan, L. *Protein Science* **2013**, *22*, 510.

(106) Kortemme, T.; Baker, D. *Current Opinion in Chemical Biology* **2004**, *8*, 91.

(107) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D. *Science* **2012**, *336*, 1171.

(108) Song, W. J.; Tezcan, F. A. *Science* **2014**, *346*, 1525–1528.

(109) Junge, W.; Nelson, N. *Annual Review of Biochemistry* **2015**, *84*, 631.

(110) Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U.; Ellis, D.; Bale, J. B.; Ke, S.; Lenz, G. H.; Yehdego, A.; Ravichandran, R.; Pun, S. H.; King, N. P.; Baker, D. *Nature* **2017**, *552*, 415.

(111) Akashi, H.; Gojobori, T. *Proceedings of the National Academy of Sciences U.S.A.* **2002**, *99*, 3695.

(112) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, I. W.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.

(113) Radford, R. J.; Tezcan, F. A. *Journal of the American Chemical Society* **2009**, *131*, 9136–9137.

(114) Radford, R. J.; Nguyen, P. C.; Ditri, T. B.; Figueroa, J. S.; Tezcan, F. A. *Inorganic Chemistry* **2010**, *49*, 4362.

(115) Montclare, J. K.; Tirrell, D. A. *Angewandte Chemie International Edition* **2006**, *45*, 4518.

(116) Mills, J. H.; Sheffler, W.; Ener, M. E.; Almhjell, P. J.; Oberdorfer, G.; Pereira, J. H.; Parmeggiani, F.; Sankaran, B.; Zwart, P. H.; Baker, D. *Proceedings of the National Academy of Sciences U.S.A.* **2016**, *113*, 15012.

(117) Lang, K.; Chin, J. W. *Chemical Reviews* **2014**, *114*, 4764.

(118) Marsh, E. N. G. *Accounts of Chemical Research* **2014**, *47*, 2878.

(119) Jäckel, C.; Kast, P.; Hilvert, D. *Annual Review of Biophysics* **2008**, *37*, 153–173.

(120) Packer, M. S.; Liu, D. R. *Nature Reviews Genetics* **2015**, *16*, 379.

(121) Renata, H.; Wang, Z. J.; Arnold, F. H. *Angewandte Chemie International Edition* **2015**, *54*, 3351.

(122) Azuma, Y.; Edwardson, T. G. W.; Hilvert, D. *Chemical Society Reviews* **2018**, *47*, 3543.

(123) Nussinov, R. *Chemical Reviews* **2016**, *116*, 6263.

(124) Klinman, J. P.; Kohen, A. *Journal of Biological Chemistry* **2014**, *289*, 30205.

(125) Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C. *Accounts of Chemical Research* **2016**, *49*, 809.

(126) Haliloglu, T.; Bahar, I. *Current Opinion in Structural Biology* **2015**, *35*, 17.

Chapter 2: *De novo* design of an allosteric metalloprotein.

2.1 Introduction

In association with proteins, metal ions fulfill central biological roles that range from structural stabilization and signaling to electron transfer and catalysis.[1, 2] At the core of these diverse roles lies an exquisite interplay between the energetics of the bound metals and the protein matrix.[3-5] This interplay is well understood at the level of the primary and secondary coordination spheres of metal ions, and has been recapitulated in biomimetic systems[6, 7] and designed or reengineered proteins.[8, 9] Considerably more difficult to model is the relationship between metal coordination/reactivity and–outer-sphere interactions at the tertiary and quaternary structure level. Allostery and

cooperativity are particularly striking manifestations of the control of metal coordination through outer sphere interactions, as embodied by hemoglobin and metalloregulatory proteins. In hemoglobin, structural changes that result from O_2 coordination to a single heme cofactor are propagated onto the entire quaternary architecture, causing the structural equilibrium of the protein to shift from the "tense" state to the "relaxed" state, thereby leading to progressively more favorable O_2 binding (positive cooperativity) by the remaining heme cofactors.[10] Similarly, in metalloregulatory proteins, the equilibria for metal ion coordination are remotely coupled to those for DNA binding, enabling the regulation of gene transcription in response to changes in metal concentrations.[11]

De novo construction of allosteric systems like hemoglobin and metalloregulatory proteins represents an outstanding goal in protein design. The challenge therein stems not only from the necessity to design at least two (re-)active sites in a single protein, but also from the requirement to efficiently couple these sites via a structural conduit. Clearly, these design criteria (>two reactive sites and remote coupling) would disfavor scaffolds that are small or possess extreme rigidity or flexibility. In fact, a large number of natural allosteric proteins are large, oligomeric complexes that contain semi-rigid subunits linked by flexible interfaces.[12-14] In this way, mechanical strain generated by a chemical event in one site can be transmitted onto the interfaces and propagated throughout the entire superstructure. Although allosteric proteins have been engineered through different strategies, these typically involve either the modification of existing proteins that undergo ligand-specific conformational changes or the fusion of protein domains with preexisting chemical functions.[15-18] In this chapter, we report the *de novo* construction of an allosteric protein assembly with flexible interfaces, in which two distinct chemical functions–metal

binding and disulfide bond formation and the structural elements that couple these functions are designed in concert. This study highlights how quaternary strain can be built in a bottom-up fashion such that a) the thermodynamics of metal ion coordination can be regulated through outer sphere interactions, and b) metal binding can be remotely coupled to reversible disulfide bond formation, a premier chemical strategy that nature uses for redox regulation.[19, 20]

Our strategy for building quaternary strain is based MeTIR (see **Figure 1.11**), an approach we developed for the construction of functional metalloprotein assemblies. Using MeTIR, we have previously reported that a monomeric protein (cyt cb_{562}) could be decorated with metal chelating groups, which direct the self-assembly of cyt cb_{562} into a discrete tetrameric architecture, Zn-M1_4, upon Zn^{II} coordination (**Figure 2.1**).[21, 22] This D_2 symmetric tetramer presents three pairs of C_2 symmetric interfaces (*i1*, *i2*, and *i3*) that are independently tailorable, forming a tunable shell around the four internalized Zn^{II} coordination sites. Applying MeTIR to incorporate favorable, non-covalent interactions into the *i1* interfaces through six, primarily hydrophobic mutations afforded the variant R1 (**Figure 2.1**).[23] This was followed by the incorporation of disulfide linkages into the *i2* and *i3* interfaces through T96C and E81C mutations.[24, 25] The resulting tetramer $^{C81/C96}R1_4$ could self-assemble efficiently from the monomeric building blocks with the proper formation of C81-C81 and C96-C96 disulfide linkages and coordinate Zn^{II} ions with high affinity as intended by the templating strategy (**Figure 2.1**).[25] Successive engineering of the interfaces has two consequences: first, it increases the overall rigidity and preorganization of the quaternary assembly, thus turning it increasingly more into a

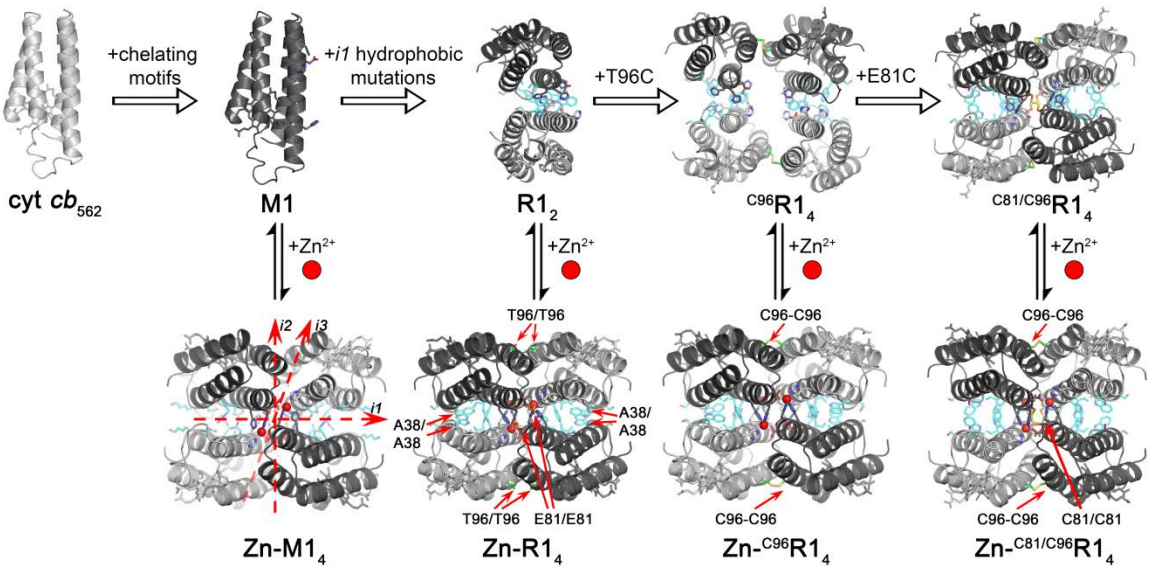

Figure 2.1 Successive engineering of cyt cb_{562} to form disulfide-linked oligomeric architectures. Using Metal Templated Interface Redesign (MeTIR), metal chelating motifs (shown as sticks) were installed on the surface of monomeric cyt cb_{562} (PDB ID: 2BC5). The resulting variant, M1 (or MBPC1, PDB ID: 2QLA), forms a Zn-dependent tetramer, Zn-$M1_4$. This tetrameric scaffold contains three C_2 symmetric interfaces, *i1*, *i2*, and *i3*. The scaffold was further engineered to include favorable protein-protein contacts at interface *i1* through the installation of hydrophobic patches (cyan). The second-generation variant, $R1_1$, self assembles into a dimer ($R1_2$, PDB ID: 3HNK) and forms a Zn-induced tetramer (Zn-$R1_4$, PDB ID: 3HNI) that is topologically identical to the parent variant. Within the tetramer are symmetry-related residues, A38/A38, E81/E81, and T96/T96, that are in close proximity. Previous work has shown that Zn-mediated R1 assembly can template C96-C96 (green sticks) disulfide bond formation to afford a tetrameric species that is maintained in the absence of metal ($^{C96}R1_4$, PDB IDs: 3IQ5 and 3IQ6). Additionally, in the case of $^{C81/C96}R1_4$, Zn-mediated R1 assembly can template the formation of C96-C96 and C81-C81 (orange sticks) disulfide bonds simultaneously (PDB IDs: 4JE9 and 4JEA).

monolithic architecture. Second, it increases the overall strain that Zn^{II} coordination and the interfacial interactions exert on one another. The strain is evident when comparing the structures of Zn-bound and apo- $^{C81/C96}R1_4$ (**Figure 2.2**). In particular, the hydrophobically stabilized *i1* interfaces (sidechains shown in cyan in **Figure 2.2**) undergo a substantial deformation upon Zn^{II} removal, whereby the average α-C distance between the symmetrically related A38 residues increases from 7.9 to 10.6 Å. This deformation is accompanied by conformational changes in the C96-C96 and C81-C81 linkages that hold the *i2* and *i3* interfaces together in a hinge-like fashion (**Figure 2.2**). In this study, we set out to examine if the *i1* interfaces in $^{C81/C96}R1_4$ could also be disulfide-crosslinked through the incorporation of a Cys residue at position 38. We envisioned that this would generate a spring-loaded quaternary structure with highly strained and potentially dissociable disulfide bonds that could allow structural coupling between Zn-binding and-disulfide bond formation equilibria.

2.2 Results and discussion

Toward this end, we generated the triple Cys mutant $^{C38/C81/C96}R1$, which was expressed in *E. coli* and isolated as a monomer in high yield. Given the presence of three Cys residues on the surface of $^{C38/C81/C96}R1$, the formation of unstructured, disulfide-linked aggregates over the desired tetramer ($^{C38/C81/C96}R1_4$) becomes a potential concern. Nevertheless, we found that the self-assembly of $^{C38/C81/C96}R1$ under air oxidation preferentially gave rise to a tetrameric species in relatively high yield (47±4%, see **Fig-**

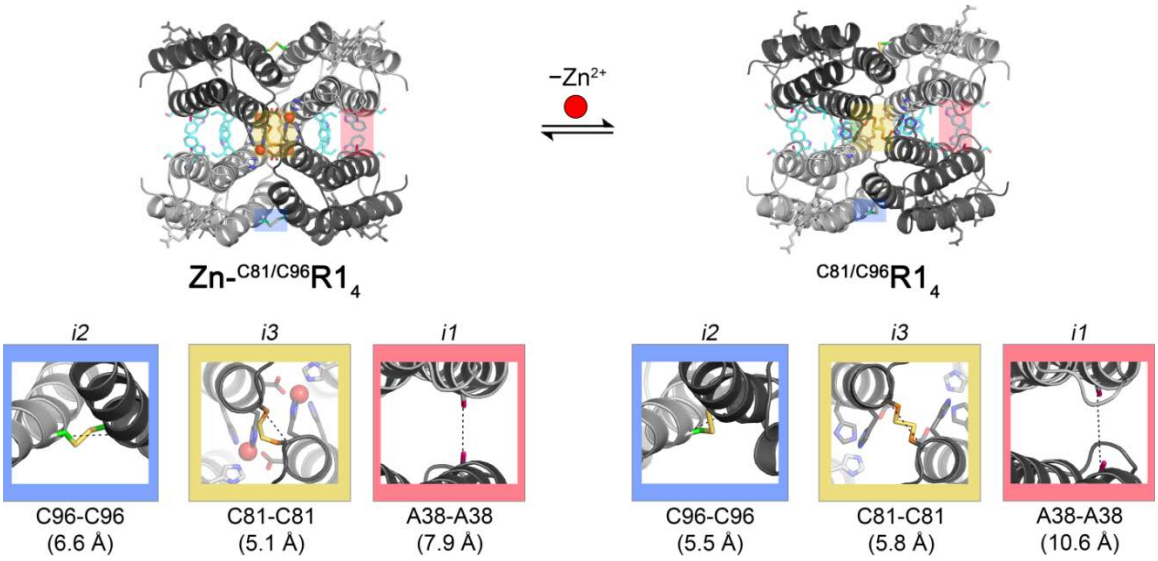

Figure 2.2 Structural rearrangements upon Zn^{II} removal from Zn-$^{C81/C96}R1_4$.
The disulfide crosslinks across interfaces *i2* and *i3* (C96-C96 and C81-C81, respectively), undergo conformational changes in a hinge-like fashion when Zn^{II} is removed from the tetramer. The accompanying structural rearrangements increase the separation of the symmetry-related A38-A38 residues from 7.9 Å to 10.6 Å. Dashed lines and values in parentheses denote the average α-carbon separations of the indicated residues in the Zn- and apo-$^{C38/C81/C96}R1_4$ structures.

ure 2.3). The inclusion of 1.5-fold molar excess of Zn^{II} ions gave a slight but reproducible increase in the yield of the tetrameric species (53±7%, see **Figure 2.3**). These observations suggest that the hydrophobic mutations engineered to stabilize the *i1* interface likely align the $^{C38/C81/C96}$R1 monomers in the proper orientation for the desired disulfide bonds (C38-C38, C81-C81, C96-C96 rather than any hetero-pair) to form, and that Zn^{II} ions have a modest templating effect. Indeed, oxidative self-assembly of the variant $^{C38/C81/C96}$M1 lacking the hydrophobic mutations led primarily to the population of higher-order aggregates (**Figure 2.3**) with no preference for the formation of the tetrameric species (16% yield, see **Figure 2.3**).

The $^{C38/C81/C96}R1_4$ tetramer, self-assembled in the presence of Zn^{2-} ions, was isolated in high purity (>90%, **Figure 2.4**) by size-exclusion chromatography and treated with ethylenediaminetetraacetic acid (EDTA) to produce fully demetallated stock solutions. Upon treatment with the reducing agent dithiothreitol (DTT), $^{C38/C81/C96}R1_4$ fully dissociated into a monomeric form (**Figure 2.4**), confirming that its self-assembly is mediated by disulfide bond formation. The narrow and symmetric sedimentation-velocity profile of $^{C38/C81/C96}R1_4$ (S_{max} = 4.3) is consistent with a closed tetrameric architecture that undergoes a slight structural rearrangement upon binding Zn^{II} (S_{max} = 4.4) (**Figure 2.5**). We examined the Zn^{II} binding properties of $^{C38/C81/C96}R1_4$ using the fluorophore Fura-2, fit by a thermodynamic model with two pairs of equivalent Zn-binding sites and two corresponding dissociation constants (K_d = 8.1 ± 0.4 and 400 ± 100 nM; overall ΔG_{Zn} = – 164 ± 1 kJ/mol) (**Figure 2.6** and **Table 2.1**). These values indicate a significantly diminished Zn-binding affinity for $^{C38/C81/C96}R1_4$ compared to the previously

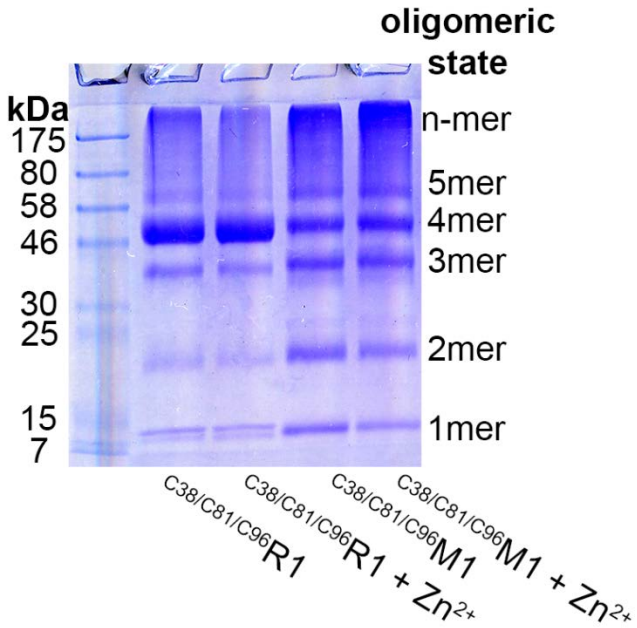

Figure 2.3 SDS-PAGE gel of the products of $^{C38/C81/C96}$R1 and $^{C38/C81/C96}$M1 self-assembly reactions in the presence or absence of Zn^{II}.
The $^{C38/C81/C96}$M1 variant lacks the hydrophobic mutations engineered to stabilize the *i1* interfaces in $^{C38/C81/C96}$R1. The gel was run in the absence of any reductants to keep disulfide bonds intact.Templating was carried out using 50 μM monomer in a buffered solutoin of 20 mM Tris, pH 7 and 150 mM NaCl. Samples were incubated overnight at 37 °C under ambient atmosphere.

Table 2.1 Thermodynamic parameters for Zn^{II} binding to disulfide-linked $R1_4$ tetramers.
Binding affinities were determined by Zn^{II} titration in the presence of Fura-2. Errors denote fitting errors of a single measurement. Binding enthalpies were measured by ITC for the titration of Zn^{II} with disulfide-linked $R1_4$ variants. Reported enthalpies are not corrected for buffer effects, and reported errors are the standard deviation among 3-4 replicate measurements. Entropies of binding are inferred from the other measured quantities. All measurements were carried out at 22 °C and in the presence of 20 mM MOPS, pH 7 + 150 mM NaCl. [a]Data taken from reference 24. [b]Data taken from reference 25.

	$^{C96}R1_4$	$^{C81/C96}R1_4$	$^{C38/C81/C96}R1_4$
Binding affinities			
K_{d1} (nM)	1.3 ± 0.3[a]	2.6 ± 0.3 [b]	8.1 ± 0.4
K_{d2} (nM)	0.53 ± 0.07[a]	25 ± 4 [b]	400 ± 100
K_{d3} (nM)	33 ± 8 [a]	-	-
K_{d4} (nM)	58 ± 8 [a]	-	-
Total ΔG_{Zn} (kJ/mol)	-186 ± 2 [a]	-183 ± 1 [b]	-164 ± 1
Enthalpy of binding			
$\Delta H_{1,ITC}$ (kJ/mol)	18 ± 3	14 ± 4	4 ± 2
$\Delta H_{2,ITC}$ (kJ/mol)	14 ± 10	33 ± 6	5 ± 3
$\Delta H_{3,ITC}$ (kJ/mol)	33 ± 15	11 ± 10	2 ± 3
$\Delta H_{4,ITC}$ (kJ/mol)	−3 ± 4	1 ± 2	−1 ± 2
Total ΔH_{ITC} (kJ/mol)	62 ± 13	59 ± 1	10 ± 0.3
Entropy of binding			
Total ΔS_{ITC} (kJ/mol•K)	0.84 ± 0.05	0.82 ± 0.01	0.59 ± 0.01

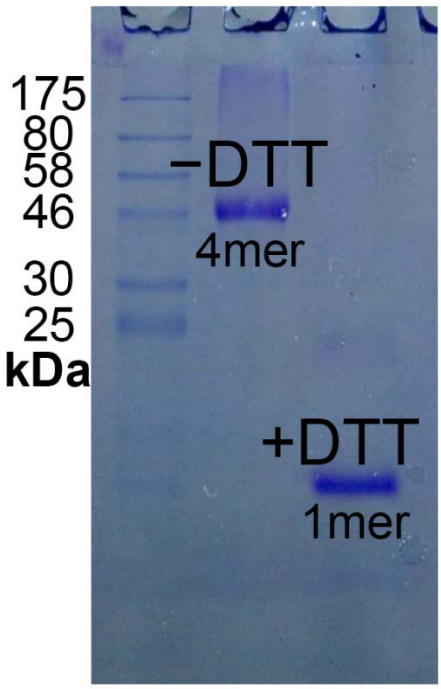

Figure 2.4 Non-reducing SDS-PAGE of $^{C38/C81/C96}R1_4$ in the absence and presence of DTT.
Isolated $^{C38/C81/C96}R1_4$ protein runs as a tetramer (~48 kDa) by SDS-PAGE in the absence of reductant. In the excess reductant (DTT), gives a single band consistent with a monomeric species.

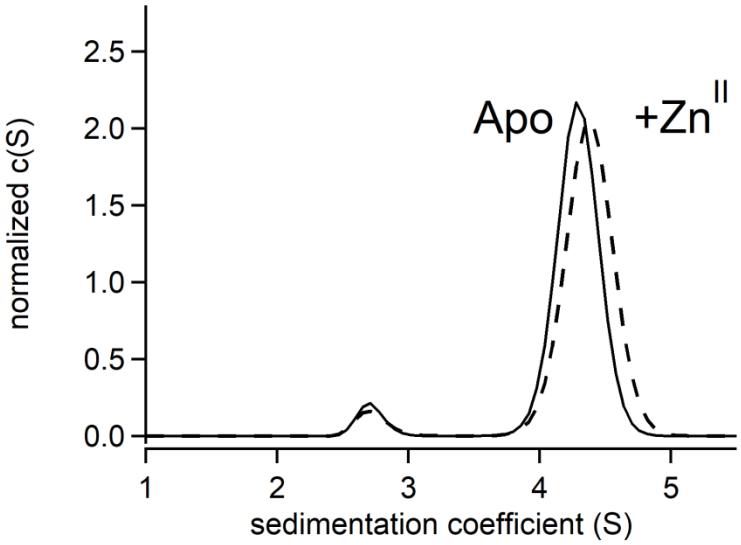

Figure 2.5 Sedimentation velocity profiles of $^{C38/C81/C96}R1_4$.
Representative c(S) distributions of $^{C38/C81/C96}R1_4$ (1.25 μM tetramer) samples treated with either 5 μM $ZnCl_2$ ($+Zn^{II}$; dashed trace) or 1 mM EDTA (apo; solid trace). The distributions are noramlized such that the total area under each curve equals one.

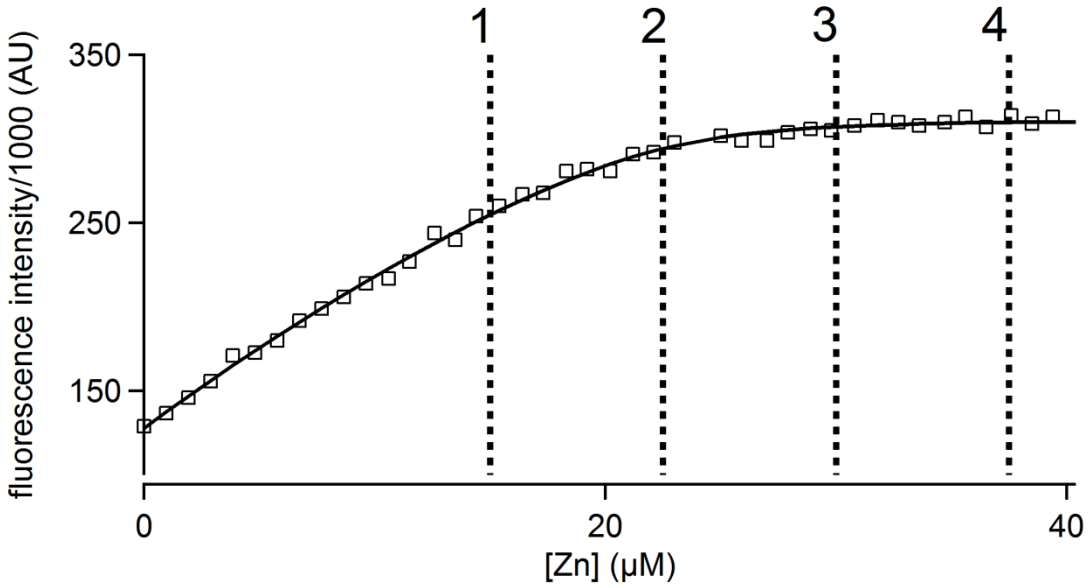

Figure 2.6 Binding isotherm of Zn^{II} to Fura-2 in the presence of $^{C38/C81/C96}R1_4$.
A sample of $^{C38/C81/C96}$R1 and Fura-2 was titrated with Zn^{II}, and the fluorescence intensity was monitored throughout the titration (open squares). The data were fit to a model where four binding sites of $^{C38/C81/C96}R1_4$ were described by two dissociation constants (2 + 2 model). Vertical lines denote the molar equivalents of Zn^{II} added (with respect to $^{C38/C81/C96}R1_4$ monomer) in excess of Fura-2.

characterized tetramers $^{C96}R1_4$ (ΔG_{Zn} = –186 ± 2 kJ/mol) and $^{C81/C96}R1_4$ (ΔG_{Zn} = –183 ± 1 kJ/mol),[24, 25] which feature one and two pairs of disulfide bonds, respectively. These findings indicate that the presence of all three pairs of interfacial disulfide bonds imparts a strain at the level of the quaternary architecture, manifested in the destabilization of Zn^{II} coordination by ~20 kJ/mol.

To get a deeper insight into how the interfacial disulfide linkages influence Zn-binding thermodynamics, we conducted isothermal titration calorimetry (ITC) experiments with $^{C96}R1_4$, $^{C81/C96}R1_4$ and $^{C38/C81/C96}R1_4$ scaffolds, bearing one, two and three pairs of disulfide bonds, respectively. In all cases, the ITC thermograms were consistent with the coordination of four Zn^{II} ions per tetramer (**Figure 2.7**), and allowed us to estimate the apparent Zn binding enthalpies (ΔH_{ITC}) in combination with the thermodynamic parameters derived from fluorescence titrations (**Table 2.1**). Since the primary Zn^{II} coordination spheres are conserved in all three tetramers (see below), their measured ΔH_{ITC} values can be meaningfully compared without deriving their condition-independent Zn-binding enthalpies (ΔH_{Zn}).[26] Our analysis reveals that $^{C96}R1_4$ and $^{C81/C96}R1_4$ display nearly identical ΔH_{ITC} values of 62 and 59 kJ/mol, respectively. Strikingly, Zn^{II} coordination by $^{C38/C81/C96}R1_4$ is significantly more exothermic (ΔH_{ITC} = 10 kJ/mol). This 50 kJ/mol decrease in enthalpy suggests that there may be a distinct structural change associated with the $^{C38/C81/C96}R1_4$ architecture that is coupled to Zn^{II} coordination. Conversely, the $^{C38/C81/C96}R1_4$ tetramer exhibits a less-favorable entropy gain upon Zn^{II} binding compared to $^{C81/C96}R1_4$ and $^{C96}R1_4$ (**Table 2.1**). Though changes in entropy can arise from multiple sources, the observed trend is consistent with decreased conformational flexibility of the $^{C38/C81/C96}R1_4$ scaffold.

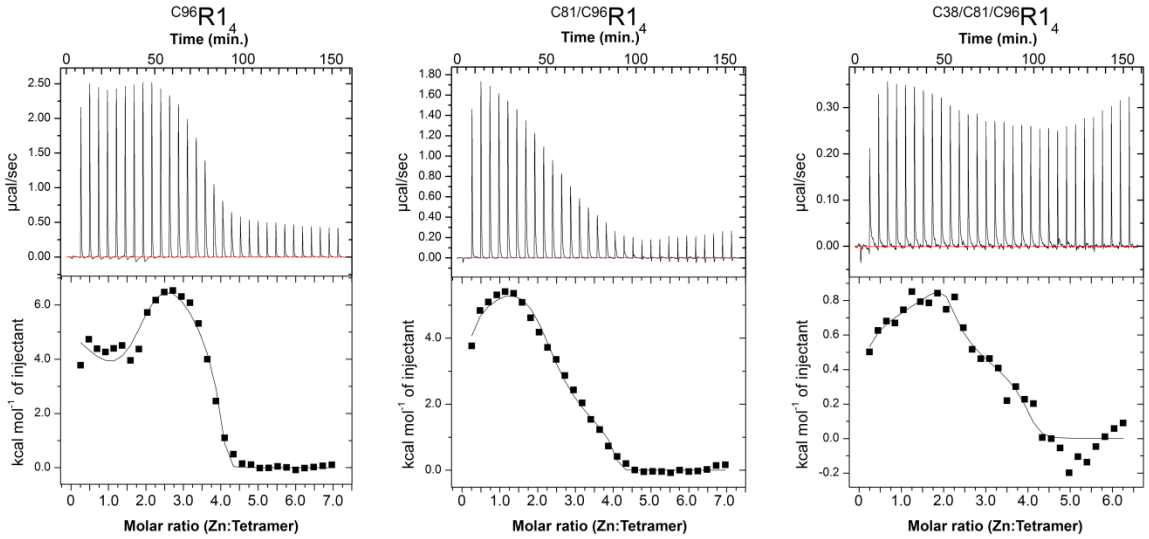

Figure 2.7 ITC thermograms of Zn^{II} titrated into disulfide-linked R1 tetramers.
Representative baseline-corrected and integrated ITC thermograms for disulfide-linked $R1_4$ variants titrated with $ZnCl_2$. Solid lines are for fits to four independent binding enthalpies using the dissociation constants derived from fluorescence competition assays.

Table 2.2 X-ray crystallography data collection and refinement statistics for $^{C38/C81/C96}R1_4$ structures.
Values in parentheses denote statistics for the highest resolution shell.

‡Rsym = $\Sigma\Sigma_j|I_j - \langle I \rangle|/\Sigma\Sigma_j|Ij|$.
§R = $\Sigma||F_{obs}| - |F_{calc}||/\Sigma |F_{obs}|$.
[II]Free R calculated against 5% of the reflections removed at random for both structures.
¶Root mean square deviations from bond and angle restraints.
*Values in parentheses correspond to the highest resolution shell.

	$^{C38/C81/C96}R1_4$	Zn-$^{C38/C81/C96}R1_4$
Data collection		
Space group	$P\ 2_12_12_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	62.59, 77.53, 88.06	48.06, 62.53, 72.44
α, β, γ (˚)	90.00, 90.00, 90.00	90.00, 99.01, 99.00
Resolution (Å)	38.77-2.40	37.81-2.10
R_{sym} ‡ (%)*	6.7 (23.0)	4.9 (46.3)
I/σ*	14.9 (5.7)	15.0 (3.0)
$CC_{1/2}$ (%)*	99.9 (96.8)	99.9(79.6)
Completeness (%)*	99.8 (99.6)	99.5 (99.3)
Redundancy	6.9	3.7
Refinement		
Resolution (Å)	2.40	2.10
No. unique reflections	17325	24815
R_{work} §/R_{free} II (%)	22.5/26.5	21.4/25.0
No. atoms		
Protein	3224	3264
Ligand/ion	172	195
Water	64	73
***B*-factors**		
Protein	56.54	47.19
Ligand/ion	54.31	43.45
Water	47.83	39.94
R.m.s. deviations		
Bond lengths (Å) ¶	0.008	0.018
Bond angles (˚) ¶	1.30	1.57

We determined the X-ray crystal structures of Zn-bound and apo-$^{C38/C81/C96}R1_4$ at 2.1 Å and 2.4 Å resolution, respectively (PDB IDs: 5L32 and 5L31) (**Table 2.2**). As designed, Zn-$^{C38/C81/C96}R1_4$ contains four Zn-coordination sites and is structurally very similar to its predecessors Zn-$^{C96}R1_4$ and Zn-$^{C81/C96}R1_4$ both in terms of over topology (α-C RMSD = 1.78 Å and 1.02 Å, respectively) and primary/secondary Zn^{II} coordination spheres (**Figures 2.8**, **Figure 2.9**, and **Figure 2.10**). The presence of all planned interfacial disulfide bonds is clearly evident from the electron density maps (**Figure 2.10**). To our knowledge, the Zn-$^{C38/C81/C96}R1_4$ scaffold represents the first designed protein that self-assembles into a closed architecture through the formation of six covalent bonds. All three pairs of disulfides adopt distinct conformations in Zn-$^{C38/C81/C96}R1_4$, with the C38-C38 disulfide pairs possessing the highest strain energy as calculated using the method reported by Katz and Kossiakoff (**Table 2.3**).[27]

Indeed, the apo-$^{C38/C81/C96}R1_4$ structure reveals that, upon removal of Zn^{II}, one of the two C38-C38 disulfide pairs is broken. This yields a tetrameric architecture that now contains five disulfide bonds and does not possess the original D_2 symmetry of the Zn-bound assembly any longer (**Figure 2.8 a-c**). Despite extensive disulfide crosslinking, there is a sizable conformational shift from the Zn-bound to the apo-$^{C38/C81/C96}R1_4$ structure (RMSD = 2.55 Å) (**Figure 2.11**). This shift is paralleled by the rearrangement of all existing disulfide bonds whose calculated strain energies decrease upon Zn^{II} removal (**Table 2.3**). These observations suggest that Zn^{II} coordination spring-loads the quaternary architecture through the formation of strained disulfide bonds, and that this strain is relieved through the dissociation of a single C38-C38 bond which is >14 Å away from the nearest Zn^{II} coordination site. The fact that the Zn-$^{C38/C81/C96}R1_4$ crystals were

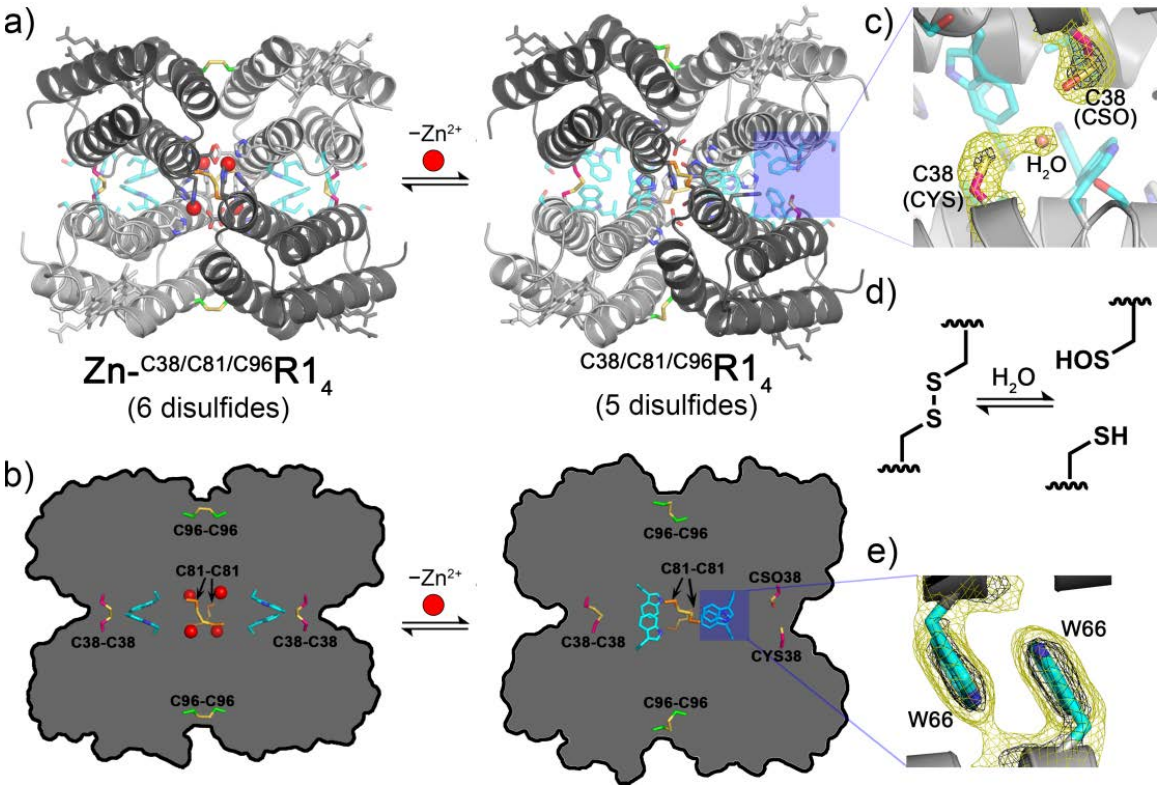

Figure 2.8 Overview of Zn-induced structural changes in $^{C38/C81/C96}R1_4$.
(a) X-ray crystal structures of Zn-$^{C38/C81/C96}R1_4$ and the metal-free $^{C38/C81/C96}R1_4$. The broken C38-C38 disulfide bond is highlighted with a blue box. Disulfide bonds, Zn-binding sites, and the engineered hydrophobic mutations in *i1* (cyan) are shown as sticks. (b) Scheme highlighting rearrangements of interfacial residues in Zn-$^{C38/C81/C96}R1_4$ upon Zn^{II} removal. (c) Close-up view of the broken C38-C38 disulfide bond, forming a cysteine (Cys) and cysteine sulfenic acid (Cso). A water molecule H-bonded to Cys 38 is shown as a pink sphere. (d) Hydrolytic dissociation of a disulfide bond into a sulfenic acid-thiol pair. (e) Close-up view of the W66-W66 π-stacking interaction. All $2F_o-F_c$ electron density maps are contoured at 1σ (yellow) and 2.5 σ (black).

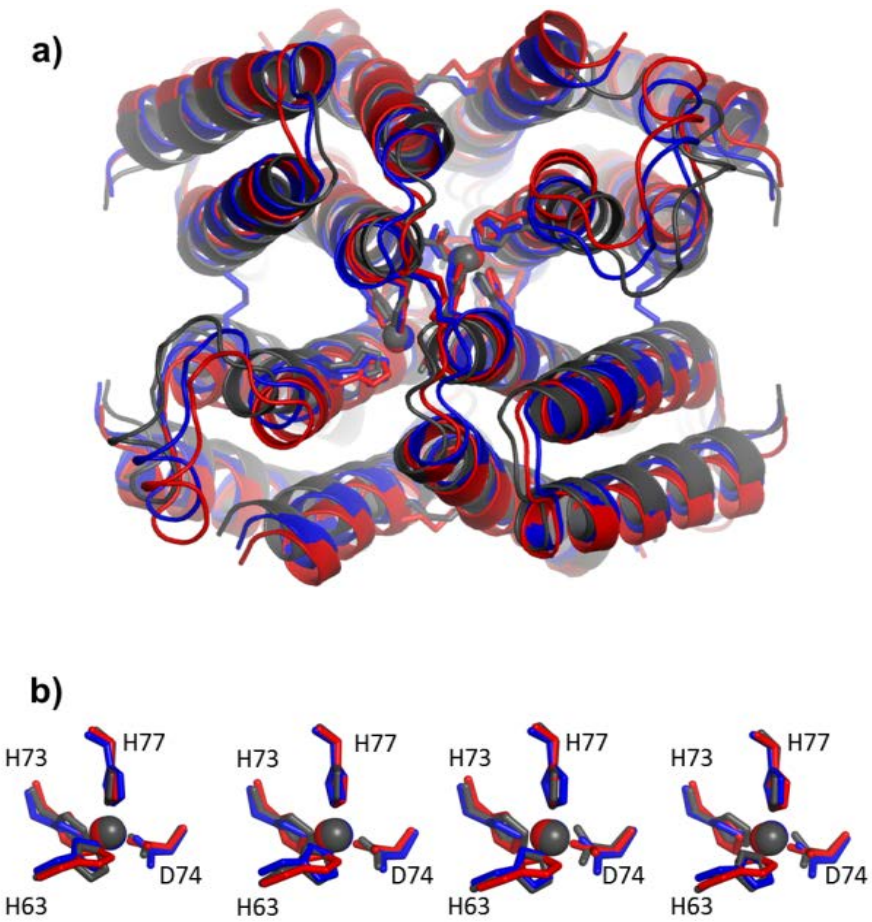

Figure 2.9 Structural alignments of metal-bound disulfide-linked R1 tetramers.
(a) Structural overlay of Zn-$^{C96}R1_4$ (black), Zn-$^{C81/C96}R1_4$ (red), and Zn-$^{C38/C81/C96}R1_4$ (blue) aligned over all α-carbon atoms. (b) Individual Zn^{II} sites within each structure aligned to the ligand atoms only.

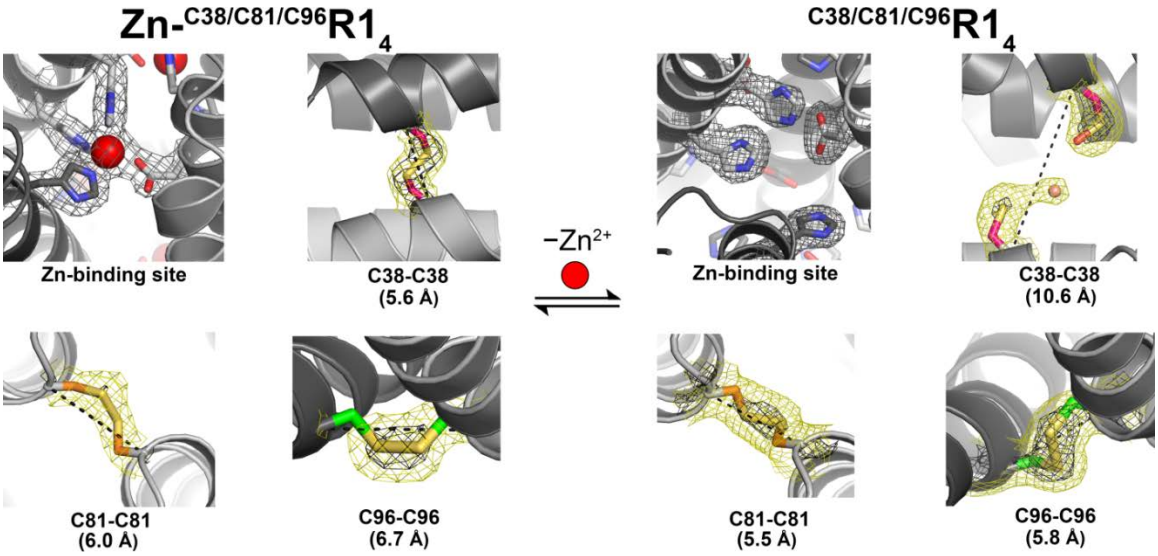

Figure 2.10 Electron density maps of key features in $^{C38/C81/C96}R1_4$ structures.
Close-up views of the Zn^{II} coordination sites and disulfide bonds in Zn-$^{C38/C81/C96}R1_4$ and $^{C38/C81/C96}R1_4$ structures. $2F_o$-F_c electron density maps are contoured at 1 σ (yellow) and 2.5 σ (black). Dashed lines and values in parentheses denote the α-carbon separation between the indicated residue pairs.

Table 2.3 Dihedral strain energies of disulfide bonds in $^{C38/C81/C96}R1_4$ structures.
Calculated strain energies of disulfide bonds in the available $^{C83/C81/C96}R1_4$ structures. Reported energies are averaged over disulfide pairs within each structure, with the exception of the single C38-C38 disulfide in apo-$^{C83/C81/C96}R1_4$. Strain energies were calculated as described in reference 27.

	Average strain energy (kJ/mol)	
Disulfide site	**Zn-$^{C38/C81/C96}R1_4$**	**apo-$^{C38/C81/C96}R1_4$**
C38-C38	25	19
C81-C81	23	13
C96-C96	17	6

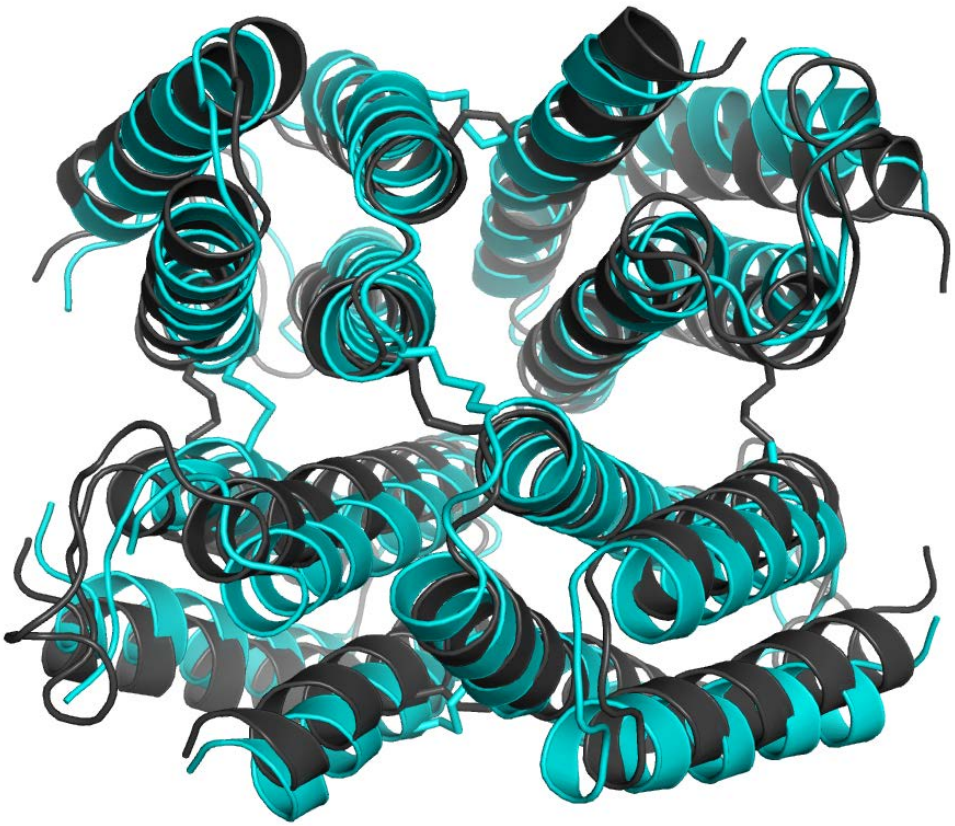

Figure 2.11 Structural alignment of Zn-bound and metal-free $^{C38/C81/C96}R1_4$.
Structural overlay of Zn-$^{C38/C81/C96}R1_4$ (black) and $^{C38/C81/C96}R1_4$ (cyan) assemblies aligned over all α-carbon atoms (RMSD = 2.55 Å). Disulfide bonds and the broken C38-C38 disulfide are shown as sticks.

obtained by adding Zn^{II} to apo-$^{C38/C81/C96}R1_4$ stock solutions indicates that C38-C38 bond formation is reversible. Moreover, the direct coupling between Zn^{II} coordination and reversible disulfide bond formation in $^{C38/C81/C96}R1_4$ is consistent with its significantly more exothermic Zn^{II} binding enthalpy compared to $^{C96}R1_4$ and $^{C81/C96}R1_4$, which do not contain dissociable disulfide bonds.

The breakage of the C38-C38 linkage is accompanied by a considerable distortion of the *i1* interface, whereby the distance between the α-Cs of the C38 residues increases from 5.5 Å to 10.6 Å. This distortion largely involves the sliding of the two protein monomers with respect to another, which is enabled by the fluidity of the hydrophobic interactions in *i1*. Notably, two sets of Trp residues (W41 and W66) that form the core of the *i1* interface adopt new arrangements in the vicinity of the cleaved C38-C38 disulfide bond. The W66 sidechains are now found in a π-stacking arrangement with an interplanar distance of ~3 Å (in contrast to a distance of >10 Å in the Zn-$^{C38/C81/C96}R1_4$ structure), likely making a contribution to the stability of the *i1* interface in the apo-state (**Figure 2.8d-e**).

Given that there are no reducing agents in any of the protein solutions, the question arises as to how the C38-C38 bond is broken upon Zn^{II} dissociation. A mechanistic clue is provided by the elongated electron density associated with one of the C38 side chains (**Figure 2.8c** and **Figure 2.12**), suggesting the formation of a cysteine sulfenic acid. Although the reversible 2-electron/2-proton reduction is the most common reaction manifold for disulfide bonds, they can also undergo reversible conversion into a sulfenic acid-thiol pair through hydrolysis (**Figure 2.8c**).[19, 28] Whether through redox chemistry or hydrolysis, the reactivity of disulfides is thought to depend on mechanical

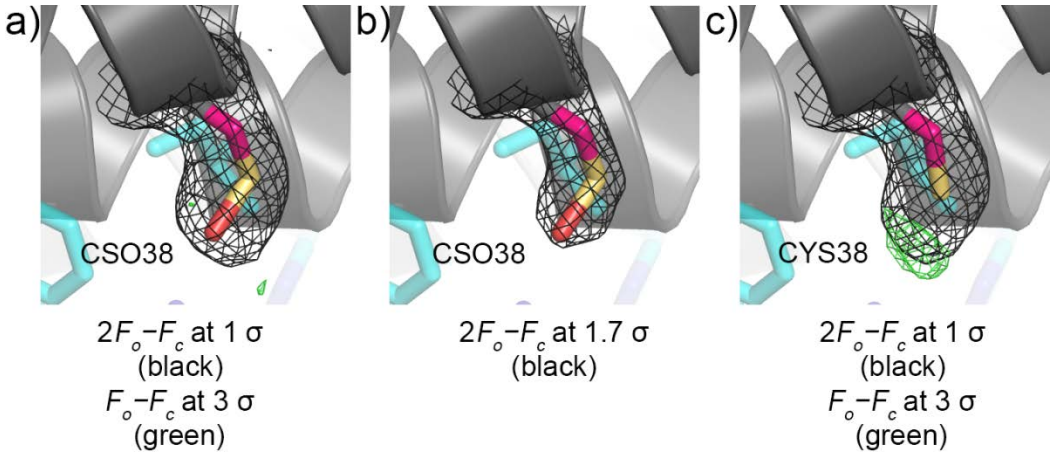

Figure 2.12 Electron density maps of alternative models for CSO38 in apo-$^{C38/C81/C96}$R1$_4$. Electron density maps highlighting residue 38 modeled as a cysteine sulfenic acid (CSO) or a cysteine (CYS) in the apo-$^{C38/C81/C96}$R1 structure. The $2F_o-F_c$ maps are shown in black, and are contoured at 1 σ, 1.7 σ, and 1 σ in panels a, b, and c, respectively. The F_o-F_c maps are shown in green, and contoured at 3 σ in panels a and c.

strain.[29] However, based on recent work with model compounds, it has been suggested that disulfide reactivity may instead be correlated with increased solvent accessibility.[30, 31] An analysis of the relative accessible surface areas[32] of the disulfide bonds indicates that the C38-C38 disulfide bonds in Zn- and apo-$^{C38/C81/C96}R1_4$ are the least solvent-exposed (**Table 2.4**). These data suggest that it is the quaternary strain imposed by the $^{C38/C81/C96}R1_4$ scaffold on the C38-C38 disulfide bonds that poises one of them to undergo hydrolytic cleavage upon Zn removal.

In this chapter, we discuss the *de novo* construction of a protein complex, $^{C38/C81/C96}R1_4$, through a metal-templating strategy, whereby quaternary strain was built through the formation of additional disulfide bonds across protein interfaces. This strain has two important consequences: first, it has enabled the modulation of metal coordination thermodynamics exclusively through outer sphere interactions and led to an destabilization of metal binding by 20 kJ/mol. This stands in contrast to most examples in molecular design where metal coordination and reactivity are exclusively handled at the primary and secondary coordination spheres.[6, 9] Second, the increased strain has allowed metal coordination to be directly and remotely linked to the formation/breakage of a distinct disulfide bond within the quaternary scaffold, thereby creating an allosteric system. Individually, metal coordination and disulfide bond formation are commonly employed as allosteric effectors in biological systems.[11, 19, 20, 33] The $^{C38/C81/C96}R1_4$ scaffold represents a unique system that links these two effectors to one another. More generally, this system illustrates the power of supramolecular protein design in creating artificial biological systems with coupled and externally controllable functions.

Table 2.4 Relative accessible surface areas of cysteine residues in $^{C38/C81/C96}R1_4$.
Reported values are averaged over disulfide pairs within each structure, with the exception of the single C38-C38 disulfide in apo-$^{C83/C81/C96}R1_4$. Values were calculated using the ASAView server (http://www.abren.net/asaview/).

	Relative accessible surface areas	
Disulfide site	**$^{C38/C81/C96}R1_4$**	**Zn-$^{C38/C81/C96}R1_4$**
C38-C38	0.08	0.10
C81-C81	0.21	0.29
C96-C96	0.14	0.21

2.3 Materials and methods

Protein sample preparation. All constructs used were housed in a pet20b vector (Novagen), and coding sequences included the N-terminal periplasmic localization sequence of *R. palustris* cytochrome c_{556}.[34] The A38C and L38C mutations were introduced into vectors (Novagen) harboring the genes encoding $^{C81/C96}$R1 and $^{C81/C96}$M1, respectively, by site-directed mutagenesis using oligonucleotide primers obtained from Integrated DNA Technologies. The primer pair 5'-cgctccacgcatcgcacgctgcggccgc-3' and 5'-ggcggccgcagcgtgcgatgcgtggagcg-3' was used with the pET20b/$^{C81/C96}$R1 template to yield the gene encoding $^{C38/C81/C96}$R1, and the primer pair 5'-gcgcgccgcagcgtgcgatgcgtggagcg-3' and 5'- cgctccacgcatcgcacgctgcggcgcgc-3' was used with the pET20b/$^{C81/C96}$M1 template to yield the gene encoding $^{C38/C81/C96}$M1. DNA amplification was carried out using PfuTurbo DNA Polymerase (Agilent Technologies). The amplified DNA was transformed into XL-1 Blue competent *E. coli* cells (Agilent Technologies), and the cells were grown at 37 °C on lysogeny broth (LB) agar plates containing 60 μg/mL ampicillin. Plasmids were purified with the QIAprep Spin Miniprep Kit (QIAGEN), verified by sequencing (Retrogen), and then transformed into BL21(DE3) *E. coli* cells (New England Biolabs) housing the cytochrome c maturation (ccm) plasmid cassette.[35] The cells were grown at 37 °C on lysogeny broth (LB) agar plates containing 34 μg/mL chloramphenicol and 60 μg/mL ampicillin. Single colonies were used to inoculate 5 mL of liquid LB medium containing chloramphenicol and ampicillin. Cultures were shaken at 37 °C until visibly turbid (typically about 8 h, or to an OD_{600}>0.6) at which time 50 μL of inoculum was transferred to 1 L of LB medium, typically to 15 cultures in parallel, and shaken for 16-20 h at 37 °C with protein

expression occurring by auto-induction. Cells were harvested by centrifugation (4,000 rpm, 4 °C, 10 min), resuspended in 100 mL of 5 mM sodium acetate (pH 5), frozen, thawed, and sonicated for 12 min in pulses of 30 s on and 60 s off in the presence of lysozyme and excess DTT while on ice. The lysate was titrated with sodium hydroxide to pH >10 and acetic acid to pH 5, and cleared by centrifugation (10,000 rpm, 4 °C, 10 min). The red cleared lysate was decanted and diluted to 2 L in 5 mM sodium acetate (pH 5) containing 2 mM DTT, and was manually applied to a CM Sepharose column (GE Healthcare) equilibrated in assembly buffer. The sample was washed with 5 mM sodium acetate containing 2 mM DTT (pH 5), and eluted in a 0-1 M gradient of NaCl. The clear red eluate was concentrated and buffer-exchanged into 10 mM sodium phosphate (pH 8) containing 2 mM DTT using a Diaflow concentrator (Amicon) fitted with a 3-kDa cutoff membrane. The sample was loaded onto a DuoFlow fast protein liquid chromatography station fitted with a Macroprep High Q-cartridge column (BioRad) with 10 mM sodium phosphate (pH 8) running buffer containing 2 mM DTT, and eluted using a 0-0.5M NaCl gradient. Fractions with Reinheitzahl values (A_{421}/A_{280}) above 3 were combined and concentrated, and excess EDTA and DTT (>10 molar equivalents) was added to the protein samples to remove any bound metal ions and prevent disulfide formation. These stock solutions were then flash frozen in liquid nitrogen for storage at –80 °C.

Tetramer self-assembly and purification. Concentrated stock solutions of $^{C38/C81/C96}$R1 (or $^{C38/C81/C96}$M1) were thawed, reduced by the addition of excess DTT, transferred to glass vials sealed with rubber septa, cycled under vacuum and argon atmosphere, and transferred to a glove box (Coy Lab) under an anaerobic (<10 ppm O_2) atmosphere of argon with 10% hydrogen. Here, the samples were exchanged into an

assembly tris(hydroxymethyl)aminomethane-, or Tris-buffered solution (50 mM Tris, pH 7, and 150 mM NaCl) using Econo-Pac 10DG pre-packed columns (BioRad). Protein concentration was determined spectrophotometrically ($\varepsilon_{415,\ ox}$ = 148,000 M^{-1} cm^{-1}),[34] and samples were diluted to a final concentration of 50 μM in assembly buffer supplemented with 75 μM $ZnCl_2$. Samples were incubated overnight under ambient atmosphere at 37 °C with shaking. Self-assembly reactions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the absence of added reductant to keep disulfide bonds intact. The crude self-assembly reaction mixture was applied to a preparative scale Superdex 75 gel filtration column (GE Healthcare) equilibrated in assembly buffer. Purities of the tetramer-containing fractions were assessed by non-reducing SDS-PAGE. The purest fractions were combined, concentrated, and treated with >10-fold excess EDTA to remove bound metal, and applied to an Econo-Pac 10DG prepacked column equilibrated in 20 mM MOPS (3-(N-morpholino)propanesulfonic acid) (pH 7) and 150 mM NaCl. The resulting stock solutions of Zn-free tetramer were flash frozen in liquid nitrogen, and stored at –80 °C.

*Sedimentation velocity analytical ultracentrifugation***.** Solutions of 1.25 μM tetramer ($^{C38/C81/C96}R1_4$) in 20 mM MOPS (pH 7) and 150 mM NaCl were treated with either 5 μM $ZnCl_2$ or 1 mM EDTA to prepare metallated or apo samples. Sedimentation velocity measurements were made on a XL-1 Analytical Ultracentrifuge (Beckman-Coulter) equipped with an An-60 Ti rotor at 41,000 rpm for 400 scans at 25 °C, and monitored at 415 nm. The endpoint of sedimentation was determined using the match function in HeteroAnalysis (http://biotech.uconn.edu/auf/). Scans were processed in Sedfit[36] using buffer viscosity (0.01002 poise), density (1.007 g/mL) and partial specific

volume (0.7316 ml/g) parameters calculated by SEDNTERP (http://sednterp.unh.edu/). The final c(S) distributions are reported at a confidence level of 0.95 in **Figure 2.5**.

Competitive Zn^{II} titration assay. The concentration of a light-protected stock solution of Fura-2 (Invitrogen) was determined spectrophotometrically (ε_{362} = 27,000 M^{-1} cm^{-1}).[37] Samples of $^{C38/C81/C96}R1_4$ and Fura-2 were prepared at a final concentration of 7.5 μM in a buffered solution of 20 mM MOPS (pH 7) and 150 mM NaCl treated with Chelex 100 resin (BioRad). The sample was titrated with a $ZnCl_2$ solution while thermostatted at 22 °C, and fluorescence measurements were made after 5-min equilibration periods. Fura-2 fluorescence emission at 510 nm was monitored to obtain an excitation scan over 250-450 nm on a Horiba Fluorolog 2 fluorimeter (**Figure 2.13**). Binding isotherms were generated from the changes in emission intensity plotted as a function of Zn^{II} concentration, and were fit using Dynafit[38] as previously reported.[24, 25] Briefly, we fit the four Zn-binding equilibria using three different models: a single dissociation constant model (or 4 × 1) where all sites were treated as equivalent, a two dissociation constant model (or 2 + 2) with two inequivalent pairs of binding sites, or a four dissociation constant model (or 1+1+1+1) with each site treated as inequivalent. We found the two dissociation constant model (2 + 2) adequately described the data. The measured dissociation constants and total ΔG_{Zn} are reported in **Table 2.1**.

X-ray crystallography and structural analysis. For crystallization of apo-$^{C38/C81/C96}R1_4$, a stock of 700 μM tetramer was prepared in 20 mM MOPS, pH 7, with 150 mM NaCl. For crystallization of Zn-$^{C38/C81/C96}R1_4$, a 375 μM stock of apo-protein in assembly buffer (20 mM Tris, pH 7, with 150 mM NaCl) was prepared and preincubated

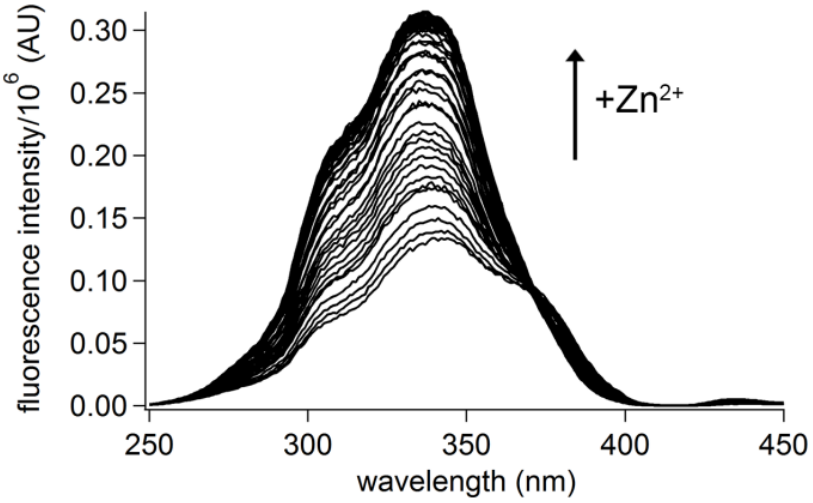

Figure 2.13 Fluorescence scans of Zn^{II} titrated into a solution of Fura-2 and $^{C38/C81/C96}R1_4$.
Competitive Zn^{II} binding titration of $^{C38/C81/C96}R1_4$ in the presence of Fura-2. Zn competition was monitored by excitation scans at a fixed emission wavelength of 510 nm. The corresponding changes in the intensity at 335 nm (shown above) were used to generate the Zn binding isotherms shown in **Figure 2.6**

with 2250 μM $ZnCl_2$. Screens were carried out by sitting-drop vapor diffusion at room temperature with 500 μL reservoirs and droplets consisting of 1.5 μL of protein and 1 μL of precipitant solution. Crystals of metal-free $^{C38/C81/C96}R1_4$ were obtained with a precipitant solution consisting of 0.1 M Bis-Tris (pH 6.5) and 45% 2-methyl-2,4-pentanediol, and crystals of Zn-$^{C38/C81/C96}R1_4$ were obtained with a precipitant consisting of 0.2 M $MgCl_2$, 0.1 M Bis-Tris (pH 6.5), and 35% PEG 400. Crystals were harvested, cryoprotected in perfluoropolyether cryo oil (Hampton Research), and stored in liquid nitrogen. X-ray diffraction data for apo-$^{C38/C81/C96}R1_4$ and Zn-$^{C38/C81/C96}R1_4$ were collected at 100 K on beamlines BL9-2 and BL14-1, respectively, with 0.99 Å radiation at the Stanford Synchrotron Radiation Lightsource. Diffraction data were integrated using Web-Ice.[39] Integrated datasets were scaled using SCALA[40] within the CCP4 suite. Molecular replacement was carried out using Phaser[41] with a $^{C96}R1$ monomer (PDB ID: 3IQ6) as the search model. Data refinement and model building were carried out using REFMAC[42] and COOT[43]. Four-fold non-crystallographic symmetry restraints were applied in the initial stages of refinement and gradually relaxed. Ramachandran plots were calculated using PROCHECK.[44] All structure figures and structural alignments were generated using PyMOL (www.pymol.org).

Isothermal titration calorimetry. Experiments were carried out on a VP-ITC instrument (MicroCal) with a cell volume of 1.4125 mL. Samples of $^{C96}R1_4$, $^{C81/C96}R1_4$ and $^{C38/C81/C96}R1_4$ were buffer-exchanged into a solution of 20 mM MOPS (pH 7) and 150 mM NaCl previously treated with Chelex-100 Resin (BioRad). Protein samples were then dialyzed twice at 4 °C for >8 h against the indicated buffer, with the second dialysis buffer retained for diluting protein samples and preparing $ZnCl_2$ titrant solutions. The

calorimeter cell contained 25-50 μM of apo-tetramer and a water reference. The cell was equilibrated at 22 °C for 180 s before an initial injection of 1 μL of 1.4 mM $ZnCl_2$ titrant, followed by a schedule of 30 injections of 5.5 μL $ZnCl_2$ delivered at 5 min intervals. Data analysis was carried out using the ORIGIN data analysis plugin from MicroCal. The initial injection was omitted from data analysis. Integrated data were background-subtracted using the final injections of the thermogram to account for metal dilution effects, and were fit to a four-site sequential binding model. ITC thermograms were fit using dissociation constants determined by fluorescence competition assays, and enthalpy values were allowed to vary. The dissociation constants could not be directly determined by ITC owing to the tight binding constants (requiring ~40 μM protein to measure the weakest binding events, and even lower concentrations for tighter binding) and low enthalpies of binding (requiring ~50 μM binding site to achieve 1 μcal/injection). Therefore, the reported experimental conditions gave a value of $c > 1000$ (where c is the concentration of binding sites divided by the dissociation constant). This limitation has also been reported in other metalloprotein systems with high metal-binding binding affinities, where an ancillary experiment (*e.g.*, equilibrium dialysis) can be used instead to determine the dissociation constants.[45] As previously noted,[45] the use of two independent experiments has the added benefit of avoiding correlated errors in determining ΔG and ΔH from measuring both parameters by ITC. Finally, we note that the complexity of Zn^{II} binding to the $R1_4$ scaffolds (four ΔH_{ITC}'s and four K_d's, giving eight free parameters in total) necessitate a complex model that may not robustly measure these binding parameters on a per-site basis. Therefore, we employed an analysis of the global binding parameters supported by additional experiments. Representative raw and

integrated thermograms, as well as their best-fit curves, are reported in **Figure 2.7**, and the measured ΔH_{ITC} values in **Table 2.1**.

Chapter 2 was reproduced in part, with permission, from Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A. "*De novo* design of an allosteric metalloprotein." *J. Am. Chem. Soc.* **2016**, *138*, 13163–13166. Copyright 2016, American Chemical Society.

The dissertation author is primary author on all reprinted materials.

References

(1) Lippard, S.; Berg, J. *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, 1994.

(2) Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valentine, J. S. *Biological Inorganic Chemistry, Structure & Reactivity*; University Science Books: Sausalito, 2007.

(3) Wittung-Stafshede, P. *Accounts of Chemical Research* **2002**, *35*, 201.

(4) Barker, P. D. *Current Opinion in Structural Biology* **2003**, *13*, 490.

(5) Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A. *Annual Review of Biophysics* **2014**, *43*, 409.

(6) Cook, S. A.; Hill, E. A.; Borovik, A. S. *Biochemistry* **2015**, *54*, 4167.

(7) Que, L.; Tolman, W. B. *Nature* **2008**, *455*, 333.

(8) Zastrow, M. L.; Pecoraro, V. L. *Coordination Chemistry Reviews* **2013**, *257*, 2565.

(9) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M. *Nature* **2009**, *460*, 855.

(10) Yuan, Y.; Tam, M. F.; Simplaceanu, V.; Ho, C. *Chemical Reviews* **2015**, *115*, 1702.

(11) Reyes-Caballero, H.; Campanello, G. C.; Giedroc, D. P. *Biophysical Chemistry* **2011**, *156*, 103.

(12) Cui, Q.; Karplus, M. *Protein Science* **2008**, *17*, 1295.

(13) Changeux, J.-P. *Annual Review of Biophysics* **2012**, *41*, 103.

(14) Goodsell, D. S.; Olson, A. J. *Annual Review of Biophysics and Biomolecular Structure* **2000**, *29*, 105–153.

(15) Raman, S.; Taylor, N.; Genuth, N.; Fields, S.; Church, G. M. *Trends in Genetics* **2014**, *30*, 521.

(16) Choi, J. H.; Laurent, A. H.; Hilser, V. J.; Ostermeier, M. *Nature Communications* **2015**, *6*, 6968.

(17) Ha, J.-H.; Loh, S. N. *Chemistry* **2012**, *18*, 7984.

(18) Vallee-Belisle, A.; Plaxco, K. W. *Current Opinion in Structural Biology* **2010**, *20*, 518.

(19) Hogg, P. J. *Trends in Biochemical Sciences* **2003**, *28*, 210.

(20) Messens, J.; Collet, J. F. *Antioxidant & Redox Signaling* **2013**, *18*, 1594.

(21) Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A. *Journal of the American Chemical Society* **2007**, *129*, 13374–13375.

(22) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. *Methods in Enzymology* **2016**, *580*, 223.

(23) Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A. *Proceedings of the National Academy of Sciences U.S.A* **2010**, *107*, 1827–1832.

(24) Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A. *Journal of the American Chemical Society* **2010**, *132*, 8610–8617.

(25) Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A. *Journal of the American Chemical Society* **2013**, *135*, 12013.

(26) Campanello, G. C.; Ma, Z.; Grossoehme, N. E.; Guerra, A. J.; Ward, B. P.; Dimarchi, R. D.; Ye, Y.; Dann, C. E.; Giedroc, D. P. *Journal of Molecular Biology* **2013**, *425*, 1143.

(27) Katz, B. A.; Kossiakoff, A. *Journal of BIological Chemistry* **1986**, *261*, 15480.

(28) Kettenhofen, N. J.; Wood, M. J. *Chemical Research in Toxicology* **2010**, *23*, 1633.

(29) Vogel, V. *Annual Review of Biophysics and Biomolecular Structure* **2006**, *35*, 459.

(30) Dopieralski, P.; Ribas-Arino, J.; Anjukandi, P.; Krupicka, M.; Kiss, J.; Marx, D. *Nature Chemistry* **2013**, *5*, 685.

(31) Kucharski, T. J.; Huang, Z.; Yang, Q. Z.; Tian, Y.; Rubin, N. C.; Concepcion, C. D.; Boulatov, R. *Angewandte Chemie International Edition* **2009**, *48*, 7040.

(32) Ahmad, S.; Gromiha, M.; Fawareh, H.; Sarai, A. *BMC Bioinformatics* **2004**, *5*, 51.

(33) Philips, S. J.; Canalizo-Hernandez, M.; Yildirim, I.; Schatz, G. C.; Mondragon, A.; O'Halloran, T. V. *Science* **2015**, *349*, 877.

(34) Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R. *Biochemistry* **2006**, *45*, 10504.

(35) Arslan, E.; Schulz, H.; Zufferey, R.; Kunzler, P.; Thony-Meyer, L. *Biochemical and Biophysical Research Communications* **1998**, *251*, 744.

(36) Schuck, P. *Biophysical Journal* **2000**, *78*, 1606.

(37) Grynkiewicz, G.; Poenie, M.; Tsien, R. Y. *Journal of Biological Chemistry* **1985**, *260*, 3440.

(38) Kuzmic, P. *Analytical Biochemistry* **1996**, *237*, 260.

(39) Gonzalez, A.; Moorhead, P.; McPhillips, S. E.; Song, J.; Sharp, K.; Taylor, J. R.; Adams, P. D.; Sauter, N. K.; Soltis, S. M. *Journal of Applied Crystallography* **2008**, *41*, 176.

(40) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta crystallographica Section D: Biological crystallography* **2011**, *67*, 235.

(41) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Journal of Applied Crystallography* **2007**, *40*, 658.

(42) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallographica Section D: Biological Crystallography* **1997**, *53*, 240.

(43) Emsley, P.; Cowtan, K. *Acta Crystallographica Section D: Biological Crystallography* **2004**, *60*, 2126.

(44) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. *Journal of Applied Crystallography* **1993**, *26*, 283.

(45) DiTusa, C. A.; Christensen, T.; McCall, K. A.; Fierke, C. A.; Toone, E. J. *Biochemistry* **2001**, *40*, 5338.

Chapter 3: Biochemical dissection of an engineered allosteric protein

3.1 Introduction

The grand challenge posed to protein engineering is to create systems that can rival, and even surpass, those of natural proteins. A large portion of naturally-occurring proteins function as oligomers,[1, 2] which is thought to grant access key functional advantages, such as coordinated action of distal regions in a protein scaffold through allosteric function.[3] To this end, a number of groups have explored strategies for achieving designed supramolecular protein assemblies,[4] including protein-DNA composite lattices,[5] self-assembly of fused oligomers,[6] computationally-designed protein cages,[7, 8] and disulfide-linked nanomaterials. [9] These strategies have made great strides in

their predictive accuracy, leading to the design of protein assemblies of varying scales and complexities. However, less attention has been given to strategies that can generate dynamic and responsive protein assemblies. This latter case of designed system successfully captures a key functional aspect of Nature's proteome, given the growing appreciation of protein dynamics as enabling allosteric function[10] or efficient catalytic function[11].

Our lab has employed a metal-centered synergistic design strategy that uses interfacial chelating motifs[12] to enable the subsequent design of self-assembling protein complexes.[13, 14] The strength and geometric preferences of metal coordination drive the formation of new protein-protein interactions, termed MDPSA.[14] We have employed MDPSA to as a first step toward the design oligomeric metalloprotein assemblies[15] and metal-mediated nanomaterials[16, 17] of the monomeric protein cytochrome cb_{562} .[18] The small footprint interfacial of the interfacial chelating sites allows for further design of the nascent interfaces to form favorable protein-protein contacts, termed MeTIR (see **Figure 1.11**).[13, 14] Applying MDPSA and MeTIR to cyt cb_{562} afforded a variant, R1, that can bind Zn^{II} to form the D_2-symmetric tetramer Zn-$R1_4$. Each protomer forms contacts along three C_2-symmetric axes: one engineered interface (*i1*) and two additional non-designed interfaces (*i2* and *i3*)[13] (**Figure 3.1**). This design approach has also allowed for the introduction of such as designed disulfide crosslinks [19, 20] and interfacial active sites.[21]

As discussed in **Chapter 2**, the MDPSA and MeTIR approaches have been extended to the design of an allosteric metalloprotein assembly, Zn-$^{C38/C81/C96}R1_4$. This homotetrameric protein assembly contains four core Zn^{II} binding sites and six peripheral disulfide bonds, with one disulfide pair spanning each of the three protein-protein inter-

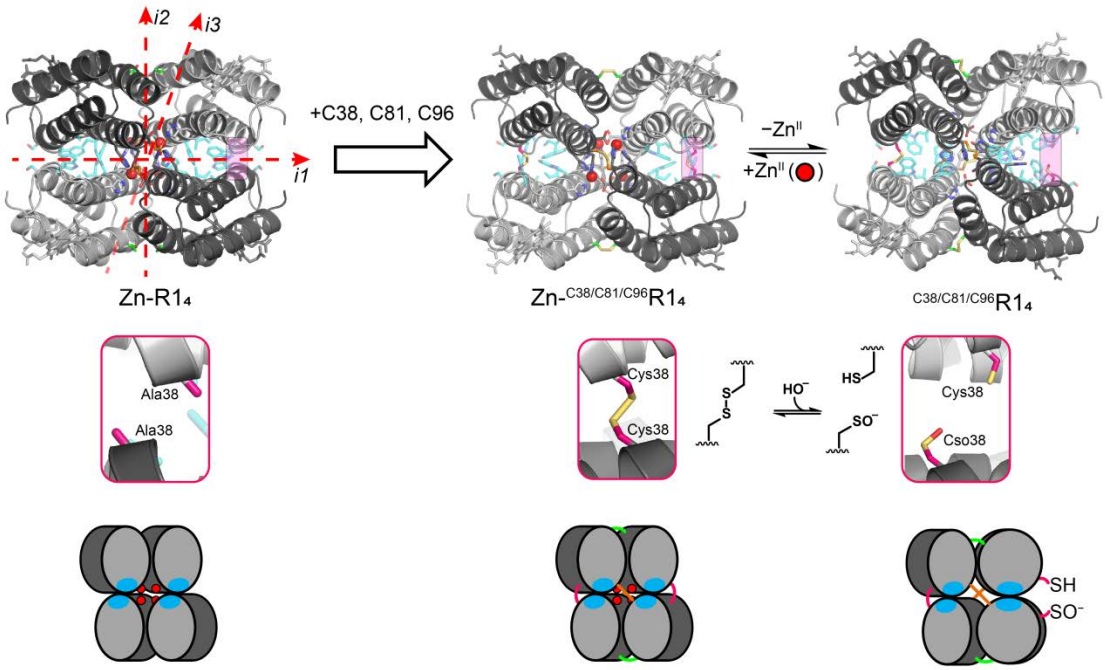

Figure 3.1 Interfaces and disulfide crosslinks in the allosteric protein $^{C38/C81/C96}$R1$_4$.
The Zn-R1$_4$ tetramer bears three distinct interfaces. Installing pairs of cysteine residues at these interfaces affords the hextuply disulfide bonded Zn-$^{C38/C81/C96}$R1$_4$ complex. Notably, upon removal of Zn^{II}, one member of the C38-C38 crosslinks located in the i1 interface is observed to cleave, presumably by hydrolysis.

faces (**Figure 3.1**). When Zn^{II} was removed from the protein, one member of the C38-C38 disulfide pair spanning the *i1* interface was cleaved, presumably via hydrolysis, to yield a thiol-sulfenic acid pair. This bond cleavage event is remotely coupled to Zn^{II} binding, and occurs over 14 Å away from the nearest metal binding site. The net result of Zn^{II} removal and bond breakage is significant structural rearrangement of the tetrameric scaffold, including relaxation of the remaining disulfide bonds and rearrangement of the *i1* interface. The apo-$^{C38/C81/C96}R1_4$ protein was competent to bind Zn^{II}, but exhibited a 5 kJ/mol/site penalty to the free energy of binding relative to $^{C81/C96}R1_4$ (see **Table 2.1**).

In the present study, we sought to understand the structural requirements for the allosteric behavior of $^{C38/C81/C96}R1_4$. We generated cysteine-deleted analogs of $^{C38/C81/C96}R1$ that still bear C38 residues, $^{C38}R1$ and $^{C38/C96}R1$, and examined their (dis)assembly behavior in the presence and absence of Zn^{II}. These variants allowed us to examine whether C38-C38 hydrolysis is an intrinsically fragile bond that breaks upon Zn^{II} release, or if there is a larger structural requirement that leads to the emergence of the observed bond breakage, and to probe the origin of the decrease in metal binding affinity of $^{C38/C81/C96}R1_4$. Collectively, these experiments show that C38-C38 hydrolysis only arises in the context of the crosslinked tetrameric architecture. Additionally, it is the extent of crosslinking in $^{C38/C81/C96}R1_4$, rather than the presence of C38-C38 disulfide bonds, that can be attributed to the penalty in Zn^{II} binding affinity, and which gives rise to other non-additive effects on tetramer stability and the rate of Zn^{II} release. This suggests that rearrangement of a single *i1* interface provides insufficient driving force for C38-C38 disulfide bond hydrolysis, which is only observed in constructs couple two *i1* dimers to one another. These results show how the function of dynamic and responsive

protein features can be coordinated, giving rise to emergent functional properties. We anticipate that the improved understanding of this designed system will inform future efforts to engineer functional protein assemblies that approach the complexity and sophistication of their counterparts in nature.

3.2 Results and Discussion

Protein scaffold design and preliminary characterization. The net result of Zn^{II} removal from Zn-$^{C38/C81/C96}R1_4$ is C38-C38 bond scission and concomitant structural rearrangements. However, the causal relationship of these changes is not clear from the available crystal structures of the protein. For example, C38-C38 hydrolysis could arise directly from the Zn-induced structural changes observed in the dimeric interface of the parent R1 scaffold's metal-free (PDB 3HNK) and Zn-bound (PDB 3HNI) structures.[13] Here, Zn^{II} release increases the α-C separation of residue A38 from 5.0 Å to 12.2 Å, outside a reasonable disulfide bonding distance.[22] Similar structural rearrangements of A38 are observed for Zn^{II} removal from $^{C96}R1_4$ and $^{C81/C96}R1_4$,[19, 20] as well as for the broken C38-C38 disulfide bond of $^{C38/C81/C96}R1_4$ (**Table 3.1**). The structural rearrangement across the dimeric interface could render the associated C38-C38 disulfide bond susceptible to hydrolysis. However, from this explanation alone, it is not clear why only a single C38-C38 disulfide bond hydrolyzes in $^{C38/C81/C96}R1_4$. The breakage of one, rather than both members of the pair, suggests that this allosteric behavior is a global phenomenon that arises from the tetrameric architecture in its entirety, rather than solely from local changes experienced by a disulfide bond in its embedded interfaces. In order understand the dependence of C38-C38 disulfide scission on the structural context of $^{C38/C81/C96}R1_4$, we generated the protein variants $^{C38/C96}R1$ and $^{C38}R1$. Zn-mediated assem-

Table 3.1 C-α separations of residue 38 pairs in available crystal structures of R1 variants.
Values are the measured distance between the C-α atoms of pairs located at the *i1* interface of the indicted R1 variants. Values are averages of all residue pairs within the assymmetric unit of the crystal strurcture, except in the case of metal-free $^{C38/C81/C96}$R1. ‡Reported value corresponds to the distance between the residues in the broken C38-C38 disulfide bond, and the parenthetical value corresponds to the distance of the intact disulfide bond.

		C-α separations (Å)	
Protein Variant	Residue	Zn-bound	Metal-free
R1	A38	5.0	12.2
^{C96}R1	A38	5.1	11.2
$^{C81/C96}$R1	A38	7.9	10.6
$^{C38/C81/C96}$R1	C38	5.5	10.7 (5.7) ‡

bly of these variants was expected to yield tetramers that were isotopological to $Zn^{C38/C81/C96}R1_4$ (and one another) but with pairs of disulfide bonds selectively deleted (**Figure 3.2**). Additionally, due to the low extent of crosslinking, $^{C38}R1$ was expected to undergo partial disassembly upon removal of Zn^{II}, and to dissociate even further in the event of C38-C38 disulfide bond breakage (**Figure 3.2**).

We first set out to isolate the disulfide-linked species of interest for $^{C38}R1$ and $^{C38/C96}R1$. Since $^{C38/C96}R1$ bears more than one surface-exposed cysteine, it was possible for undesirable mixed disulfide bonds to form. Therefore, the $^{C38/C96}R1$ protein was isolated under reducing conditions for subsequent *in vitro* templating. Oxidative self-assembly of $^{C38/C96}R1$ showed a modest enrichment for the tetrameric species (27 ± 2%), which was further enhanced in the presence of Zn^{II} (34 ± 2%) (**Figure 3.3** and **Table 3.2**). Oxidation of a variant lacking the engineered *i1* contacts showed decreased tetramer formation in the absence or presence of one equivalent of Zn^{II} (6 ± 2% and 16 ± 1%, respectively), suggesting that the engineered protein contacts were involved in $^{C38/96}R1_4$ assembly (**Figure 3.3** and **Table 3.2**). The $^{C38/C96}R1_4$ species was isolated by size-exclusion chromatography, and the ~50 kDa species observed by non-reducing SDS-PAGE dissociated into a ~12 kDa species in the presence of excess DTT, consistent with formation of a monomer (**Figure 3.4**). The $^{C38}R1$ variant bears a single cysteine residue, and therefore the formation of the correct disulfide bonds is site-specific by default. Since the protein was localized to the periplasm of *E. coli* during expression, we were able to isolate disulfide-linked species $^{C38}R1_2$ without the need for *in vitro* templating by omitting reductant from the purification (**Figure 3.5**).

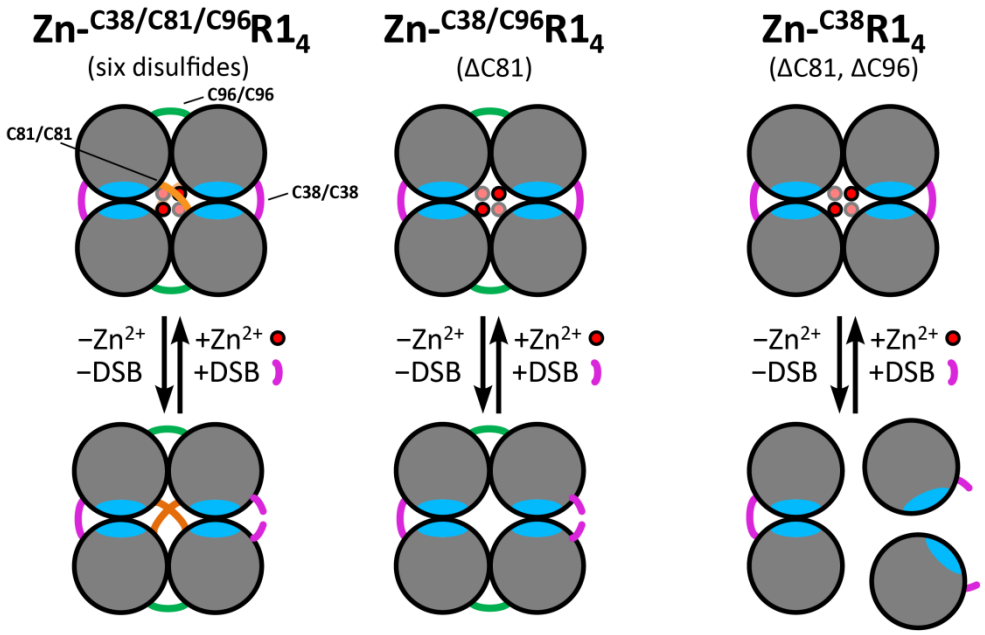

Figure 3.2 Hypothetical speciation of C38-linked R1 variants following Zn-removal.
Cartoon representations Zn-induced tetramers of C38-bearing R1 variants are shown with each monomer represented as a gray circle, and Zn^{II} represented as a red circle. The engineered *i1* residues are shown as cyan patches. Pink lines represent C38 residues, while C81 and C96 are shown as orange and green lines, respectively. The expected speciation resulting from removal of all Zn^{II} ions and breakage of a single C38-C38 disulfide bond is depicted for each construct.

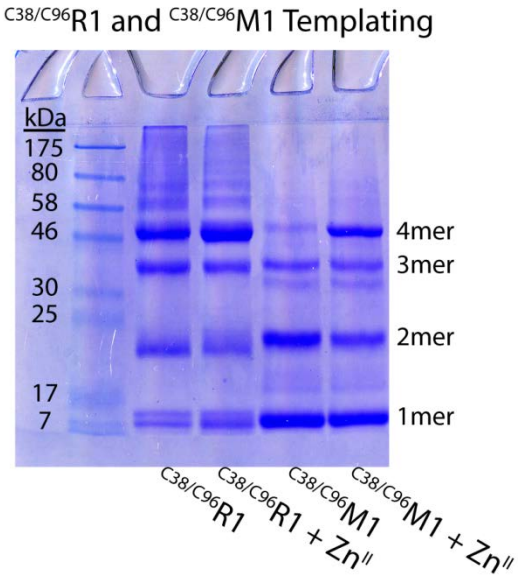

Figure 3.3 Representative templating reactions of $^{C38/C96}$R1 and $^{C38/C96}$M1.

Table 3. 2 Speciation of templating reactions of $^{C38/C96}$R1 and $^{C38/C96}$M1.
Errors denote standard deviations from triplicate measurements of relative band intensities in Coomassie-stained non-reducing SDS-PAGE gels.

	$^{C38/C96}$R1	
	Zn-free	$+Zn^{II}$
1mer	15.6 ± 0.2%	20 ± 2%
2mer	20 ± 5%	18 ± 1%
3mer	17.1 ± 0.2%	13.3 ± 0.5%
4mer	27 ± 2%	34 ± 2%
n-mer	22 ± 4%	15 ± 1%

	$^{C38/C96}$M1	
	Zn-free	$+Zn^{II}$
1mer	44 ± 3%	43 ± 7%
2mer	33 ± 5%	20 ± 2%
3mer	15 ± 4%	16 ± 1%
4mer	6 ± 2%	19 ± 4%
n-mer	2 ± 3%	3 ± 4%

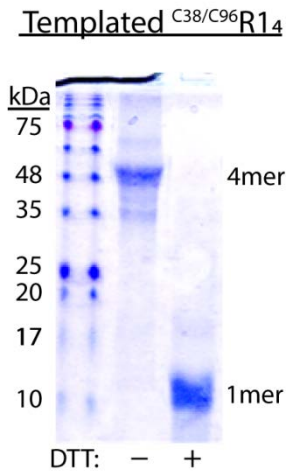

Figure 3.4 SDS-PAGE of isolated $^{C38/C96}R1_4$ in the presence and absence of reductant.

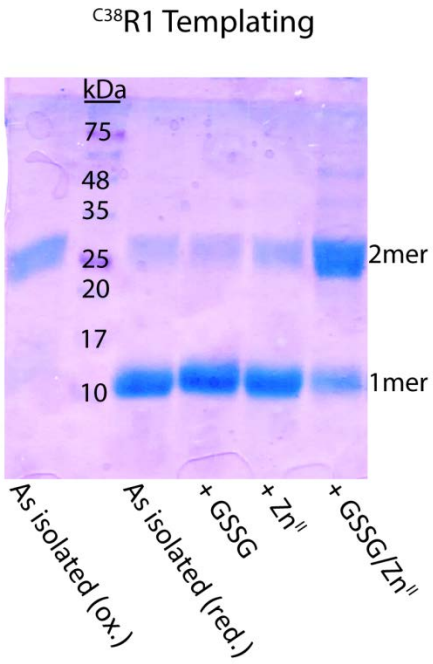

Figure 3.5 SDS-PAGE of ^{C38}R1 protein under varied templating conditions.
^{C38}R1 protein was isolated under oxidizing condtions (with no added reductant), or reducing conditions (in the presence of 2 mM DTT). Reduced protein was subjected to *in vitro* templating following removal of the DTT. Templating reactions were incubated overnight at 37 °C with shaking.

We also purified ^{C38}R1 in the presence of reductant to explore its ability to undergo *in vitro* C38-C38 disulfide bond formation. When we removed the reductant and incubated samples under ambient atmosphere overnight, we did not observe the formation of appreciable disulfide-linked protein (**Figure 3.5**). Under these conditions, significant crosslinking of ^{C38}R1 was only observed in the presence of both excess Zn^{II} and oxidized glutathione (GSSG) (**Figure 3.5**). By contrast, $^{C38/C96}$R1 under comparable templating conditions was comparatively rapid, and the formation of trimeric, tetrameric and larger assemblies observed by SDS-PAGE (**Figure 3.3**) suggested that the C38 residues, in addition to the C96 residues, must be participating in this rapid self-assembly. The difference in crosslinking ability of C38 in the case of ^{C38}R1 self-assembly and $^{C38/C96}$R1 implies a difference in structural/contextual effect exerted by each protein. The previously-characterized structures of C38-lacking R1 variants provide some indication of how the C38-bearing variants would look upon *i1*-mediated self-assembly without forming C38-C38 crosslinks. In these structures, Ala38 displays modestly increased relative accessible surface area (ASA) in apo-^{C96}R1$_4$ (0.47) relative to apo-R1$_2$ (0.45). This difference in accessibility holds true when modeling in a C38 at this site as well, yielding different relative ASAs for the models R1 + C38 (0.33) and ^{C96}R1 + C38 (0.51) (**Table 3.3**). This structural effect is consistent with our understanding of the *i1* contacts constituting fluid interface,[13] and the observed differences in accessibility correlate with the observed differences in templating behavior. Moreover, they suggest that crosslinking elsewhere in the protein scaffold can influence the orientation of the proto-

Table 3.3 Relative accessibility of residue 38 in metal-free R1 and ^{C96}R1 structures and models.
Relative accessible surface areas of residue A38 in the Zn-free crystal structures of R1 and ^{C96}R1 are reported, along with the relative accessible surface area estimated for an uncrosslinked C38 residue modeled in these structures.

	Relative Accessible SA's	
	A38 structure	C38 model
apo-R1	0.45	0.33
apo-^{C96}R1	0.47	0.51

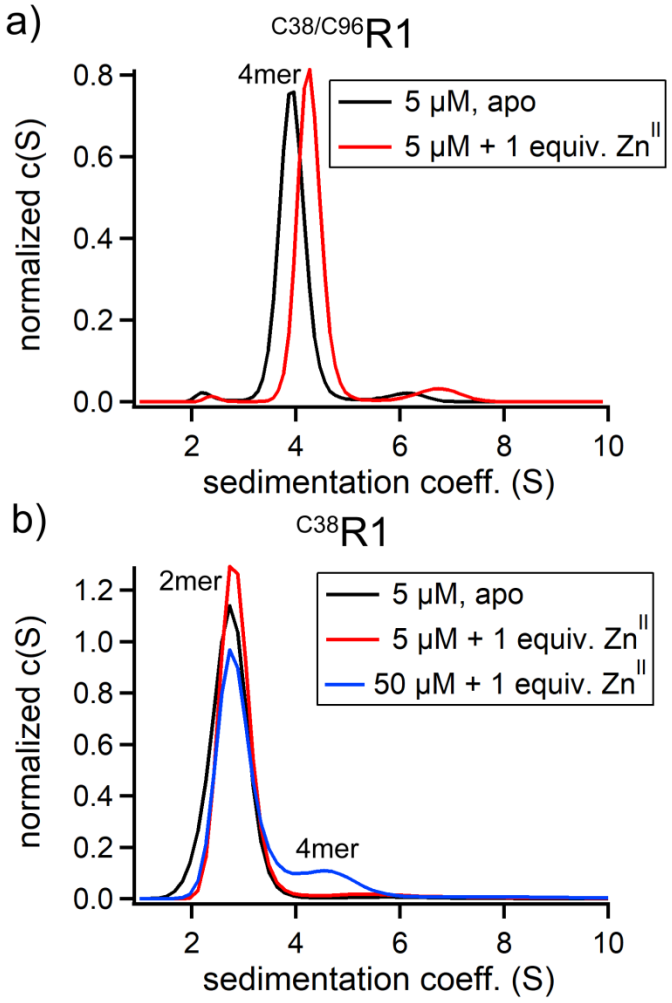

Figure 3.6 Sedimentation velocity profiles of templated ^{C38}R1 and $^{C38/C96}$R1.
Sedimentation velocity profiles of (a) $^{C38/C96}$R1 and (b) ^{C38}R1 protein in under metal-free conditions (black traces) or in the presence of stoichiometric Zn^{II} (red and blue traces). The data were normalized such that the total area under each curve is equal to one.

mers along their *i1*contact in such a way that modifies the accessibility (and reactivity) of the C38 residues embedded at this interface.

We determined the solution-phase oligomerization behavior of both variants through sedimentation-velocity analytical ultracentrifugation (AUC). The sedimentation velocity profile of $^{C38/C96}$R1 in the presence of EDTA confirmed the formation of a tetramer that persists in the absence of bound metal (S_{max} = 4.0 S) (**Figure 3.6a**). The sedimentation velocity profile showed little change in the presence of Zn^{II} (S_{max} = 4.3 S), consistent with a tetrameric species that undergoes minimal structural rearrangement (**Figure 3.6a**). The sedimentation-velocity profile of the isolated $^{C38}R1_2$ species gave a single symmetric peak (S_{max} = 2.7 S), suggesting that the protein was predominantly dimeric in solution (**Figure 3.6b**). Surprisingly, the sedimentation velocity profile of a low concentration of $^{C38}R1_2$ protein (5 μM monomer) in the presence of a stoichiometric amount of Zn^{II} was almost exclusively dimeric (S_{max} = 2.7 S) as well. At 50 μM ^{C38}R1 monomer in the presence of stoichiometric Zn^{II}, only ~15% of the sample is tetrameric (S_{max} = 4.5 S), which is much less than the observed abundance of Zn-$R1_4$ under similar experimental conditions.[13] These data reveal that the presence of a C38-C38 bond disfavors tetramer formation. Conversely, this implies that a covalently enforced a tetrameric structure, such as in $^{C38/C81/C96}R1_4$ or $^{C38/C96}R1_4$, is likely disfavored by the C38-C38 disulfide linkages within the complex.

Probing Zn^{II} binding affinity by competitive titrations. We next examined the metal-binding thermodynamics of the templated $^{C38}R1_2$ and $^{C38/C96}R1_4$ proteins. In order to estimate the Zn^{II} binding affinities of the two variants, we carried out competitive binding titrations of Zn^{II} to each protein in the presence of Fura-2 ($K_{d,Zn}$ = 5.7 nM)[19] as a

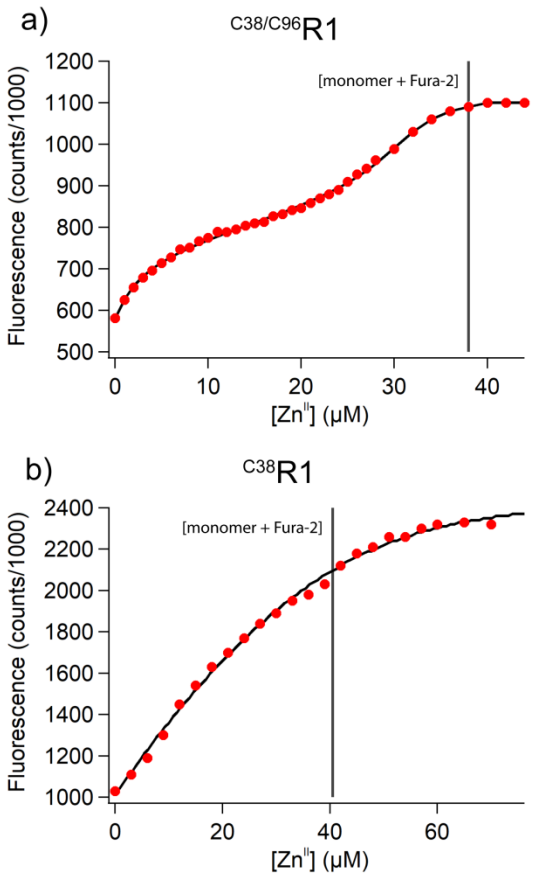

Figure 3.7 Binding isotherm of Zn^{II} to Fura-2 in the presence of templated C38-bearing R1 variants. Sample of templated (a) $^{C38/C96}$R1 and (b) ^{C38}R1 were mixed with Fura-2, and the solutions were titrated with Zn^{II}. The fluorescence intensity was monitored throughout the titration and plotted as a function of [Zn^{II}] (red circles). The data were fit to models where the individual sites (two for $^{C38}R1_2$ and four for $^{C38/C96}R1_4$) were fit to independent dissociation constants. Additionally, the superstoichiometric binding observed for titrating Zn^{II} into $^{C38}R1_2$ was accounted for with by including an additional apparent dissociation constant (K^*) with the binding site concentration modeled as equal to the concentration of monomer.

competing ligand. The Zn^{II} binding behavior of $^{C38/C96}R1_4$ showed the expected stoichiometry of one Zn^{II} per protomer, or four per tetramer (**Figure 3.7a**), that has been previously observed for disulfide-linked R1 tetramers.[19, 20] The data was well-described by a model with four different dissociation constants (K_d = 4 ± 2 nM; 3 ± 1 nM; 2 ± 1 nM; and 40 ± 10 nM) (**Figure 3.7a** and **Table 3.4**), which we used to estimate the per-site free energy of binding ($\Delta G_{Zn/site}$ = −47.0 ± 0.4 kJ mol^{-1} $site^{-1}$) (**Table 3.5**). The $^{C38}R1$ variant, by contrast, displayed appreciable non-specific binding beyond the expected stoichiometry of one Zn^{II} per protomer. We fit the data to a thermodynamic model with two different dissociation constants (K_d = 4.4 ± 0.8 nM; 12 ± 3 nM) and an apparent dissociation constant (K_{app} = 90 ± 40 nM) to capture the supersoichiometric binding (**Figure 3.7**). With this treatment, we estimated the per-site free energy of binding to $^{C38}R1_2$ as $\Delta G_{avg.}$ = −46.0 ± 0.2 kJ mol^{-1} $site^{-1}$ (**Table 3.5**).

Alongside previously the reported measurements,[19, 20] these data show that C38-C38 disulfide bond formation does not impair metal binding. For example, the free energy penalty for Zn^{II} binding to $^{C38/C81/C96}R1_4$ ($\Delta G_{avg.}$ = −41.0 ± 0.3 kJ/mol•site) relative to that of $^{C81/C96}R1_4$ ($\Delta G_{avg.}$ = 45.8 ± 0.3 kJ/mol•site)[20] and $^{C38/C96}R1_4$ cannot be attributed to the presence of the C38 disulfide bond. Rather, the extensive crosslinking afforded by all three pairs of disulfide pairs in $^{C38/C81/C96}R1_4$ seems to uniquely result in the ~20 kJ/mol decrease in overall binding affinity due to non-additive effects among the multiple disulfide pairs (**Table 3.5**). Additionally, these data reveal that $^{C38}R1_2$ binds Zn^{II} with comparable affinity to that of $^{C96}R1_4$ ($\Delta G_{avg.}$ = 46.5 ± 0.5 kJ/mol•site)[19] despite the decreased propensity to form a tetramer for $^{C38}R1_2$. Crucially, this implies that the formation of a tetrameric species perturbs the constituent dimers in the C38-linked $R1_2$

Table 3.4 Dissociation constants of Zn^{II} binding to dislufide-linked ^{C38}R1 and $^{C38/C96}$R1.

Binding affinities were determined by Zn^{II} titration in the presence of Fura-2. Errors denote fitting errors of a single measurement. *K** is the apparent dissociation constant that resulted from modelling the superstoichiometric Zn^{II} binding observed for $^{C38}R1_2$. All measurements were carried out at 22 °C and in the presence of 20 mM MOPS, pH 7 + 150 mM NaCl.

	^{C38}**R1$_2$**	$^{C38/C96}$**R1$_4$**
Binding affinities		
K_{d1} (nM)	4.4 ± 0.8	4 ± 2
K_{d2} (nM)	12 ± 3	3 ± 1
K_{d3} (nM)	-	2 ± 1
K_{d4} (nM)	-	40 ± 10
*K**(nM)	90 ± 40	-

Table 3.5 Average free energies of Zn^{II} binding to disulfide-linked R1 variants.
Binding affinities of Zn^{II} to disulfide-linked R1 variants. [a] Data taken from reference 22. [b] Data taken from reference 19.[c] Data taken from reference 20. Reported errors correspond to fitting errors of a single measurement.

Protein Variant	$\Delta G_{avg.}$ (kJ/mol•site)	$\Delta\Delta G_{avg.}$ (kJ/mol•site)
^{C38}R1	−46 ± 0.2	–
$^{C38/C96}$R1	−47 ± 0.4	−1
$^{C38/C81/C96}$R1	−41 ± 0.2[a]	+5
^{C96}R1	−46.5 ± 0.4 [b]	−0.5
$^{C81/C96}$R1	−45.8 ± 0.3[c]	−0.2

species, and implies some degree of structural communication between the dimers along their shared *i2* interface when assembled into a tetramer.

Measuring Zn^{II} release rates by stopped-flow mixing spectroscopy. We examined the effect of disulfide bonds on Zn^{II} binding lability in the disulfide-linked R1 tetramers: $^{C96}R1_4$, $^{C38/C81}R1_4$, $^{C81/C96}R1_4$, and $^{C38/C81/C96}R1_4$. The four variants contain chemically identical Zn^{II} binding sites and comparable Zn^{II} binding affinities (see above). Any difference in the kinetics of Zn^{II} release would, therefore, likely arise to some outer-sphere or global effects of the disulfide bonds on the scaffold as a whole. We pre-incubated each protein with stoichiometric amounts of Zn^{II}, and carried out asymmetric mixing with a vast excess (400 equivalents) of the metal-binding chromophore 4-(2-Pyridylazo)resorcinol (PAR).[23] The absorbance measurements were fit to pseudo first-order release kinetics with one or two release rates (**Table 3.6**). Overall, the Zn-release curves (**Figure 3.8**) reveal markedly different behaviors for the four protein scaffolds. The slowest overall release occurs from $^{C38/C96}R1_4$ and $^{C38/C81/C96}R1_4$, implying that Zn^{II} binding is relatively inert for these scaffolds. This is consistent with binding sites that are less exposed and exhibit roughly compensatory decrease in the association and dissociation rates. Strikingly, we also observe a context-dependence to the role of the C81-C81 disulfide bonds: addition of these crosslinks to $^{C96}R1_4$ appears to accelerate Zn^{II} release, while they slow Zn^{II} release when added to the $^{C38/C96}R1_4$ scaffold. These data illustrate the subtleties in overall behavior that can arise from a single perturbation in a protein scaffold

Determining tetramer stabilities by thermal denaturation. Disulfide bonds are commonly used in natural proteins to stabilize tertiary and quaternary structures.[24, 25]

Table 3. 6 Pseudo first-order Zn^{II} release rates from disulfide-linked $R1_4$ tetramers.
Data were determined by stopped-flow mixing of Zn-loaded protein with excess PAR at 22 °C. Rate constants are from fits to single- or double-exponential pseudo first-order rate laws. Binding stoichiometry (N_1 and N_2) is estimated from normalizing the pre-exponential factors in the rate equations. Values in parentheses denote errors in fitting of data averaged over eight replicate measurements.

$R1_4$ Variant	$k_{off,1}$ (s^{-1})	N_1	$k_{off,2}$ (s^{-1})	N_2
$^{C96}R1_4$	0.0247(2)	4	-	-
$^{C81/C96}R1_4$	0.34(2)	2.7(1)	0.044(4)	1.3(1)
$^{C38/C96}R1_4$	0.14(1)	1.01(1)	0.0051(1)	2.9(1)
$^{C38/C81/C96}R1_4$	0.0212(3)	1.25(3)	0.0031(2)	2.75(7)

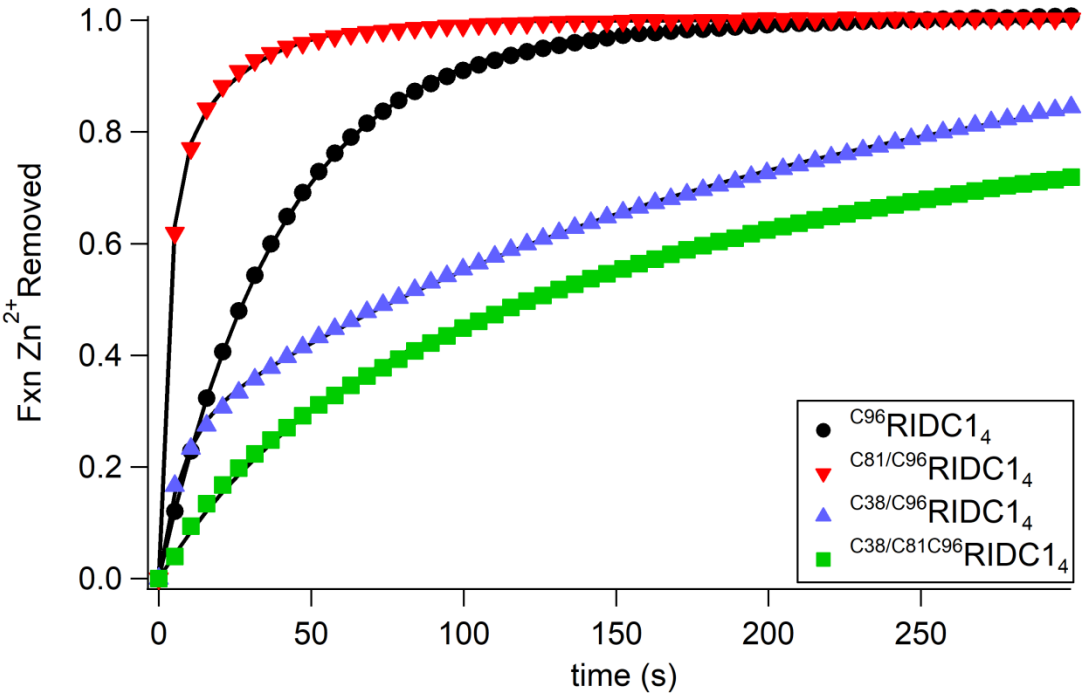

Figure 3.8 Zn^{II} binding curves of PAR mixed with Zn-loaded $R1_4$ variants.
Data collected at 5 s intervals are plotted for clarity (filled shapes). Data collected at 1 ms intervals were fit to pseudo first-order release rates (black lines). Data shown are averages of an eight-measurement series, and normalized to the expected Zn^{II} binding stoichiometry.

While disulfide bonds are generally thought to increase protein stability through limiting the conformational freedom of the unfolded state,[26-28] their net energetic influence is context-dependent, and they can sometimes even be destabilizing.[29] We investigated the effects of the different disulfide connectivities and supramolecular rigidities on protein stability in the isotopological disulfide-linked R1 tetramers. The tetrameric complexes do not fold/unfold reversibly, which precludes a quantitative thermodynamic analysis. Instead, we determined the resistance of these complexes to thermal denaturation as a measure of their stability, where the apparent melting temperature (T_m) served as an indicator of relative protein stability. Thermal unfolding was monitored by monitoring the ellipticity at 222 nm as a function of temperature (**Figure 3.9**).

$^{C96}R1_4$, with a single pair of disulfides in its *i2* interfaces, exhibited a T_m of ~56 °C. Addition of a second set of disulfides either across *i1* ($^{C38/C96}R1_4$) or *i3* ($^{C81/C96}R1_4$) increased the T_m to 64.6 and 66.8 °C, respectively (**Table 3.7** and **Figure 3.9**). However, the addition of the third set of disulfide bonds in the context of either scaffold, to give $^{C38/C81/C96}R1_4$, decreased the T_m to 62.2 °C (**Table 3.7** and **Figure 3.9**). The decrease in T_m was accompanied by an apparent decrease in the cooperativity of unfolding, indicated by a gentler slope for the melting transition. These observations point to the presence of the mutually destabilizing interactions in $^{C38/C81/C96}R1_4$ due to its extensive crosslinking. The series as a whole also points to non-additive and context-dependent effects exerted by the disulfide bonds on stabilization of the protein complex. Additionally, these data reveal the potentially undesirable effects of extensive crosslinking from generating an overly rigid protein scaffold.

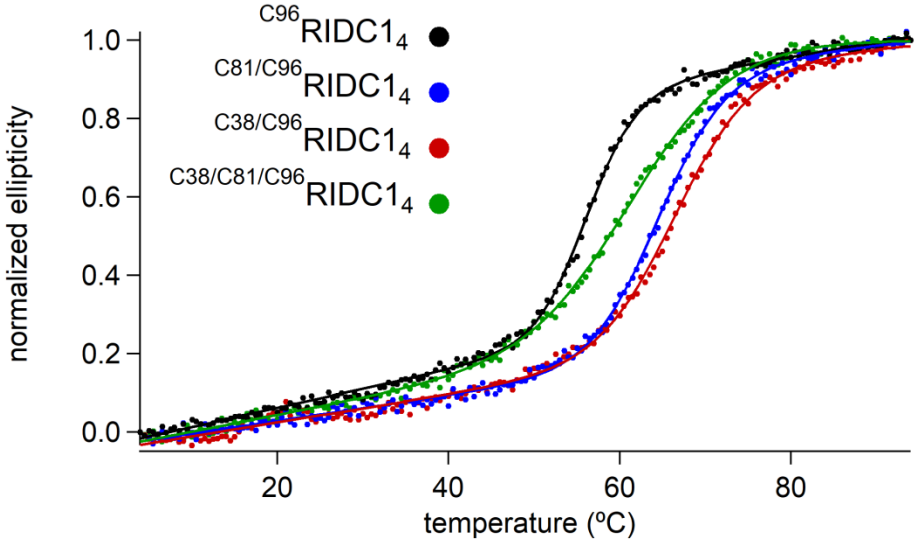

Figure 3.9 Thermal denaturation of disulfide-linked R1 tetramers.

Table 3.7 Apparent melting temperatures of disulfide-linked R1 tetramers.
Reported errors correspond to fitting errors of a single measurement.

Protein Variant	T_M (° C)
$^{C96}R1_4$	56.2 ± 0.1
$^{C81/C96}R1_4$	66.8 ± 0.3
$^{C38/C96}R1_4$	64.7 ± 0.2
$^{C38/C81/C96}R1_4$	62.2 ± 0.3

Assessing C38-bearing R1 variants for disulfide breakage. We next sought to determine if Zn^{II} removal and C38-C38 hydrolysis were allosterically coupled in $^{C38/C96}R1_4$ using ITC. We found that formation of a C38-C38 disulfide bond could be inferred from the Zn-binding thermogram of a C38-bearing variant relative to its C38-lacking counterpart, as the apparent enthalpy of binding (ΔH_{ITC}) was far more exothermic ($\Delta\Delta H_{+C38}$ = −49 ± 1 kJ/mol for $^{C38/C81/C96}R1_4$). Therefore, we carried out ITC measurements for Zn^{II} binding to $^{C38/C96}R1_4$, and fit the data using the dissociation constants determined by fluorescence competition titrations (**Table 3.4**) to estimate ΔH_{ITC} for Zn-binding to $^{C38/C96}R1_4$ (ΔH_{ITC} = +17 ± 7 kJ/mol) (**Table 3.8** and **Figure 3.10**). Comparing these data with the previously obtained Zn-binding enthalpy of $^{C96}R1_4$ reveals a significantly more exothermic heat of binding in $^{C38/C96}R1_4$ ($\Delta\Delta H_{+C38}$ = −45 ± 15 kJ/mol), consistent with formation of a disulfide bond upon Zn^{II} binding (**Table 3.8**). The extensive crosslinking of $^{C38/C81/C96}R1_4$, therefore, is not necessary for allosteric coupling of Zn^{II} binding to disulfide bond cleavage.

We also examined whether C38-C38 hydrolysis occurs upon Zn^{II} removal for $^{C38}R1_2$. Due to the minimal crosslinking of the scaffold, this question could be probed using non-reducing SDS-PAGE. Hydrolysis of the sole disulfide bond in a crosslinked $^{C38}R1_2$ dimer can be readily inferred from the formation of monomeric species. Treating $^{C38}R1_2$ with a stoichiometric amount of Zn^{II}, as expected, afforded a single dimeric species (**Figure 3.11**). We then exposed this sample to iodoacetamide (IAA), an alkylating agent for capable of capping both cysteine thiol and cysteine sulfenic acid [30]. In the presence of this capping agent, we added the chelator EDTA to remove the protein-bound Zn^{II}. The observed dimer persisted under these conditions, with no concomitant

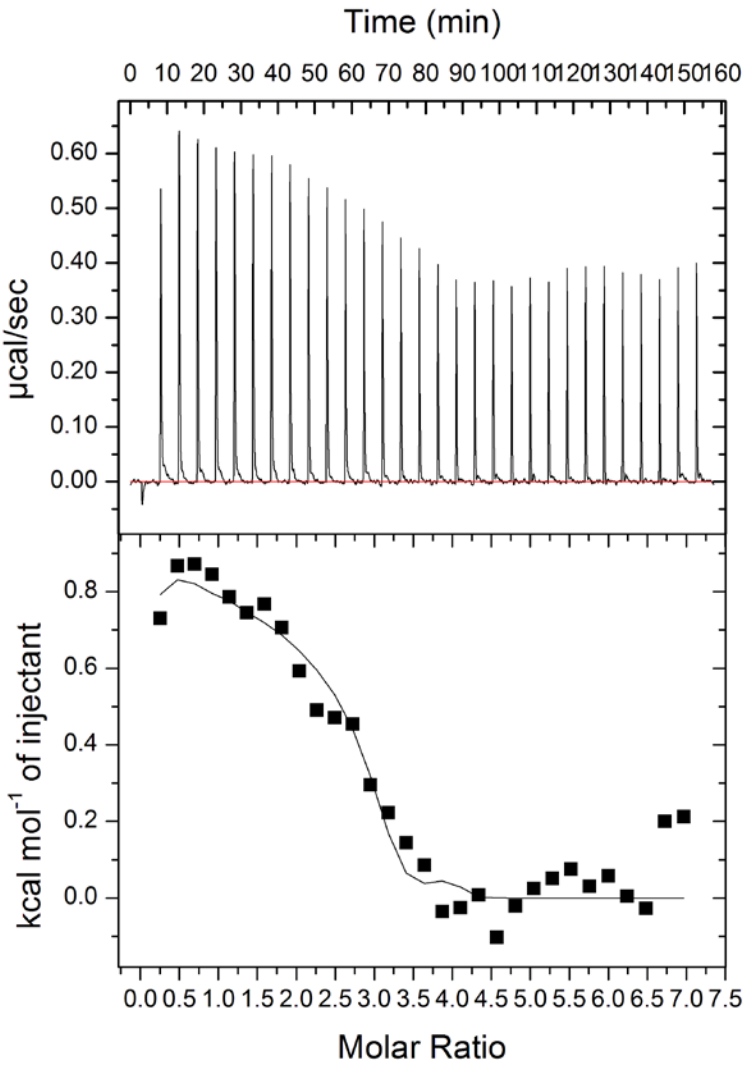

Figure 3.10 Representative thermogram of Zn^{II} binding to $^{C38/C96}R1_4$.
A solution of $ZnCl_2$ was titrated into $^{C38/C96}R1_4$. Upper panel depicts the baseline corrected heats of injection. Lower panel depicts the integrated heats of injection (black squares) and best-fit curve (black line) for the data. The data were fit usin the dissociation constants determined by fluorescence competition, which were held fixed, and fit to a four-site model for the enthalpies of binding, which were allowed to vary.

Table 3.8 Binding enthalpies of disulfide-linked R1 tetramers
Errors are derived from three or more replicate measurements. *Data taken from reference 22.

$R1_4$ Variant	**Total ΔH_{ITC}** (kJ/mol)	**$\Delta\Delta H_{+C38}$** (kJ/mol)
$^{C81/C96}R1_4$	59 ± 1*	-
$^{C38/C81/C96}R1_4$	10 ± 0.3*	−49 ± 1
$^{C96}R1_4$	62 ± 13*	-
$^{C38/C96}R1_4$	17 ± 7	−45 ± 15

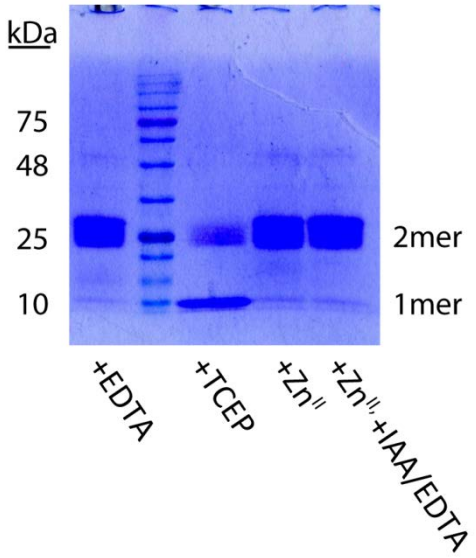

Figure 3.11 Assessment of disulfide breakage in $^{C38}R1_2$ by SDS-PAGE.
Isolated, disulfide-linked protein was treated with excess EDTA, affording a single dimeric species. Treatment with the reductant TCEP afforded predominantly monomeric species, consistent with reduction of a disulfide bond. Treating the oxidized, metal-free protein with Zn^{II}, or stepwise treatment with Zn^{II}, IAA, and EDTA, both also afforded a dimeric species on a non-reducing gel.

formation of a monomer (**Figure 3.11**), showing that Zn^{II} removal from $^{C38}R1_2$ does not result in C38-C38 hydrolysis. Allosteric coupling of Zn^{II} removal and C38-C38 hydrolysis cannot arise from intrinsic fragility due, or solely from local structural rearrangement within the *i1*-mediated dimer. Rather, the observed allosteric behavior emerges from the larger structural context afforded by $^{C38/C81/C96}R1_4$. We reasoned that the favorable relaxation that occurs upon Zn^{II} release from a single dimer is not sufficient to promote C38-C38 hydrolysis, but that coupling the relaxation of two dimers within a tetramer would provide an increased driving force for hydrolysis.

We sought to confirm the allosteric coupling of Zn^{II} removal and C38-C38 hydrolysis in $^{C38/C96}R1_4$ by x-ray crystallography. Surprisingly, we observed the *in crystallo* formation of a disulfide-linked octameric species consisting of four *i1*-mediated dimers connected by four C38-C38 and four C96-C96 disulfide bonds (**Figure 3.12** and **Table 3.9**). When metal-free $^{C38/C96}R1_4$ was allowed to stand, formation of an octameric species was never observed by SDS-PAGE or AUC, suggesting that this crystallographically-observed state is artefactual, a phenomenon that can complicate the oligomeric assignment of many proteins [31]. In obtaining crystals, we used the amphiphilic precipitant 2-methyl-2,4-pentanediol (MPD), which is known to lower the dielectric concentration of the crystallization solution [32]. The high concentration of MPD employed here (15% v/v) could weaken the *i1* interactions within $^{C38/C96}R1_4$ as it crystallizes. A $^{C38/C96}R1_4$ bearing a single broken C38-C38 disulfide would be capable of opening up significantly if the *i1* contacts were loosed, effectively exposing the C38 residues to make them available to form octamerizing crosslinks (**Figure 3.12**). Such a species need not be the prevailing state in solution for this to occur, as crystallization contains inherent

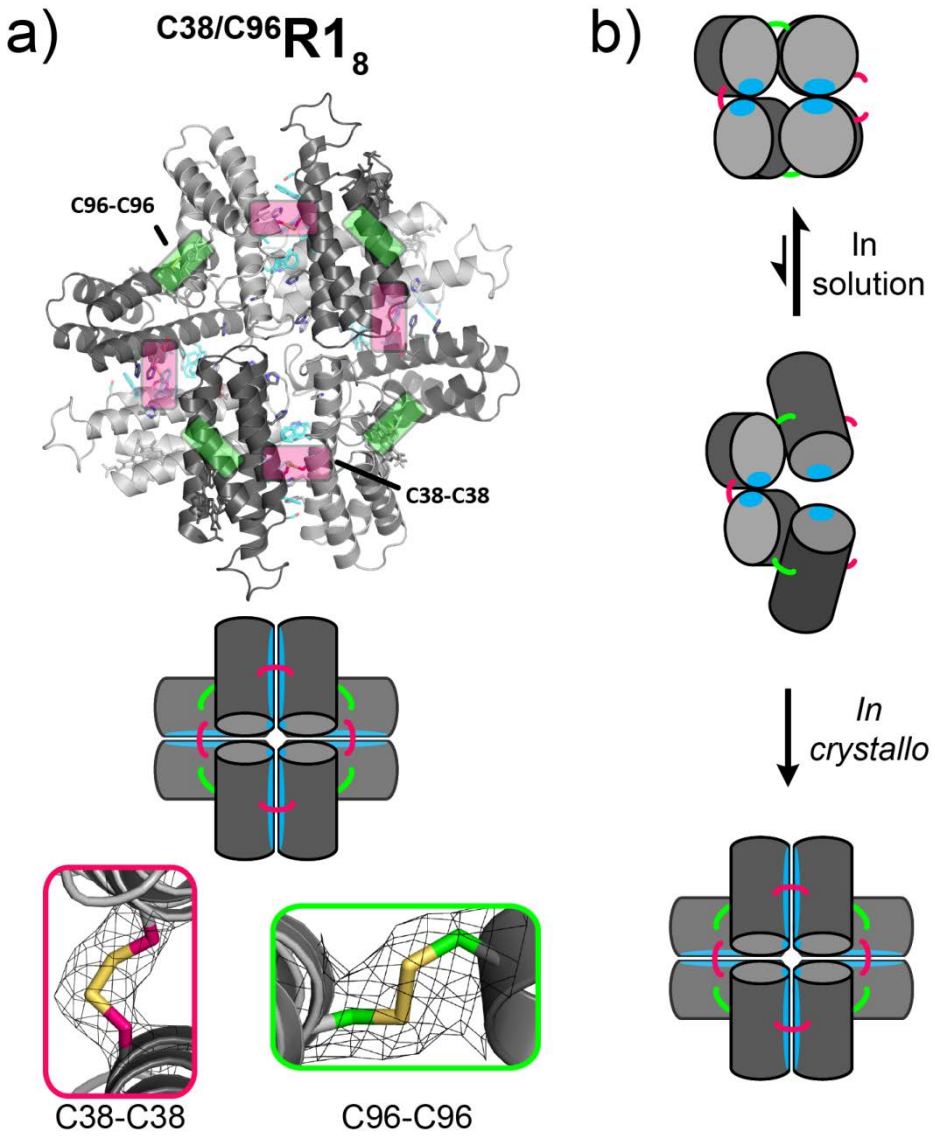

Figure 3.12 Crystallization of metal-free disulfide-linked $^{C38/C96}$R1.
(a) X-ray crystal structures of the *in crystallo* octamer of metal-free $^{C38/C81/C96}$R1 (PDB: 6DHZ). Insets show the interfacial disulfide bonds: C38-C38 (pink) and C96-C96 (green). $2F_o-F_c$ electron density maps (black mesh) are contoured at 1 σ. (b) Proposed scheme for the formation of the crystallographic octamer $^{C38/C96}$R1 from a tetrameric species that bearing a single broken disulfide bond. Due to the low extent of crosslinking, an intermediate butterfly tetramer could form by breaking and reforming the designed interfacial contacts (cyan) during crystallization.

Table 3.9 X-ray crystallography data collection and refinement statistics for $^{C38/C96}$R1 structures. Values in parentheses denote statistics for the highest resolution shell.

‡Rsym = $\Sigma\Sigma_j|I_j - \langle I \rangle|/\Sigma\Sigma_j|Ij|$.
§R = $\Sigma||F_{obs}| - |F_{calc}||/\Sigma |F_{obs}|$.
[II]Free R calculated against 5% of the reflections removed at random for both structures.
¶Root mean square deviations from bond and angle restraints.
*Values in parentheses correspond to the highest resolution shell.

	Zn-$^{C38/C96}$R1$_4$	$^{C38/C96}$R1$_8$
Data collection		
Space group	P 6_1	P 4_122
Cell dimensions		
a, *b*, *c* (Å)	52.5, 52.5, 255.7	78.6, 78.6, 168.3
α, β, γ (˚)	90, 90, 120	90, 90, 90
Resolution (Å)	2.22	2.80
R_{sym}‡* (%)	5.8 (19.6)	8.8 (46.9)
I/σ*	23.2 (9.2)	16.8 (3.5)
$CC_{1/2}$* (%)	99.9 (97.7)	99.9 (87.9)
Completeness* (%)	99.8 (99.9)	99.7 (98.1)
Redundancy*	4.3 (3.7)	9.0 (5.8)
Refinement		
Resolution (Å)	2.22	2.80
No. unique reflections	19755	13669
$R^{§}/R_{free}$ [II] (%)	22.5/26.4	20.9/28.0
No. atoms		
Protein	3271	3168
Ligand/ion	185	179
Water	41	32
***B*-factors**		
Protein	44.01	60.99
Ligand/ion	33.37	56.45
Water	29.26	39.85
R.m.s. deviations		
Bond lengths¶ (Å)	0.0085	0.0135
Bond angles¶ (˚)	1.385	1.590

energetic bias. Therefore, we take the formation of this unusual species as further, albeit indirect, evidence that hydrolysis of a C38-C38 disulfide bond occurs upon Zn^{II} removal from $^{C38/C96}R1_4$.

We also determined the structures of the Zn-bound form of $^{C96/C38}R1_4$ by x-ray crystallography (**Figure 3.13** and **Table 3.8**). The electron density maps showed Zn^{II} ions at the conserved H63/H73/D74/H77 metal-binding sites, and a 1:1 Zn:protomer stoichiometry. Additionally, the homotetrameric architecture resembling that of the parent Zn-$R1_4$ variant was preserved. These structural models allowed us to quantify the strain of the C38-C38 disulfide bonds using their dihedral angles,[33] which were known to be the most-strained disulfide pair in Zn-$^{C38/C81/C96}R1_4$ (**Table 2.3**). By this measure, the C38-C38 disulfide bonds of Zn-$^{C38/C96}R1_4$ were slightly more strained ($E_{strain\ avg.}$ = 27 kJ/mol) than those of Zn-$^{C38/C81/C96}R1_4$ ($E_{strain\ avg.}$ = 25 kJ/mol) (**Table 3.10**).[34] However, the octamer in the apo-$^{C38/C96}R1$ crystal structure contained C38-C38 disulfide bonds with even greater strain energy ($E_{strain\ avg.}$ = 30 kJ/mol). Clearly, disulfide strain alone is not a sufficient energetic descriptor of the structural changes that Zn^{II} removal and C38-C38 hydrolysis. Moreover, in comparing a Zn-bound and metal-free structure pair, we cannot decouple structural changes that follow Zn^{II} release from those that only occur following disulfide bond hydrolysis. We reasoned that decoupling these factors would help to implicate energetic drivers of disulfide hydrolysis in $^{C38/C81/C96}R1_4$, and reveal the degree to which structural relaxation, including relaxation of the disulfide bonds, occurs following metal removal or following bond breakage.

The functional prowess of proteins arises from their complex and dynamic structures. We have demonstrated that this holds true of the allosteric behavior that

Table 3.10 Dihedral strains of disulfide bonds in crystal structures of disulfide-linked R1 variants. Strain energies were calculated as reported in reference 33. Entries for individual disulfide bonds were averaged over all intact disulfide bonds within each crystallgraphic asymmetric unit. Entries for the ΔC38 sum were calculated from the total disulfide bond energies in the $^{C38/C81/C96}R1_4$ structure. For all other structures, the ΔC38 sum was taken as the total strain energy of all intact disulfide bonds, less the strain energy of a single C38 disulfide bond.

	Average strain energy (kJ/mol)			
Disulfide	**Zn-$^{C38/C96}R1_4$**	**$^{C38/C96}R1_8$**	**Zn-$^{C38/C81/C96}R1_4$**	**$^{C38/C81/C96}R1_4$**
C38-C38	30	27	25	19
C81-C81	-	-	23	13
C96-C96	13	12	17	6
ΔC38 sum	56	51	105	57

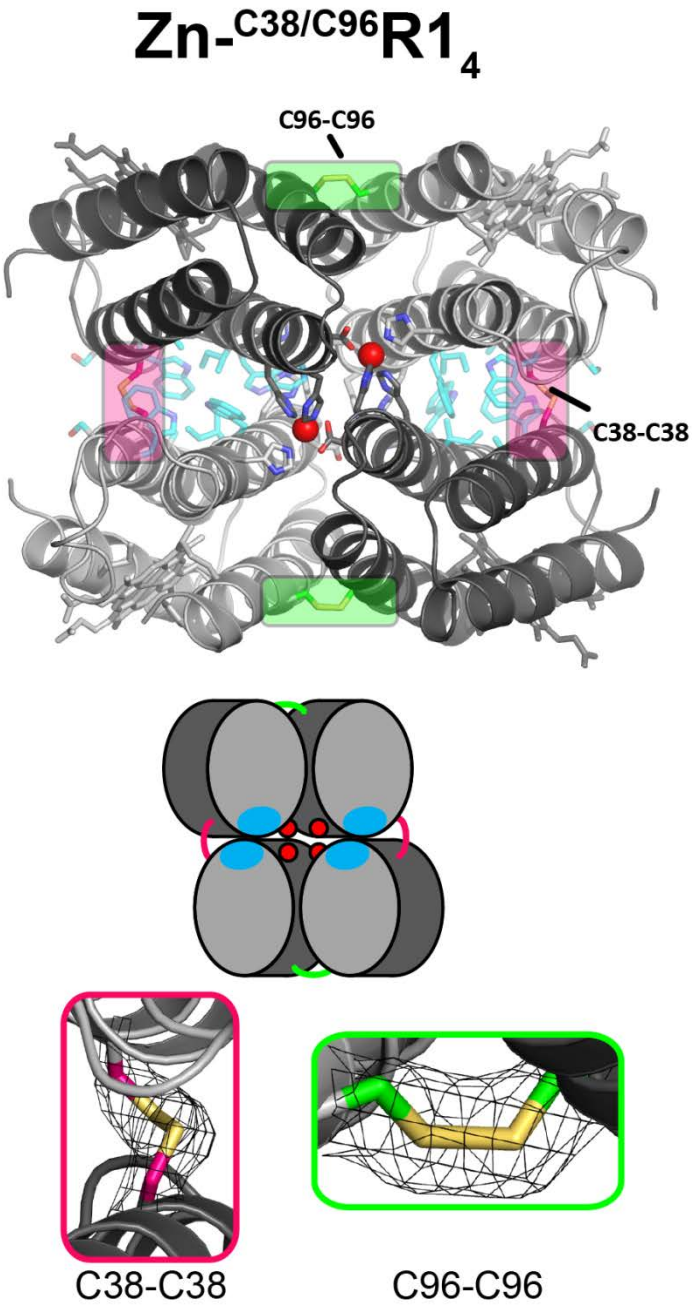

Figure 3.13 Crystal structure of Zn-$^{C38/C96}$R1$_4$.
X-ray crystal structures of Zn-bound $^{C38/C96}$R1$_4$ (PDB: 6DHY) with insets showing the C38-C38 (pink) and C96-C96 (green) disulfide bonds. $2F_o-F_c$ electron density maps (black mesh) are contoured at 1 σ.

occurs in the engineered protein $^{C38/C81/C96}R1_4$. By characterizing assemblies of the disulfide-deleted variants $^{C38}R1$ and $^{C38/C96}R1$, we have shown that the coupling of Zn^{II} removal to C38-C38 hydrolysis does not arise automatically from the Zn-induced structural rearrangement at a single C38-containing dimeric interface. Instead, these data imply that coupling of both interfaces within a crosslinked tetramer is necessary to develop a sufficient driving force for C38-C38 hydrolysis to occur. This family of variants also revealed that the extensive crosslinking arising from the three pairs of disulfide bonds in $^{C38/C81/C96}R1_4$ is not required for allostery. In fact, incorporation of all six disulfide bonds in $^{C38/C81/C96}R1_4$, rather than inclusion of the C38-C38 disulfide bond irrespective of structural context, results in the observed penalty to Zn^{II} binding affinity. Moreover, while each pair of disulfide bonds had a stabilizing effect on the crosslinked tetramers, the extensive crosslinking in $^{C38/C81/C96}R1_4$ actually led to a destabilization of the complex relative to $^{C38/C96}R1_4$ and $^{C81/C96}R1_4$. Collectively, these results point to the importance of structural plasticity as an element of design in engineered protein systems. We anticipate that, as protein engineers continue to make strides toward designing complex and functional protein assemblies, structural dynamics will emerge as a crucial aspect of achieving these goals.

3.3 Materials and Methods

Protein sample preparation and purification. Protein samples were prepared and purified essentially as described elsewhere.[14] The genes for all protein variants used in this work were housed in a pet20b vector (Novagen), and included an N-terminal leader sequence[18] for periplasmic localization. A codon encoding residue C38 was installed on

genes encoding R1, ^{C96}R1, and ^{C96}M1 by site-directed mutagenesis to generate genes encoding ^{C38}R1, $^{C38/C96}$R1, and $^{C38/C96}$M1, respectively. Mutagenic primers were obtained from Integrated DNA Technologies, and the pair 5'-cgctccacgcatcgcacgctgcggccgc-3' and 5'-ggcggccgcagcgtgcgatgcgtggagcg-3' was used to generate the R1-based genes, while the primer pair 5'-gcgcgccgcagcgtgcgatgcgtggagcg-3' and 5'-cgctccacgcatcgcacgctgcggcgc-gc-3' was used to generate the M1-based gene of interest. DNA amplification was carried out using Pfu Turbo DNA Polymerase (Agilent Technologies). The amplified DNA was transformed into XL-1 Blue competent *E. coli* cells (Agilent Technologies), and the cells were grown at 37 °C on lysogeny broth (LB) agar plates containing 60 μg/mL ampicillin. Plasmids were purified with the QIAprep Spin Miniprep Kit (QIAGEN), verified by sequencing (Retrogen), and then transformed into BL21(DE3) *E. coli* cells (New England Biolabs) housing the cytochrome c maturation (ccm) plasmid cassette.[35] The cells were grown at 37 °C on LB agar plates containing 34 μg/mL chloramphenicol and 60 μg/mL ampicillin. Single colonies were used to inoculate 5 mL of liquid LB medium containing chloramphenicol and ampicillin. Cultures were shaken at 37 °C until visibly turbid (typically about 8 h, or to an OD_{600}>0.6) at which time 50 μL of inoculum was transferred to 1 L of LB medium, typically to 15 cultures in parallel, and shaken for 16-20 h at 37 °C with protein expression occurring by auto-induction. Cells were harvested by centrifugation (5,000 × g, 4 °C, 10 min), resuspended in 100 mL of 5 mM sodium acetate (pH 5), frozen, thawed, and sonicated for 12 min in pulses of 30 s on and 60 s off in the presence of lysozyme (~100 mg). For purifications under reducing conditions, excess DTT (~300 mg) was added prior to sonicating.

The lysate was titrated with sodium hydroxide to pH >10 and acetic acid to pH 5, and cleared by centrifugation (10,000 rpm, 4 °C, 10 min). Purifications carried out under reducing conditions were done in the presence of 2 mM DTT, and DTT was excluded for the purification of oxidized protein. The red cleared lysate was decanted and diluted to 2 L in 5 mM sodium acetate (pH 5) containing 2 mM DTT, and was manually applied to a CM Sepharose column (GE Healthcare) equilibrated in 5 mM sodium acetate (pH 5), and the protein sample was eluted by a manually applied 0-1 M gradient of NaCl. The clear red eluate was concentrated using a Diaflow concentrator (Amicon) fitted with a 3-kDa cutoff membrane, and exchanged into a solution of 10 mM sodium phosphate (pH 8) and 2 mM DTT. The sample was loaded onto a DuoFlow fast protein liquid chromatography station fitted with a Macroprep High Q-cartridge column (BioRad) with 10 mM sodium phosphate (pH 8) running buffer containing 2 mM DTT, and eluted using a 0-0.5M NaCl gradient. Fractions of R1-based variants with Reinheitzahl values (A_{421}/A_{280}) above 3 were retained, and fractions of M1-based variants were assessed for purity by SDS-PAGE. Retained fractions were combined, concentrated, and mixed with an excess of EDTA and DTT (>10 molar equivalents). These stock solutions were then flash frozen in liquid nitrogen for storage at –80 °C.

Protein sample templating. Concentrated stock solutions of $^{C38/C96}$R1 (or $^{C38/C96}$M1) were thawed, reduced by the addition of excess DTT, transferred to glass vials sealed with rubber septa, cycled under vacuum and argon atmosphere, and transferred to a glove box (Coy Lab) under an anaerobic (<10 ppm O_2) atmosphere of argon with 10% hydrogen. Here, the samples were exchanged into an assembly buffer solution (50 mM Tris, pH 7, and 150 mM NaCl) using Econo-Pac 10DG pre-packed columns (BioRad).

Protein concentration was determined spectrophotometrically ($\varepsilon_{415,\ ox}$ = 148,000 M^{-1} cm^{-1}),[18] and samples were diluted to a final concentration of 50 μM in assembly buffer supplemented with 75 μM $ZnCl_2$. Samples were incubated overnight under ambient atmosphere at 37 °C with shaking. Self-assembly reactions were analyzed by SDS-PAGE in the absence of added reductant to keep disulfide bonds intact. The crude self-assembly reaction mixture was applied to a preparative-scale Superdex 75 gel filtration column (GE Healthcare) equilibrated in assembly buffer. Tetramer content of the eluted fractions was assessed by non-reducing SDS-PAGE. The purest fractions were combined, concentrated, and treated with >10-fold excess EDTA to remove bound metal, and applied to an Econo-Pac 10DG prepacked column equilibrated in 20 mM MOPS (pH 7) and 150 mM NaCl. The resulting stock solutions of Zn-free tetramer were flash frozen in liquid nitrogen, and stored at –80 °C.

Analytical ultracentrifugation. Solutions of 1.25 μM tetramer ($^{C38/C81/C96}R1_4$) in 20 mM MOPS (pH 7) and 150 mM NaCl were treated with either 5 μM $ZnCl_2$ or 1 mM EDTA to prepare metallated or apo samples. Sedimentation velocity measurements were made on a XL-1 Analytical Ultracentrifuge (Beckman-Coulter) equipped with an An-60 Ti rotor at 41,000 rpm for 400 scans at 25 °C, and monitored at 415 nm. The endpoint of sedimentation was determined using the match function in HeteroAnalysis (http://biotech.uconn.edu/auf/). Scans were processed in Sedfit[36] using buffer viscosity (0.01002 poise), density (1.007 g/mL) and partial specific volume (0.7316 ml/g) parameters calculated by SEDNTERP (http://sednterp.unh.edu/). The final c(S) distributions shown in **Figure 3.7** are reported at a confidence level of 0.95 and were plotted in Igor (WaveMetrics).

Competitive Zn^{II} binding titrations. Assay buffer (20 mM MOPS, pH 7 and 150 mM NaCl) was treated with Chelex 100 resin (BioRad) to remove trace transition metal impurities. The concentration of a light-protected stock solution of Fura-2 (Invitrogen) was determined spectrophotometrically (ε_{362} = 27,000 M^{-1} cm^{-1}).[37] Samples of $^{C38/C81/C96}R1_4$ (8.25 μM tetramer; 33 μM monomer) and Fura-2 (5 μM), or C38R12 (17.5 μM dimer; 35 μM monomer) and Fura-2 (7.5 μM) were prepared in assay buffer. The sample was titrated with a $ZnCl_2$ solution while thermostatted at 22 °C, and fluorescence measurements were made after 5-min equilibration periods. Fura-2 fluorescence emission at 510 nm was monitored to obtain an excitation scan over 250-450 nm on a Horiba Fluorolog 2 fluorimeter. Binding isotherms were generated from the changes in emission intensity plotted as a function of Zn^{II} concentration, and were fit using Dynafit[38] as previously reported.[19, 20, 34] Briefly, we fit the four Zn-binding equilibria using three different models: a single dissociation constant model (or 4 × 1) where all sites were treated as equivalent, a two dissociation constant model (or 2 + 2) with two inequivalent pairs of binding sites, or a four dissociation constant model (or 1+1+1+1) with each site treated as inequivalent. Additionally, the super-stoichiometric Zn^{II} binding observed for $^{C38}R1_2$ was modeled with an additional apparent dissociation constant (K*) assuming a binding site concentration equal to the concentration of protein monomer. We found a two dissociation constant model (2 + 2) adequately described the data for $^{C38}R1_2$, while $^{C38C96}R1_4$ was best described by a four dissociation constant model, which were plotted in Igor (WaveMetrics) to generate the curves shown in **Figure 3.7,** and the average free energy of binding was reported in **Table 3.4**.

Thermal denaturation of RIDC1$_4$ variants. Thermal denaturation experiments were carried out in 20 mM borate (pH 7) and 150 mM NaCl in the presence of 1 mM EDTA. Cell temperature was raised from 4-94 °C in 0.5 °C increments. Melts were carried out using 1.25 μM tetramer in a 0.2 cm quartz cuvette with 30 s equilibrations at each temperature. The ellipticity was measured at 222 nm using an Aviv 215 CD spectrometer. The ellipticity measurements were fit to a two-state model to calculate the apparent T_M, assuming no change in heat capacity ($\Delta C_p = 0$), using linear terms to correct for changes in signal pre- and post-transition, as previously reported.[39] The data were fit using Kaleidagraph (Synergy Software) and plotted in Igor (WaveMetrics) to generate the curves shown in **Figure 3.9**, and the apparent melting temperatures are reported in **Table 3.6**.

Stopped-flow mixing of PAR and Zn$_4$:R1$_4$ variants. Stopped-flow mixing experiemtns were carried out using an SX.17MV (Applied Photophysics) with a 0.2 cm path length and the bandwith set at 1 nm. All samples were prepared in Chelex 100-treated buffered solutions consisting of 20 mM MOPS, pH 7, and 150 mM NaCl. Light-protected samples of 4 mM PAR were mixed with 25 μM of Zn-R1$_4$ variant loaded with 1 equivalent of Zn^{II}. Mixing was done asymmetrically in a 10:1 volume ratio of PAR:protein solution, giving a 1600:1 molar ratio of PAR to tetramer in the final mixture. Changes to the absorbance at 500 nm were monitored for 300 s at 1 ms intervals (30,000 points total) for eight injections. Data were fit to rate equations of the form:

$$A_t = A_{final} - \sum_{1}^{n} A_n e^{-t*k_n}$$

where n = 1, 2 or 4 for single-, double-, or quadruple-exponential rate equations, respectively, using Kaleidagraph (Synergy Software). Assuming pseudo first-order conditions, k_n is off rate for one or more Zn^{II} from the $RIDC1_4$ variants. The goodness of fit statistics and fit residuals showed that the dissociation of Zn^{II} could be well-described by a single-exponential rate equation (n=1) for $^{C96}RIDC1_4$, and a double-exponential rate equation (n=2) for $^{C81/C96}RIDC1_4$, $^{C38/C81}RIDC1_4$, and $^{C38/C81/C96}RIDC1_4$. Approximate binding stoichiometry was verified by comparison with a $Zn:PAR_2$ standard curve, where samples of 0-200 μM Zn^{II} were mixed in a 1:10 ratio (v/v) with 4 mM PAR and monitored for 10 s over five injections. Binding stoichiometry for the individual terms were estimated by normalizing the pre-exponential constants to a total occupancy of four Zn^{II} ions per tetramer (*i.e.* by setting $A_t = 4$).

Chapter 3 was reproduced in part, with permission, from a manuscript currently being prepared for submission for publication: Churchfield, L.A.; Williamson, L.A.; Alberstein, R.G; Tezcan, F.A. Investigating the structural and mechanistic basis of allostery in a designed protein scaffold.

The dissertation author is primary author on all reprinted materials.

References

(1) Goodsell, D. S.; Olson, A. J. *Annual Review of Biophysics and Biomolecular Structure* **2000**, *29*, 105–153.

(2) Jones, S.; Thornton, J. M. *Proceedings of the National Academy of Sciences U.S.A.* **1996**, *93*, 13.

(3) Ali, M. H.; Imperiali, B. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 5013.

(4) Luo, Q.; Hou, C.; Bai, Y.; Wang, R.; Liu, J. *Chemical Reviews* **2016**, *116*, 13571.

(5) Brodin, J. D.; Auyeung, E.; Mirkin, C. A. *Proceedings of the National Academy of Sciences U.S.A.* **2015**, *112*, 4564.

(6) Lai, Y.-T.; Cascio, D.; Yeates, T. O. *Science* **2012**, *336*, 1129.

(7) Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U.; Ellis, D.; Bale, J. B.; Ke, S.; Lenz, G. H.; Yehdego, A.; Ravichandran, R.; Pun, S. H.; King, N. P.; Baker, D. *Nature* **2017**, *552*, 415.

(8) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R.; Yi, S.; Davis, T. N.; Gonen, T.; King, N. P.; Baker, D. *Nature* **2016**, *535*, 136.

(9) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A. *Nature* **2016**, *533*, 369–373.

(10) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014**, *508*, 331.

(11) Kohen, A. *Accounts of Chemical Research* **2015**, *48*, 466.

(12) Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A. *Journal of the American Chemical Society* **2007**, *129*, 13374–13375.

(13) Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A. *Proceedings of the National Academy of Sciences U.S.A* **2010**, *107*, 1827–1832.

(14) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. *Methods in Enzymology* **2016**, *580*, 223.

(15) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. *Accounts of Chemical Research* **2010**, *43*, 661–672.

(16) Brodin, J. D.; Ambroggio, X. I.; Tang, C. Y.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. *Nature Chemistry* **2012**, *4*, 375–382.

(17) Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A. *Journal of the American Chemical Society* **2015**, *137*, 10468–10471.

(18) Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R. *Biochemistry* **2006**, *45*, 10504.

(19) Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A. *Journal of the American Chemical Society* **2010**, *132*, 8610–8617.

(20) Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A. *Journal of the American Chemical Society* **2013**, *135*, 12013.

(21) Song, W. J.; Tezcan, F. A. *Science* **2014**, *346*, 1525–1528.

(22) Petersen, M. T. N.; Jonson, P. H.; Petersen, S. B. *Protein Engineering* **1999**, *12*, 535.

(23) McCall, K. A.; Fierke, C. A. *Analytical Biochemistry* **2000**, *284*, 307.

(24) Fass, D. *Annual Review of Biophysics* **2012**, *41*, 63.

(25) Mamathambika, B. S.; Bardwell, J. C. *Annual Review of Cell and Developmental Biology* **2008**, *24*, 211.

(26) Dombkowski, A. A.; Sultana, K. Z.; Craig, D. B. *FEBS Letters* **2014**, *588*, 206.

(27) Tidor, B.; Karplus, M. *Proteins: Structure, Function, and Bioinformatics* **1993**, *15*, 71.

(28) Dani, V. S.; Ramakrishnan, C.; Varadarajan, R. *Protein Engineering* **2003**, *16*, 187.

(29) Matsumura, M.; Becktel, W. J.; Levitt, M.; Matthews, B. W. *Proceedings of the National Academy of Sciences U.S.A.* **1989**, *86*, 6562.

(30) Reisz, J. A.; Bechtold, E.; King, S. B.; Poole, L. B.; Furdui, C. M. *FEBS Journal* **2013**, *280*, 6150.

(31) Krissinel, E. *Acta Crystallographica Section D: Biological Crystallography* **2011**, *67*, 376.

(32) Anand, K.; Pal, D.; Hilgenfeld, R. *Acta Crystallographica Section D: Biological Crystallography* **2002**, *58*, 1722.

(33) Katz, B. A.; Kossiakoff, A. *Journal of BIological Chemistry* **1986**, *261*, 15480.

(34) Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A. *Journal of the American Chemical Society* **2016**, *138*, 13163–13166.

(35) Arslan, E.; Schulz, H.; Zufferey, R.; Kunzler, P.; Thony-Meyer, L. *Biochemical and Biophysical Research Communications* **1998**, *251*, 744.

(36) Schuck, P. *Biophysical Journal* **2000**, *78*, 1606.

(37) Grynkiewicz, G.; Poenie, M.; Tsien, R. Y. *Journal of Biological Chemistry* **1985**, *260*, 3440.

(38) Kuzmic, P. *Analytical Biochemistry* **1996**, *237*, 260.

(39) Greenfield, N. J. *Nature Protocols* **2006**, *1*, 2527.

Chapter 4: Molecular dynamics simulations of $^{C38/C81/C96}R1_4$.

4.1 Introduction

It is a long-held paradigm in biological systems that structure dictates function. Increasingly, for many protein systems, it has been found that achieving a competent three-dimensional structure is a necessary requirement for protein function, it is not sufficient in explaining the functional breadth on display in Nature's proteome. Rather, the underlying protein structure dictates the dynamic behavior of a protein or protein complex, and is a critical aspect of the functional capabilities of many natural protein systems.[1-5] This chapter is an exploration of molecular dynamics (MD) simulations as a tool for studying these principles at work in $^{C38/C81/C96}R1_4$. Using MD simulations, one

can obtain atomic-level structural information from an ensemble of protein structures to determine how the structure varies over time. Additionally, with appropriate ensembles of structures, one can obtain thermodynamic information about a structural change of interest, an approach that has been successfully applied to investigations of other allosteric protein systems.[6]

This chapter is intended to serve as a reference for adapting existing computational tools to cyt cb_{562} systems in general, as well as to present specific insights into the allosteric behavior of the $^{C38/C81/C96}R1_4$ system. Additionally, we note that the MD simulations described here complement and corroborate the biochemical and biophysical analysis described in **Chapters 2-3**. Specifically, we suggest that disulfide bond relaxation is incidental to the structural rearrangement observed in $^{C38/C81/C96}R1_4$ opening, as allosteric coupling exists in the less-extensively crosslinked scaffold $^{C38/C96}R1_4$.

The available crystal structures of $^{C38/C81/C96}R1_4$ in the Zn-bound and Zn-free forms reperesent the two end states of the allosteric coupling between Zn^{II} removal and C38-C38 bond breakage. However, these structures provide little direct information about the energetic basis of this coupling, nor do they explain why a single C38-C38 bond is selectively hydrolyzed/formed among several others within the scaffold. This embodies a major obstacle to the study of any allosteric regulation, as the means by which information is communicated structurally typically involves complex (and transient) sequences of localized conformational changes which are diffuclut to infer purely from well-defined end states.[7] Zn-$^{C38/C81/C96}R1_4$ presents the added challenge of decoupling the structural changes induced by Zn^{II} removal from those caused by disulfide bond

breakage. Therefore, we set out to examine the $^{C38/C81/C96}R1_4$ complex through molecular dynamics (MD) simulations, which allowed us to evaluate the thermodynamic parameters for ensembles of structural states what would not be accessible by experiments, as well as to examine some of the intermediate states that attend C38-C38 bond breakage. Specifically, we examined a metal-free intermediate state of $^{C38/C81/C96}R1_4$ bearing six disulfide bonds, as well as the structural ensembles that arise from hypothetical bond breakage events (C81-C81 or C96-C96 hydrolysis) as well as the experimentally observed bond breakage event (C38-C38 hydrolysis). Additionally, efforts to characterize Zn-bound states of $^{C38/C81/C96}R1_4$ and the intermediates of tetramer opening, as well as the challenges facing these modeling efforts, are discussed.

4.2 Development of custom simulation parameters

Parameterization of cysteine sulfenic acid. In order to model the broken disulfide bond of metal-free $^{C38/81/C96}R1_4$, we had to develop and implement a custom set of forcefield parameters compatible with the CHARMM27 forcefield.[8, 9] This was accomplished using the Force Field Toolkit (ffTK) plugin[10] of VMD. Parameters in the CHARMM forcefield are implemented with transferability, rather than system-specific accuracy, underlying the process. For example, the backbone amino acids in the CHARMM27 forcefield use identical parameters, and distinct parameter sets are only implemented for the sidechain of each residue. With this in mind, we selected as a parameterization model a small molecule model that closely mimics the sidechain of cysteine sulfenic acid (CSO): ethane-SO-thioperoxol (**Figure 4.1**). The deprotonated form of the molecule was used to obtain parameters for the sidechain of cysteine sulfon-

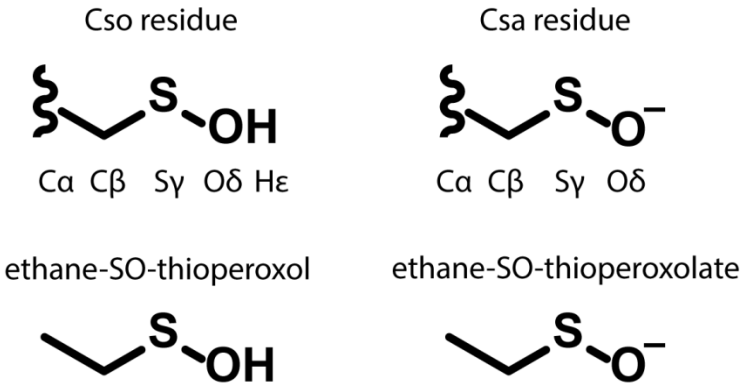

Figure 4.1 Chemical structures of sulfenic acid sidechains and paramterization models.
Cysteine sulfenic acid (Cso) and cysteine sulfonate (Csa) sidechains are shown truncated at their C-α atoms, and with sidechain atoms labeled. Parameterization was carried out using protonated and deprotonated ethane-SO-thioperoxol models, where the C-α terminates in a methyl group.

sulfonate (CSA) (**Figure 4.1**).

Initial structures of ethane-SO-thioperoxol and ethane-SO-thioperoxolate were manually built in the Molefacture plugin of VMD, which were used as the starting models for parametrization. Analogous atom types and Lennard-Jones parameters for evaluating van der Waals interactions for the β-carbon, γ-sulfur, δ-oxygen, and ε-hydrogen were assigned by analogy from existing atom types for cysteine (for C and S) and serine (for O and H) residues in the CHARMM27 forcefield. Geometry optimization was carried out in Gaussian [11] at the MP2/6-31G* level of theory. In both models, the initial arbitrary bond lengths (~1.5 Å) optimized to new lengths that appeared reasonable for the atom types involved, and the groups around the C-C bond adopted staggered configurations (**Figure 4.2**). These changes suggested that the structures had been successfully optimized, and the resulting models were used for subsequent parametrization steps.

We next carried out partial charge fitting for the new atom types in the Cso and Csa model structures. In accordance with the CHARMM convention, partial charges of +0.09 were uniformly applied to the aliphatic hydrogen atoms in the structure. This was accomplished by placing water molecules next to the parameterization model, and subjecting them to two-dimensional (distance and angle) positional refinements at the HF/6-31G* level of theory in Gaussian. This protocol omitted the sp^3 hybridized C-β atom, and placed a single water flanking S-γ atom in both residues, as well as the H-ε atom of Cso (**Figure 4.3**). A single water was placed flanking the O-γ atom of Cso, while three water molecules were

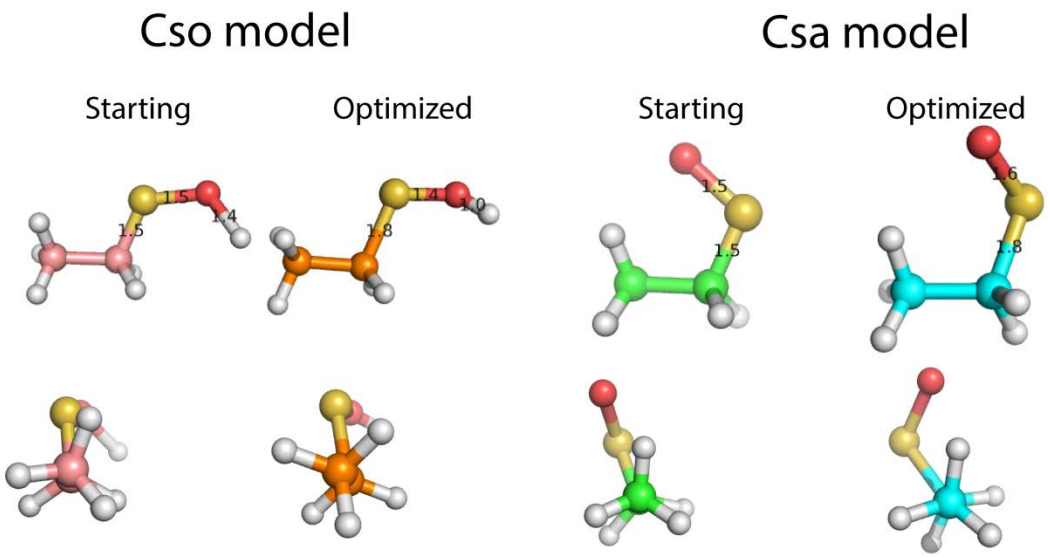

Figure 4.2 Geometry optimization of sulfenic acid parameterization models.
Starting models for Cso and Csa parameterization are shown in salmon and green, respectively. The optimized Cso and Csa models, shown in orange and blue, exhibit altered bond lengths and staggered conformations following the optimization procedure. Distances of C-S, S-O, and O-H bonds before and after optimization are reported in Ångstroms.

placed around the low-valent oxygen of Csa (**Figure 4.3**). The positions of these water molecules were used as target data for fitting CHARMM type partial charges at the molecular mechanics level of theory, which are shown alongside the previously reported parameters for Cys and Ser (**Table 4.1**).

The geometry-optimized structures were next used for parameterization of the bonds, angles, and dihedrals in the Cso and Csa models. We used these structures to compute the Hessian matrix of the potential energy surface for both structures in Gaussian at the MP2/6-31G* level of theory. Molecular mechanics parameter fitting was then carried out in the ffTK plugin to achieve bond (K_b and b_0) and angle parameters (K_θ and θ), which are reported in **Table 4.2** and **Table 4.3**. The dihedral angle parameters were obtained by carrying out torsion angle scans about the C-C-S-O dihedral in both models, and about the C-S-O-H dihedral in the Cso model. Fits of the molecular mechanics parameters to the quantum-level data are reported in **Figure 4.4.** Artefactual parameters that included the methyl hydrogens were excluded from the final dihedral parameter set (K_ψ, n, δ) reported in **Table 4.4**.

Parameterization of Zn^{II} atoms. The homotetramer $^{C38/C81/C96}R1_4$ has previously been crystallized with four Zn^{II} ions in its core metal binding sites, coordinated by residues His63, His73, Asp74 and His77. In order to probe the dynamics of the Zn-bound tetramer, we sought out Zn^{II} parameters that would represent the crystallographically observed coordination environment. However, we found that the crystallographic coordination of the Zn^{II} ions was poorly maintained in short (<10 ns) simulations in explicit solvent, during which the coordination geometry became distorted and ligands

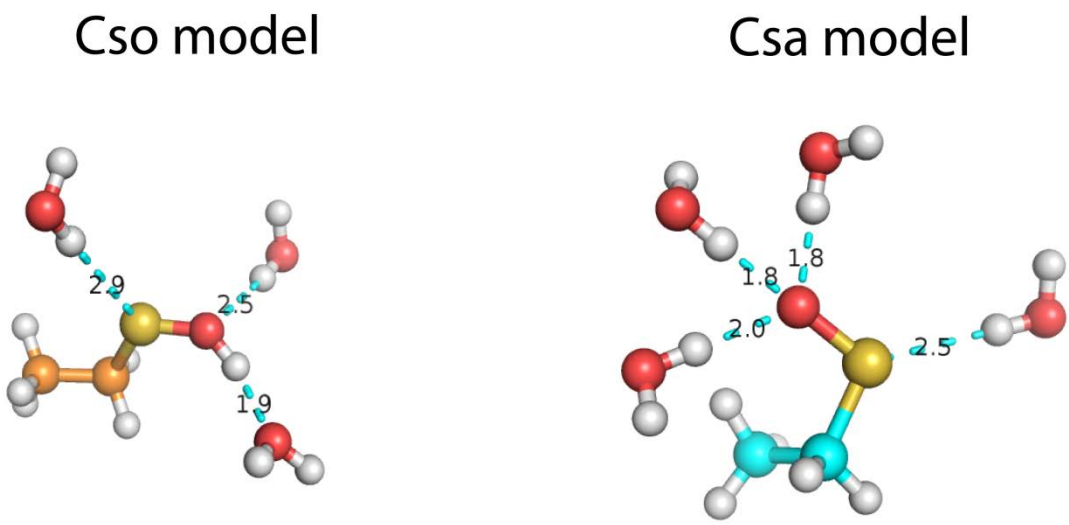

Figure 4.3 Water placement or partial charge parameterization of Cso and Csa.
Parameterization models for Cso and Csa are depicted with water molecules that were palced for partial charge parameterization. The configurations depicted are those following a 2-D optimization protocol, with the final distance drawn (cyan dashed line) and labeled in Å.

Table 4.1 CHARMM-type partial charges of sulfenic acid and chemically related residues.
Water placements depicted in **Figure 4.3** were used to establish CHARMM type partial charges for the sidechain atoms of Csa and Cso. Also shown are the partial charges used in the CHARMM27 forcefield for the chemically related Cys and Ser residues. Omitted partial charges denote atoms that are absent from the indicated residue.

Atom	Csa	Cso	Cys	Ser
C-β	-0.065	-0.04	-0.11	0.05
S	-0.195	0.01	-0.23	-
O	-0.92	-0.57	-	-0.66
H	-	0.42	0.16	0.43

Table 4. 2 Bond paramters for the sidechain atoms of the Cso and Csa residues.
The equilibrium bond lenghts (b_0) and force constants (k_b) of bonds connecting the sidechain atoms of Cso and Csa. Existing paramters were used for aliphatic C-H bonds.

	Atoms	k_b (kcal/mol/Å^2)	b_0 (Å)
Cso	C−S	214	1.808
	S−O	269	1.702
	O−H	489	0.972
Csa	C−S	200	1.839
	S−O	320	1.614

Table 4. 3 Angle parameters for the sidechain atoms of the Cso and Csa residues.
The equilibrium bond angles (θ) and force constants ($k_θ$) of angles described by the sidechain atoms of Cso and Csa.

	Atoms	$k_θ$ (kcal/mol/rad²)	$θ_0$ (°)
Cso	C-C-S	63	106
	C-S-O	60	102
	H-C-S	38	109.5
	S-O-H	55	106
Csa	C-C-S	68	109.5
	C-S-O	97	105
	H-C-S	44	111

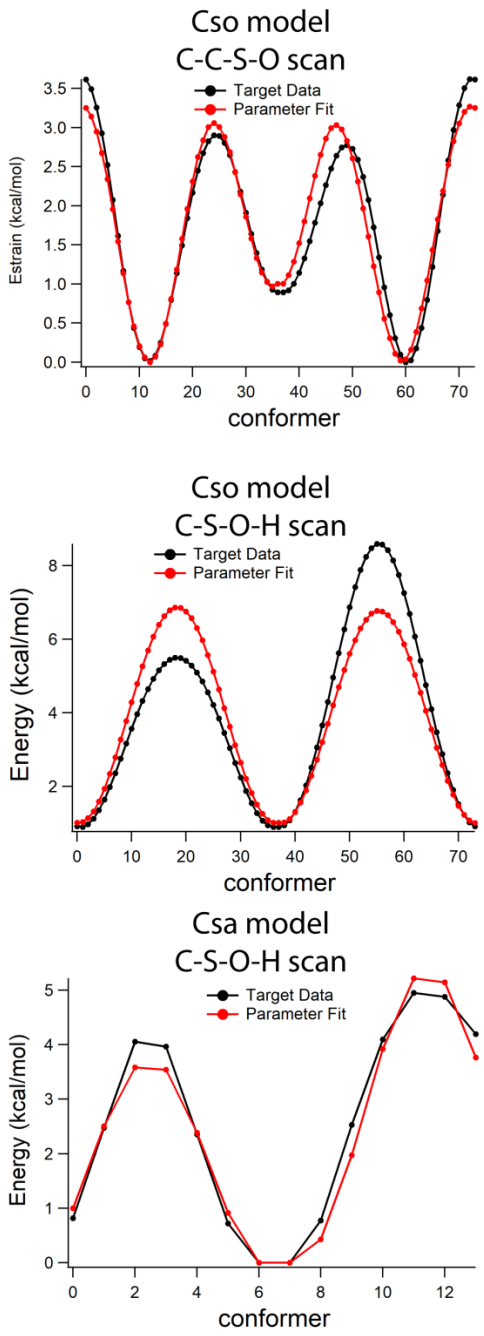

Figure 4.4 Torsion angle scan energy calculations and fits for Cso and Csa sidechains.
Cso and Csa structural models were scanned about their sidechain atom torsion angles, and the calculated potential energy of each state is plotted above (black curves). The dihedral parameters reported in **Table 4.4** are the result of fitting the data to molecular mechanics parameters.

Table 4.4 Dihedral parameters for the sidechain atoms of the Cso and Csa residues.
The phase shift (δ) periodicity (n) and force constant (k_ψ) parameters for the dihedral angles described by the sidechain atoms of Cso and Csa.

	atoms	Kψ (kcal/mol)	n	δ (deg)
Cso	H-C-S-O	0.567	1	180
	C-S-O-H	2.924	2	0
	C-C-S-O	1.193	3	0
	"	0.374	2	0
	"	1.201	1	180
Csa	C-C-S-O	2.314	1	180
	"	0.695	2	180
	"	2.301	3	0

frequently became displaced when the Zn^{II} coordinated additional water ligands. The forcefield parameters for Zn^{II} in the CHARMM27 forcefield have been successfully implemented in the past to simulate such proteins as carbonic anhydrase,[12] though in these cases the Zn^{II} centers were buried in the protein environment, and therefore better shielded from solvation.

The simple Lennard-Jones description of transition metals employed in the CHARMM forcefield are known to have limited transferability.[13] One means of circumventing this is the cationic dummy atom (CADA) description of a transition metal atom. The approach uses a tetrahedral array of low mass (3 Da) dummy atoms, each bearing a +0.5 partial charge, and surrounding an uncharged central atom (53.38 Da) (**Figure 4.5)**,[14] to better capture the tetrahedral coordination preferences of protein-associated Zn^{II}.[15] When we adapted these parameters for use in the CHARMM forcefield (see Appendix), however, we found that the Zn^{II} ions rapidly dissociated from the metal binding site in favor of coordinating water.

Merz and co-workers have previously noted that the Lennard-Jones description of transition metal ions could be readily adapted to capture different physical observables of interest, as well as different simulation protocols.[13] They developed parameters optimized with respect to the hydration free energies (HFE), and interatomic distance (IOD) of coordinated water molecules for Zn^{II} and other transition metals (**Table 4.5**). They also reported a so-called compromise parameter set, in which optimization was carried out with respect to both experimental parameters (**Table 4.5**). We found the HFE-optimized

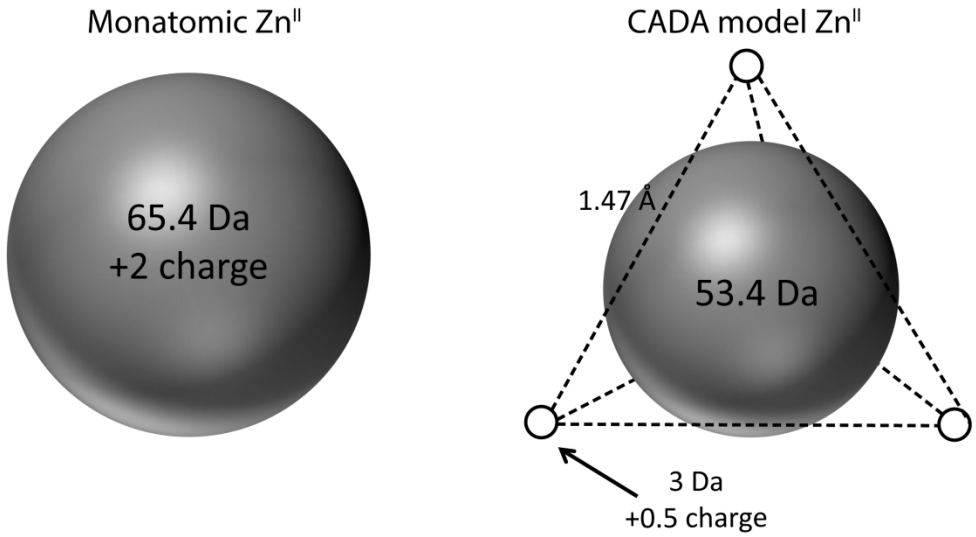

Figure 4.5 Schematic representation of the cationic dummy atom (CADA) Zn^{II} model.
The conventional treatment of Zn^{II} as a monatomic species employs a single 65.4 Da cation bearing a +2 net charge. The CADA description of Zn^{II} involves an electrostatically neutral atom with a mass of 53.4 Da which is flanked by four cationic atoms, each with a mass of 3 Da and bearing a +0.5 net charge. These so-called dummy atoms are arranged tetrahedrally about the central atom. The bonds linking the dummy atoms are 1.47 Å in length, while bonds between the dummy atoms and central atom are 0.9 Å in length.

parameters to be the most effective at maintaining the ligand set of interest when used to simulate Zn-$^{C38/C81/C96}$R1$_4$, and the least prone to coordinate additional water ligands. This treatment led to interatomic distances between the Zn^{II} ion and ligands that were shorter than reasonable, typically under 1.9 Å for His ligation and under 1.6 Å for Asp ligation (**Table 4.6**). Additionally, during a 10 ns simulation, we observed transient dissociation events for two of the 16 metal ligands (**Table 4.6)**. This suggested to us that unconstrained simulations of the Zn^{II} binding sites were unsuitable for enforcing the crystallographically observed metal binding sites.

We implemented a set of harmonic constraints on the distances between the Zn^{II} ions and the associated ligands to enforce the experimentally observed coordination environment. To do so, we determined the average Zn-ligand distances and degree of fluctuation during the 10 ns simulation, ignoring ligands for which significant dissociation events were observed (namely, His63 of chain A and His73 of chain C). We determined the force constant of a harmonic bond that would reproduce the observed level of fluctuation. By the equipartition theorem, the average potential energy of a harmonic oscillator at thermal equilibrium is:

$$\langle U \rangle = \frac{1}{2} k_B T$$

The potential energy of a harmonic oscillator is defined a harmonic oscillator is defined as follows:

$$U = \frac{1}{2} k_{HC} (d - d_0)^2$$

Table 4.5 Lennard-Jones parameter sets for Zn^{II} ions.

Parameter Set	Reference	epsilon (kcal/mol)	$R_{min}/2$ (Å)
Karplus	5	-0.25	1.09
HFE-optimized	6	-0.00071558	1.175
IOD-optimized	6	-0.014917	1.395
Compromise	6	-0.00330286	1.271

Table 4.6 Averaged Zn-ligand distances from an unconstrained Zn-$^{C38/C81/C96}R1_4$ simulation.
Zn-ligand distances above 2 Å are emphasized in grey, and denoted with a *. Manual inspection of the simulation trajectory confirmed that these corresponded to transient dissociation events. His63 of chain A and His73 of chain C were excluded from calculating the harmonic constraint parameters reported in **Table 4.7**.

His63-Zn Distances (Å)

window	A*	B	C	D
0 to 1 ns	1.91	1.86	1.88	1.97
1 to 2 ns	2.02*	1.81	1.79	1.78
2 to 3 ns	1.90	1.78	1.77	1.77
3 to 4 ns	1.86	1.80	1.78	1.78
4 to 5 ns	1.86	1.78	1.78	1.78
5 to 6 ns	1.85	1.77	1.77	1.83
6 to 7 ns	1.85	1.78	1.80	1.81
7 to 8 ns	1.91	1.78	1.76	1.78
8 to 9 ns	1.87	1.78	1.77	1.78
9 to 10 ns	2.06*	1.78	1.77	1.79

Asp74-Zn Distances (Å)

window	A	B	C	D
0 to 1 ns	1.58	1.59	1.56	1.58
1 to 2 ns	1.56	1.61	1.55	1.56
2 to 3 ns	1.55	1.60	1.56	1.56
3 to 4 ns	1.56	1.62	1.55	1.55
4 to 5 ns	1.55	1.62	1.56	1.56
5 to 6 ns	1.55	1.60	1.56	1.55
6 to 7 ns	1.55	1.61	1.56	1.56
7 to 8 ns	1.56	1.63	1.55	1.55
8 to 9 ns	1.56	1.59	1.56	1.55
9 to 10 ns	1.55	1.60	1.56	1.55

His73-Zn Distances (Å)

window	A	B	C*	D
0 to 1 ns	1.82	1.80	1.91	1.84
1 to 2 ns	1.81	1.78	1.86	1.79
2 to 3 ns	1.79	1.78	1.90	1.77
3 to 4 ns	1.79	1.78	2.12*	1.79
4 to 5 ns	1.79	1.78	2.19*	1.77
5 to 6 ns	1.79	1.78	2.25*	1.77
6 to 7 ns	1.79	1.81	2.07*	1.77
7 to 8 ns	1.78	1.79	1.96	1.76
8 to 9 ns	1.79	1.79	1.91	1.76
9 to 10 ns	1.79	1.78	1.81	1.76

His77-Zn Distances (Å)

Time	A	B	C	D
0 to 1 ns	1.78	1.79	1.82	1.83
1 to 2 ns	1.77	1.76	1.83	1.79
2 to 3 ns	1.77	1.76	1.83	1.77
3 to 4 ns	1.76	1.76	1.94	1.80
4 to 5 ns	1.77	1.76	1.93	1.77
5 to 6 ns	1.77	1.76	1.84	1.77
6 to 7 ns	1.76	1.77	1.89	1.78
7 to 8 ns	1.76	1.76	1.90	1.78
8 to 9 ns	1.76	1.76	1.84	1.78
9 to 10 ns	1.76	1.76	1.80	1.78

Setting these equations equal and solving for the spring constant, k_{HC}, gives:

$$k_{HC} = \frac{k_B T}{(d - d_0)^2}$$

Using the above relation allowed us to determine appropriate force constants from the unconstrained simulation to establish parameters for harmonic constraints to enforce the desired Zn^{II} coordination environment (**Table 4.7**). We note that, in the case of residue Asp74, both monodentate and bidentate modes were observed during coordination. For parameterization of this residue, we used distances and variations for the closest of the two oxygen atoms in a given structural conformation, noting that the identity of this atom could and did vary. These distances were used to derive a monodentate harmonic constraint (**Table 4.7**), which was applied to the oxygen atom closest in proximity to the Zn^{II} in the initial crystal structure. We found the application of harmonic constraints to be effective at enforcing the His63/His73/Asp74/His77 coordination and precluded coordination of Zn^{II} by additional water molecules, even when the tetramer was loaded with a single Zn^{II} ion. We anticipate that a similar treatment could be used to examine the interfacial coordination of metal ions on other cyt cb_{562} systems to examine their dynamic behavior.[15-20]

Solvation the protein interior. In order to effectively and efficiently capture the solution-phase dynamics of $^{C38/C81/C96}R1_4$, we sought to determine a solvation protocol that would solvate the tetramer to the extent most similar to the equilibrated state. The available crystal structures of the tetrameric $^{C38/C81/C96}R1_4$ in both its Zn-bound and metal-free states reveal large internal cavities, as well as interfacial crevices throughout the

Table 4. 7 Harmonic constraints for Zn^{II} binding to $^{C38/C81/C96}R1_4$.

Residue	Atom	d_0 (Å)	k_{HC} (kcal/mol•Å^2)
H63	εN	1.78	131
H73	εN	1.78	157
D74	δO	1.59	358
H77	εN	1.79	166

tetramer (**Figure 4.6a**). The cavity volumes identified by a 1.4 Å radius rolling probe were 212 $Å^3$ for the metal-free structure, and 575 $Å^3$ for the Zn-bound structure, as calculated by VOIDOO. This increase in internal cavity volume has been reported previously for the Zn-bound and metal-free structures of the related protein $^{C81/C96}R1_4$ as well.[16] The implications of differences in solvation on the energetics of the metal-bound and metal-free states are not investigated in this work, but it is hoped that molecular dynamics will present a useful computational tool for exploring this aspect of the disulfide-linked $R1_4$ protein assemblies.

We carried out molecular dynamics simulations of Zn-$^{C38/C81/C96}R1_4$ to determine how the number of internalized water molecules varied over time. When using only the repertoire of crystallographic waters to solvate the protein interior, and placing additional waters surrounding the protein tetramer, we found that the number of water molecules in the protein interior significantly increased over time, and was very slow to equilibrate (**Figure 4.7**). The program FLOOD[17] has been developed to solvate protein cavities and crevices, and used to place waters in the cavity of the $^{C38/C81/C96}R1_4$ interior. FLOOD placed 34 waters in the cavity of the metal-free $^{C38/C81/C96}R1_4$ structure, and 87 waters in the cavity of the Zn-bound structure (**Figure 4.6b**). Using FLOOD to place waters at the protein interior also gave an initial state that showed a less dramatic increase in the number of internalized waters over time, and more rapidly equilibrated compared to the simulation bearing only crystallographic waters in the tetramer interior (**Figure 4.7**). We estimated the number of cavity waters in these simulations using the selection criteria depicted in Figure 4.8**.**

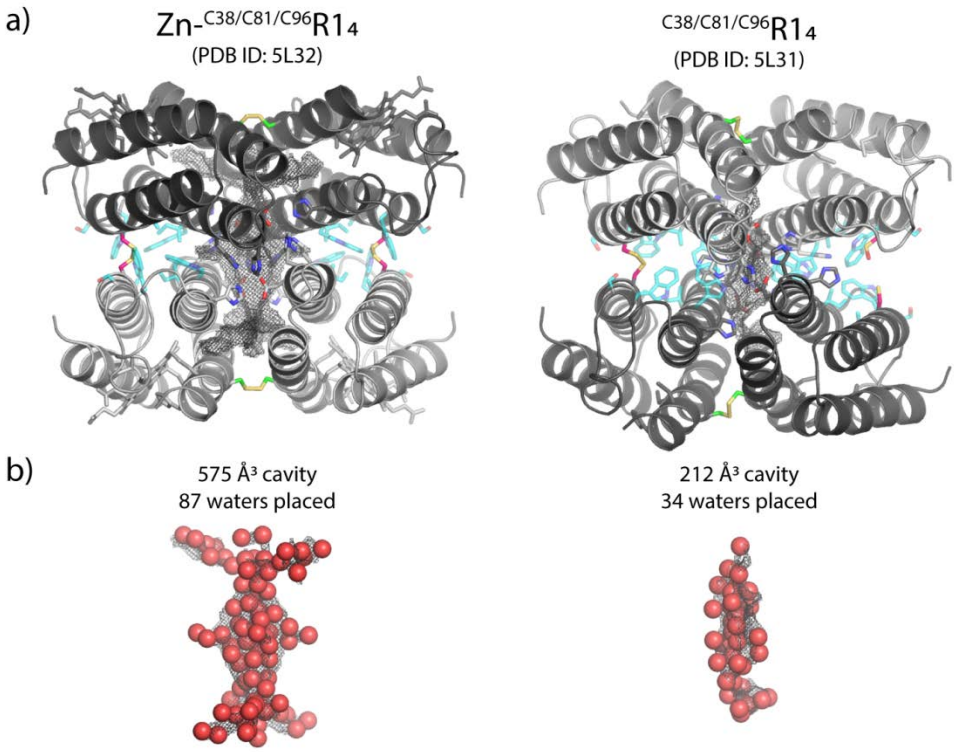

Figure 4.6 Solvating the internal of cavities in the $^{C38/C81/C96}R1_4$ structures.
(a) Central cavities of the available Zn-bound and metal-free structurs of $^{C38/C81/C96}R1_4$ (PDB ID 5L32 & 5L31, respectively). Cavities were identified using a rolling probe (r = 1.4 Å) that searched the tetramer interior beginning from a central point. The cavities identified in this manner are shown as a black mesh. (b) Cavities overlaid with the suite of waters placed in each structure by the program FLOOD.

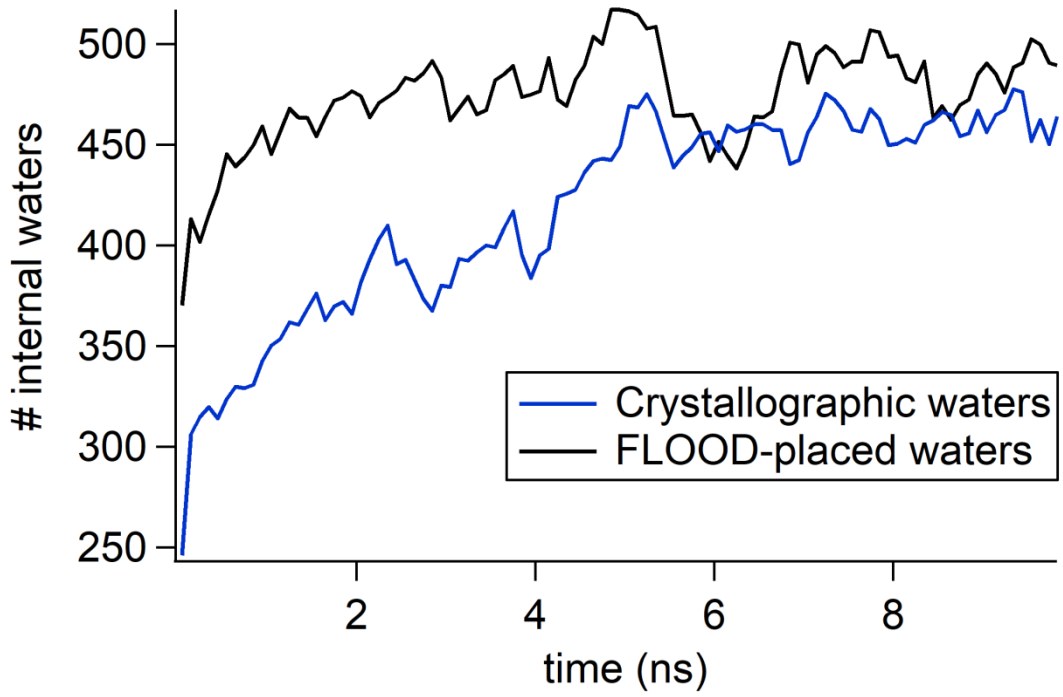

Figure 4.7 Equilibration of internal water number in Zn-$^{C38/C81/C96}$R1$_4$ MD simulations.
Simulated structures of Zn-$^{C38/C81/C96}$R1$_4$ were solvated using the SOLVATE plugin of VMD. The tetramer interiors were solvated using only the crystallographically observed waters (blue trace) or the waters placed in the internal cavity using FLOOD (black trace). Waters were assumed to be internalized if they were within 18 Å of all four protein chains, and within 18 Å of the centrally located residue 69.

4.3 Molecular dynamics simulations of C38-C38 disulfide breakage.

Identifying key structural changes in $^{C38/C81/C96}R1_4$ *opening.* We carried out molecular dynamics of two metal-free states of $^{C38/C81/C96}R1_4$, and examined the resulting ensembles of structures for significant structural changes. The first state was based on theZn-bound structure and contained the full suite of six disulfide bonds in the protein. The second was based on the metal-free structure of the protein bearing five intact disulfide bonds and a single broken C38-C38 disulfide bond (**Figure 4.9**). We reasoned that simulating these two states would allow the six-disulfide bearing structure to relax upon removal of Zn^{II} (a state not accessible by crystallography), and allow for direct examination of the structural changes that attend hydrolysis of a single C38-C38 disulfide bond in the tetramer.

We examined the structural ensembles for quantitative descriptors that could serve as reaction coordinates to describe the process of tetramer opening. In the available crystal structure of $^{C38/C81/C96}R1_4$, we observed an increase in the Cα distance between the two C38 residues ($d_{38\text{-}38}$) that undergo disulfide bond breakage, from a distance of 5.5 Å to 10.6 Å upon Zn^{II} removal. A similar trend was observed for the two metal-free simulations of $^{C38/C81/C96}R1_4$, where $d_{38\text{-}38}$ increased from 5.7 Å in the six-disulfide structural ensemble to 11.5 Å in the five-disulfide ensemble. We were interested in structural descriptors that encompassed larger collections of atoms in the two structural states, such as opposing pairs of α-helices within the tetramer. Using the naming convention outlined in **Figure 4.10**, the C38-C38 disulfide bond that undergoes breakage is contained in the Helix2-Helix2 pair of protein chains B and D. The distance between the centers of mass of these helices ($d_{H2\text{-}BD}$) was found to significantly increase from the

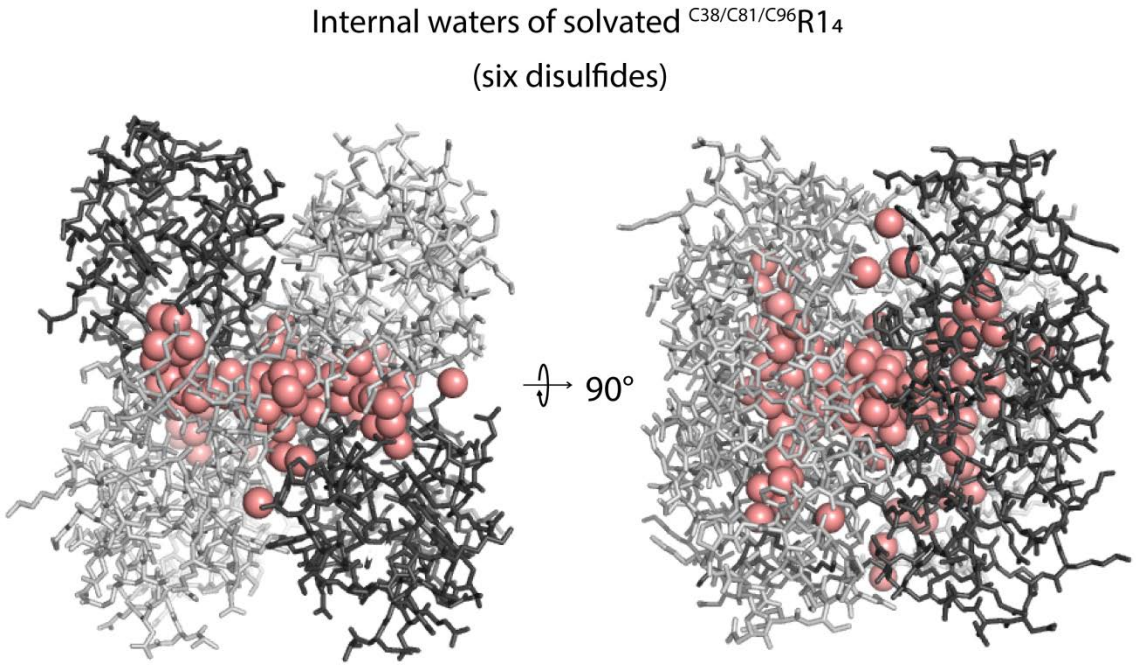

Figure 4.8 Selecting internalized waters in fully-solvated $^{C38/C81/C96}R1_4$.
The $^{C38/C81/C96}R1_4$ was solvated using flood, and placed in a cube of water 85Å to a side using SOLVATE in VMD. We regarded waters as internalized if they were within 18 Å of all four protein chains, and within 18 Å of the centrally located residue 69. Show as red spheres are water that satisfy these selection criteria.

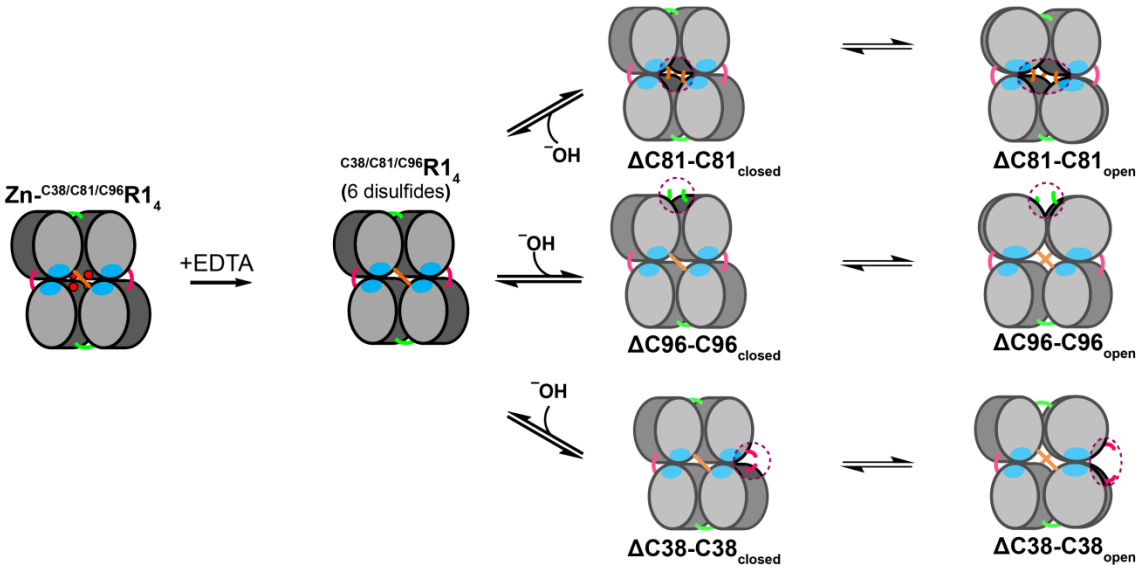

Figure 4.9 Metal-free states of $^{C38/C81/C96}R1_4$ simulated by all-atom molecular dynamics.
Removal of Zn^{II} from $^{C38/C81/C96}R1_4$ likely results in a transient metal-free intermediate state bearing six disulfide bonds. Reaction of hydroxide with a disulfide bond yields a tetrameric species bearing five disulfie bonds and in a closed state, which can further rearrange to adopt a more open configuration.

six-disulfide bond state (19.5 Å) to the five-disulfide bond state (27.1 Å) (**Table 4.8**). We measured the structural differences of the other pairs of *i1* interfacial helices, and the other largest change was found between the opposing pair of Helix3 segments in chains A and C, which contain the intact C38-C38 disulfide bond, where $d_{H3\text{-}AC}$ increased from 10.2Å to 13.4 Å (**Table 4.8**). Additionally, the centers of mass of the four proteins in the tetramer describe a dihedral angle. Applying the topology convention outlined in **Figure 4.10**, we found ψ_{4mer} significantly increased from the six-disulfide state (13°) in transitioning to the five-disulfide state (34°) (**Table 4.8**).These simple structural metrics illustrate the global nature of the structural change that occurs upon C38-C38 hydrolysis in $^{C38/C81/C96}R1_4$. All four protomers undergo structural rearrangement with respect to their *i1* binding partner. Additionally, physical intuition suggests that breakage of a single disulfide crosslink would permit a simple lever-like rearrangement, where the intact C38-C38 disulfide bond is the pivot point of the entire structure. The observed rearrangement is more complex than this, as shown by the relative magnitude of the helix 3 separations ($d_{H3\text{-}AC} > d_{H3\text{-}BD}$), which run counter to the trend expected from lever-like rearrangement. Even so, the engagement of both *i1*-bridged dimers in structural rearrang ment is consistent with the biochemical characterization of the C38-bearing R1 variants, which show that C38-C38 hydrolysis requires a crosslinked tetrameric architecture to proceed (see **Chapter 3**).

Molecular dynamics simulations of $^{C38/C81/C86}R1_4$ structural rearrangement. Using the metal-free, six-disulfide state of $^{C38/C81/C96}R1_4$ as the initial state, we set out to simulate the structural transitions that occur between the closed and open tetrameric states upon disulfide bond hydrolysis. The simulations employed in this work do not all-

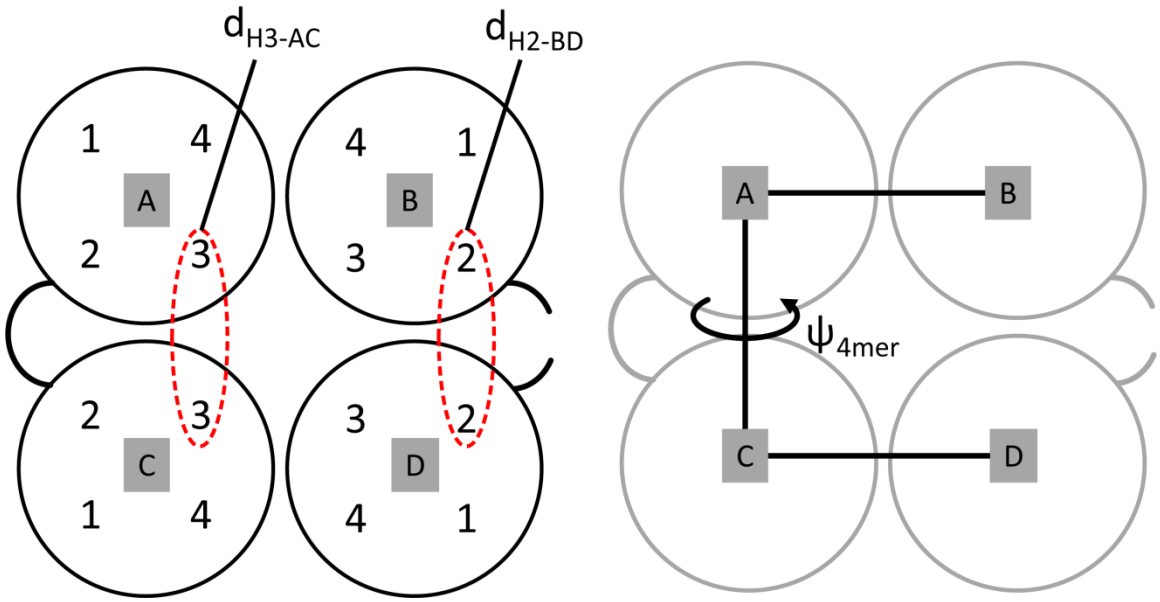

Figure 4.10 Topology and naming convention of key $^{C38/C81/C96}R1_4$ structural parameters.
The structures of $^{C38/C81/C96}R1_4$ used for molecular dynamics simulations in this work employ the topology naming convention depicted above, with gray boxes denoting the chain identities, and numbers depicting the individual α-helices of each protomer in the overall 16-helix bundle topology. Helix-helix distances that showed the pronounced difference in the two metal-free states of $^{C38/C81/C96}R1_4$ (d_{H3-AC} and d_{H2-BD}) are encircled in red. The dihedral angle described by the centers of masses of the four protomers as shown above, ψ_{4mer}, was also found to change significantly between the two states.

Table 4.8 Structural parameters of metal-free $^{C38/C81/C96}R1_4$.
Values reported in the table were averaged over 40 ns simulations of metal-free $^{C38/C81/C96}R1_4$ initialized in the indicated state. Parameter errors are the standard deviations of measurements made at 1 ps intervals during the course of the simulations.

	6-disulfide state	5-disulfide state	Difference
Cα separation (Å)			
$d_{38\text{-}38}$	5.7 ± 0.3	11.5 ± 0.4	5.8
Helix separation (Å)			
$d_{H2\text{-}BD}$	19.5 ± 0.3	27.1 ± 0.4	7.6
$d_{H3\text{-}AC}$	10.2 ± 0.4	13.4 ± 0.2	3.2
$d_{H2\text{-}AC}$	19.6 ± 0.3	19.5±0.4	-0.1
$d_{H3\text{-}BD}$	10.3 ± 0.4	11.4±0.3	1.1
Dihedral (°)			
Ψ_{4mer}	13 ± 1	34 ± 2	21

ow for modelling of bond breakage events. However, the chemical change of interest could be manually installed on the starting state, allowing us to observe the resultant structural changes *in silico*. We restarted the simulations following manual C38-C38 disulfide bond breakage, and observed rapid structural changes that corresponded to tetramer opening . The $d_{38\text{-}38}$ of the broken disulfide bond, and the distance between the flanking helices ($d_{H2\text{-}BD}$) sharply increased (**Figure 4.11a-b**). However, during the course of the simulations attempted, full tetramer opening was never observed, and the structure instead adopted a partially-opened intermediate state (**Figure 4.11a-b**). Moreover, the separation of the opposing helices, and the commensurate increase in $d_{H3\text{-}AC}$, was never observed in these simulations, and we instead noted a modest decrease in the distance between their centers of mass (**Figure 4.11c**). The dihedral described by the tetramer as a whole, ψ_{4mer}, increased toward the value expected from the 5-disulfide simulation of $^{C38/C81/C96}R1_4$, but also suggested stalling at an intermediate species partway between the two previously-observed states (**Figure 4.10d**). Though far from a full depiction of the structural transition that occurs upon C38-C38 disulfide hydrolysis, these results suggested that a significant driving force leads to a structural ensemble where the resulting Cys38 and Csa38 residue are outside of reasonable disulfide bonding distance ($d_{38\text{-}38} \geq 7$ Å).[18] Additionally, bond breakage induces some degree of local structural elasticity, as can be inferred from the large changes in $d_{38\text{-}38}$ and $d_{H2\text{-}BD}$. However, the overall transition from the six-disulfide state to the five-disulfide state suggested by the crystal structures involves changes throughout the entire tetramer. Moreover, these structural changes do not seem to be accessible by the stochastic sampling of conventional molecular dynamics, as the $d_{H3\text{-}AC}$ and ψ_{4mer} parameters do not trend signify-

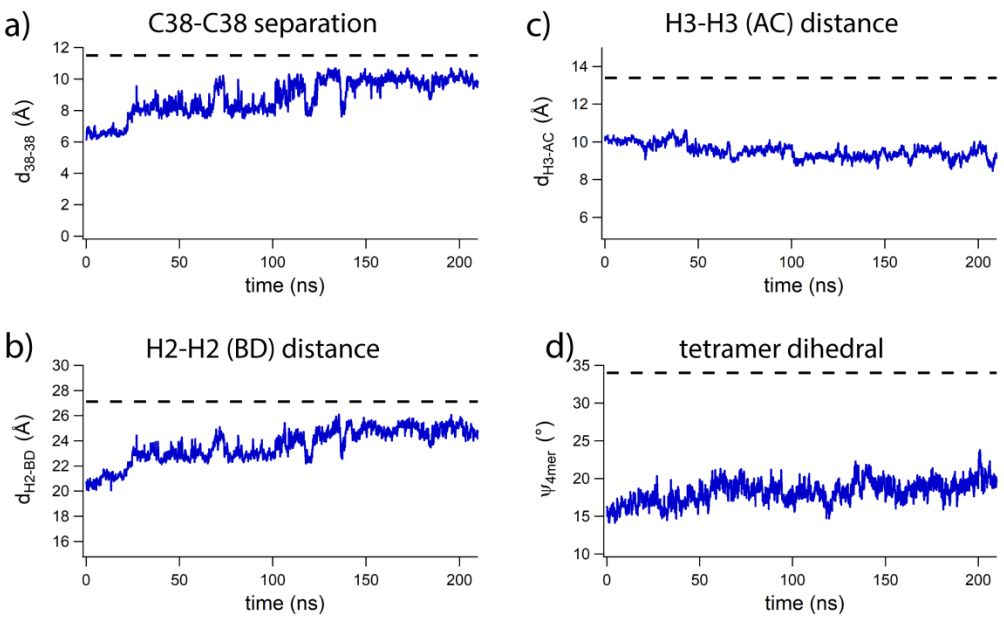

Figure 4.11 Changes in $^{C38/C81/C96}R1_4$ structural parameters following simulated bond breakage. Structural parameters of metal-free $^{C38/C81/C96}R1_4$ initially simulated in the metal-free, six-disulfide bonded state following manual breakage of the C38-C38 disulfide bond linking chains B and D. Blue lines are 100 ps averages of measurements made at 1 ps invervals. Black lines denote the target values expected from simulating metal-free $^{C38/C81/C96}R1_4$ initialized in the five-disulfide state.

cantly toward the values observed for the five-disulfide bond simulation built from the metal-free $^{C38/C81/C96}R1_4$ crystal structure.

Accelerated molecular dynamics of $^{C38/C81/C96}R1_4$ opening. Effective sampling of rare events is a fundamental problem facing molecular dynamics simulations.[19, 20] The potential energy surfaces of structural changes in proteins are frequently rough, and can feature many local minima separated by energy barriers that present kinetic barriers to effective sampling. Therefore, there have been significant efforts to address this problem through enhanced sampling methods. Steered molecular dynamics (see below) can furnish a starting series of structures that sample a transition of interest, which can be used to get thermodynamic information by umbrella sampling.[21] Replica exchange employs parallel simulations at different temperatures with defined exchange points and probabilities, giving thermodynamic information in spite of such kinetic barriers.[22] A swarm trajectory approach has been developed to examine structural transitions, where short unbiased trajectories are carried out iteratively to obtain a transition path.[23] Here, we explored the use of Gaussian Accelerated Molecular Dynamics (GaMD) as a method of enhanced sampling of $^{C38/C81/C96}R1_4$ that would allow us to recover thermodynamic information on the observed structural change.[24] Briefly, this method carries out a short unbiased trajectory to determine an appropriate boost potential for the system. The boost potential improves the transition probabilities over energy barriers, and can be accounted for to estimate the unbiased potential energy landscape of the system.[24]

We set out to use GaMD as a means to sample the transition path between the two structural states observed for $^{C38/C81/C96}R1_4$. We carried out GaMD simulations on metal-free $^{C38/C81/C96}R1_4$ following manual disulfide-bond breakage. We carried out

conventional molecular dynamics (cMD) simulations with a 400 ps preparatory phase where GaMD parameters were not estimated, followed by 2 ns of cMD for estimating GaMD parameters (**Table 4.9**). In addition to the default boost parameters that are applied to the total potential energy of the system, we explored dihedral-only and dual-boost parameters (**Table 4.9**).[24] We note that the structural changes observed between the two states include changes to the dihedral angles of the remaining disulfide bonds, and reasoned that boost potentials applied to the dihedral parameters could be effective. Significant rapid structural opening was found to occur during the preparatory phase of the GaMD when parameter estimates were made, but output structures were not written (**Figure 4.12**). Therefore, we also carried out GaMD simulations with no preparatory phase, and which applied the boost parameters determined from a separate simulation. We note that this treatment of the system likely undermines the thermodynamic rigor of the GaMD acceleration, but was appropriate for obtaining a simulation trajectory of the full structural transition. While we did observe changes in some structural parameters ($d_{38\text{-}38,}$ $d_{H2\text{-}BD}$) that suggested an approach toward the 5-disulfide structural state, we again saw the simulations stall out at a structural intermediate (**Figure 4.13**).

Steered molecular dynamics of $^{C38/C81/C96}R1_4$ *opening.* To more rapidly sample the structural transition that occurs in $^{C38/C81/C96}R1_4$ following *in silico* disulfide bond breakage, we set up pulling simulations of the protein. These simulations monitor one or more parameters of interest relative to some specified target values. Deviation of the parameter from the target value will develop a force that biases the stochastic conformational sampling toward the target value. Additionally, we varied the target value during the course of such simulations in order to carry out simulated pulling of

Table 4.9 GaMD boost parameters for simulations of $^{C38/C81/C96}R1_4$ opening.
GaMD boost parameters for metal-free $^{C38/C81/C96}R1_4$ simulations carreid out immediately following manual disulfide bond breakage. Parameters were obtained from a $1x10^6$ timestep cMD simulation that followed a $2x10^5$ timestep preparatory phase.

boost mode	E_{boost} (kcal/mol)	k_0
dihedral	2120	0.0000198
total	-187000	0.0000172
dual (dihed.)	2150	0.0027620
dual (total)	-189000	0.0000165

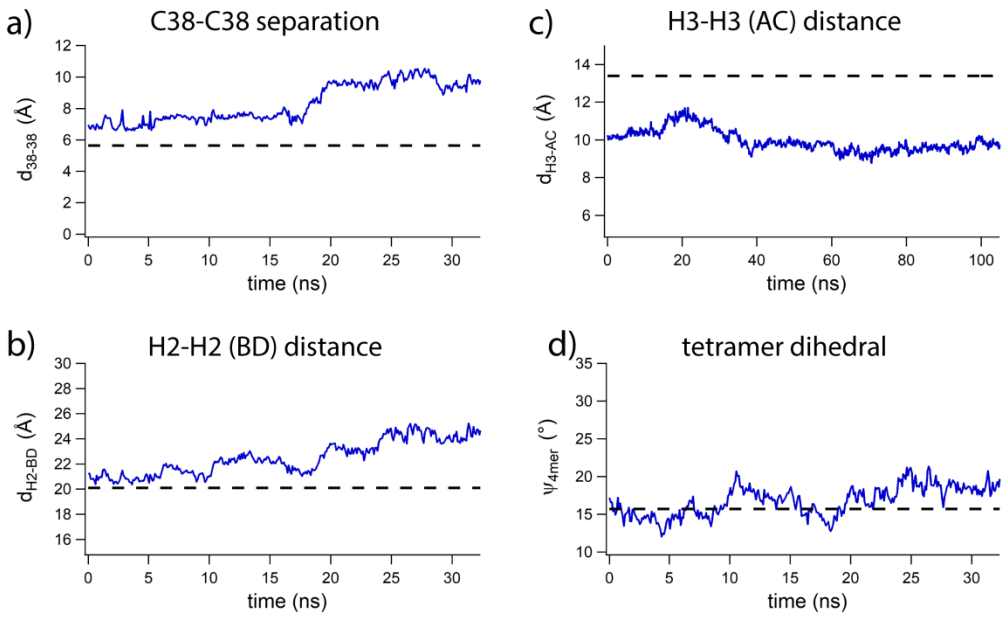

Figure 4.12 Structural changes in $^{C38/C81/C96}R1_4$ during GaMD following a preparatory simulation. GaMD simulation of metal-free $^{C38/C81/C96}R1_4$ initially simulated in the metal-free, six-disulfide bonded state following manual breakage of the C38-C38 disulfide bond linking chains B and D. Blue lines are 100 ps averages of measurements made at 1 ps invervals. Black lines denote the starting values of the structural parameters preceding the preparatory cMD simulation, and during which output structures are not saved. Data shown were collected using the total energy boost mode of GaMD.

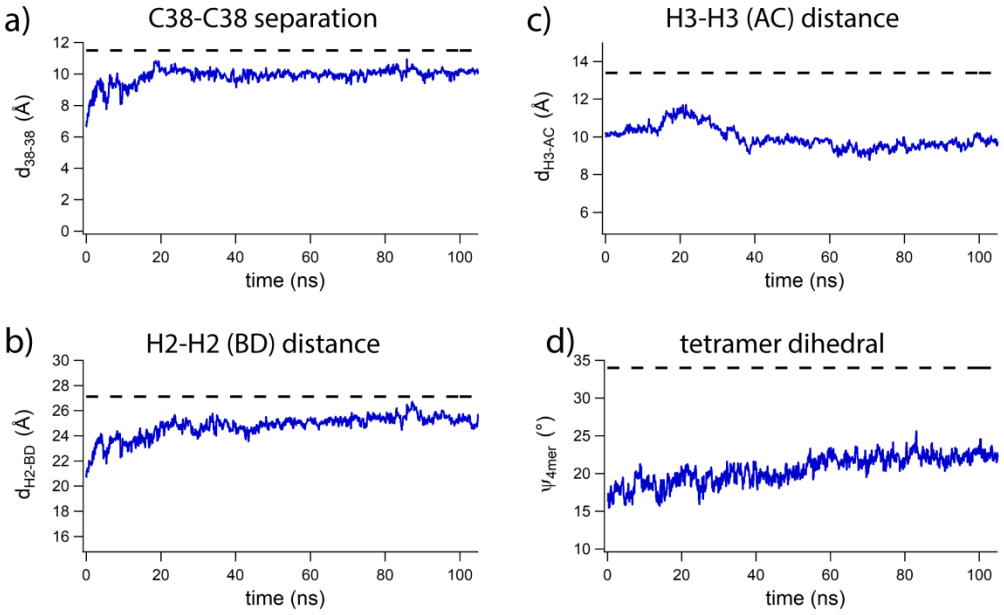

Figure 4.13 Structural changes in $^{C38/C81/C96}R1_4$ during GaMD with preparatory phase omitted. Structural parameters of metal-free $^{C38/C81/C96}R1_4$ initially simulated in the metal-free, six-disulfide bonded state following manual breakage of the C38-C38 disulfide bond linking chains B and D. Blue lines are 100 ps averages of measurements made at 1 ps invervals. Black lines denote the target values expected from simulating metal-free $^{C38/C81/C96}R1_4$ initialized in the five-disulfide state. Data shown were collected using the total energy boost mode of GaMD.

$^{C38/C81/C96}R1_4$ toward the open state observed in the five-disulfide bond simulation. Initially, we carried out pulling experiments with respect to the $d_{C38-C38}$ parameter. However, we found this led to a highly localized application of force that would lead to buckling of the α-helices. Therefore, we next used the structural parameters discussed in **Table 4.8** in order to carry out steered MD simulations. Given that the d_{H3-AC} parameter actually moved farther away from the target value in both cMD and GaMD simulations, we reasoned that monotonic pulling from the fully-closed to the fully-open state would likely be ineffective. However, when carrying out a steered MD simulation using a less localized structural parameter, d_{H2-BD}, we found that the unconstrained parameters, in this case d_{H3-AC} and ψ_{4mer}, showed no sign of converging toward the target values prescribed by the five-disulfide state simulation (**Figure 4.14a-c**). Similarly, constraining d_{H3-AC} could drive the distances toward the desired target value, but the unconstrained d_{H2-BD} and ψ_{4mer} did not improve appreciably, suggesting that the structure was being driven toward an artefactual end state (**Figure 4.14 d-f**). Attempts to constrain all three parameters (d_{H2-BD}, d_{H3-AC}, and ψ_{4mer}) could, as expected, steer each toward its respective target value (**Figure 4.15**). However, removal of one of the constraints, such as enforcement of ψ_{4mer}, revealed such end states to be far from equilibrated (**Figure 4.15c**), indicating that the structure was satisfying the criteria of the constraints in an artefactual manner. In all the cases above, inspection of the structures in these simulations revealed that the desired structural ensemble was not achieved. Emblematic of this was the failure to attain the correct *i1* interfacial packing between chains B and D, which was expected to include the π- stacking interaction between the Trp66/Trp66 pair observed *in crystallo* (see **Figure 2.8e**). As such, these trajectories were not regarded as suitable for obtaining thermodyn-

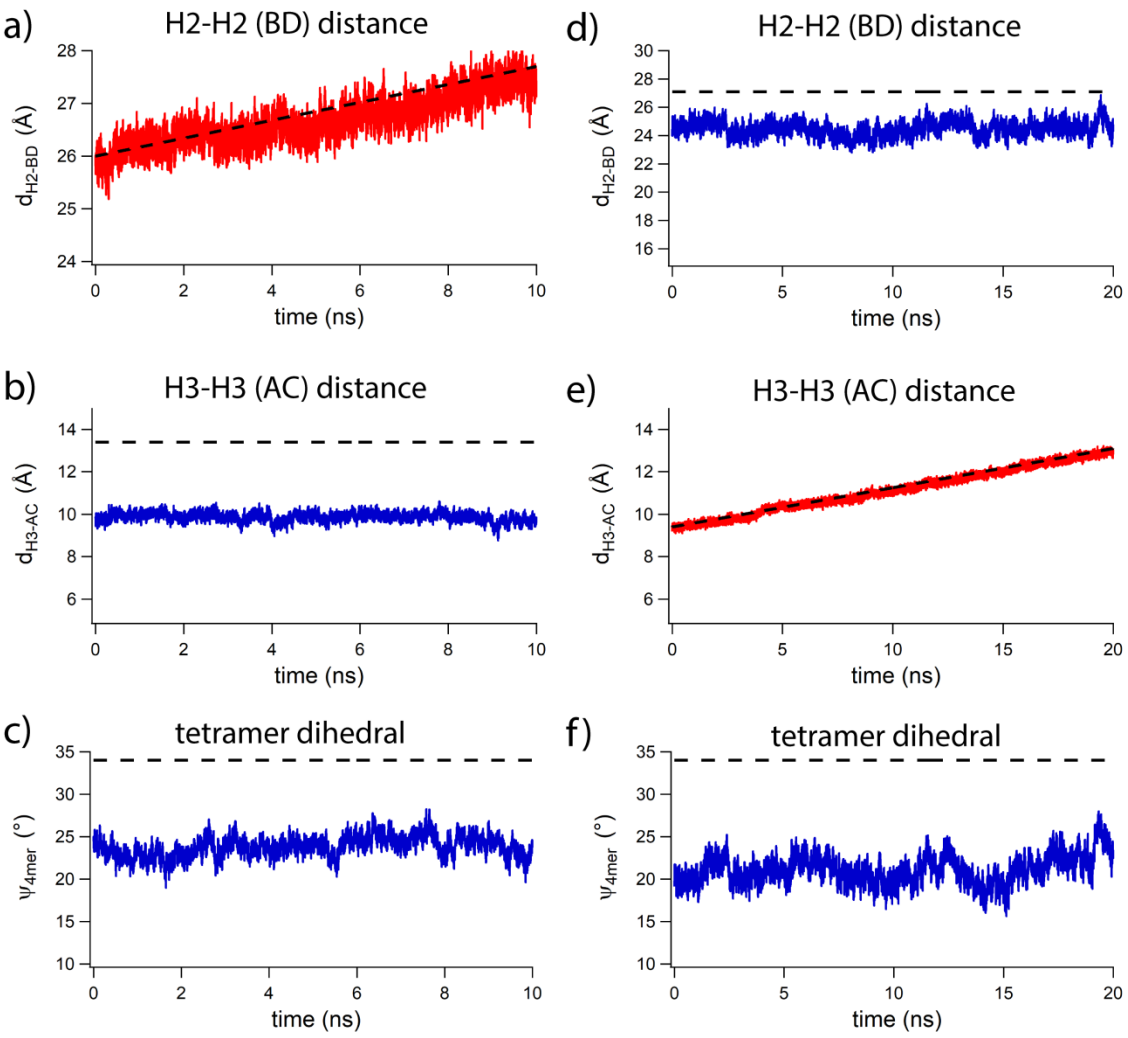

Figure 4.14 Structural changes in $^{C38/C81/C96}R1_4$ from constant-velocity pulling of a single helix pair. Simulations were carried out on $^{C38/C81/C96}R1_4$ from a partially opened intermediate achieved during MD simulations following manual disulfide bond breakage. (a-c) Changes to structural parameters arising from constant-velocity pulling with respect to d_{H2-BD}, where k_{pull} = 10 kcal/mol. (d-f) Changes to structural parameters arising from constant-velocity pulling with respect to d_{H2-BD}, where k_{pull} = 50 kcal/mol. Constrained structural parameters are plotted in red, and unconstrained parameters are plotted in blue. Dashed lines show the target value of constrained parameters, or the end-state value expected from the five-disulfide simulation for unconstrained parameters.

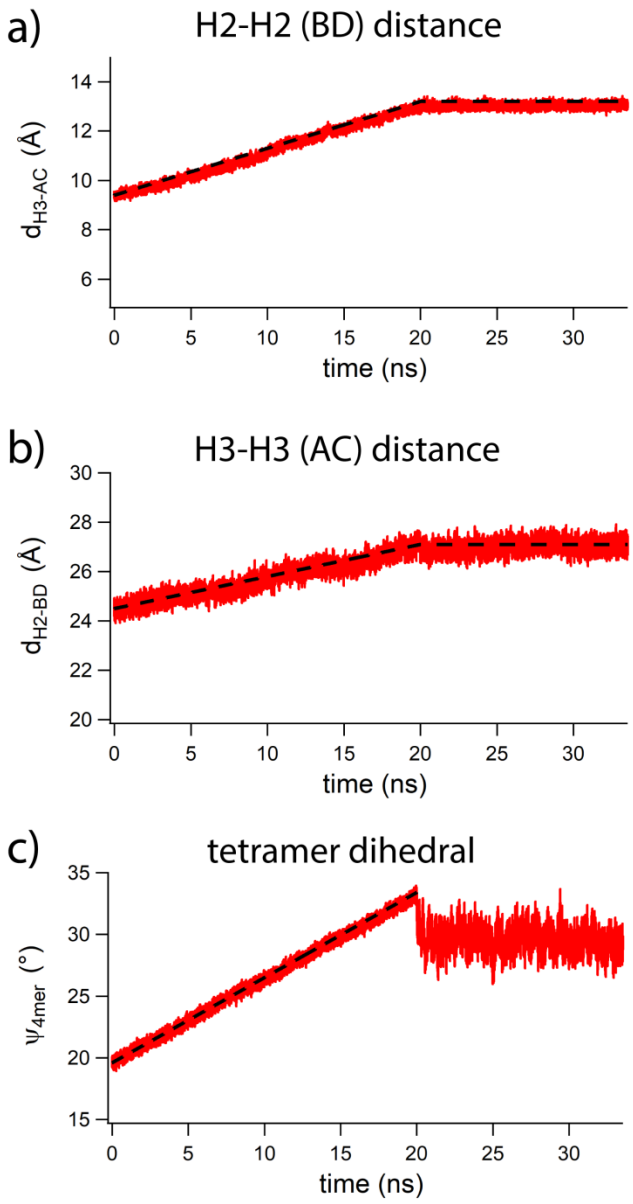

Figure 4.15 Structural changes in $^{C38/C81/C96}R1_4$ from multi-parameter constant-velocity pulling. Simulations were carried out on $^{C38/C81/C96}R1_4$ from a partially opened intermediate achieved during MD simulations following manual disulfide bond breakage. Constant velocity pulling was carried out simultaneously with respect to d_{H2-BD} (k_{pull} = 10 kcal/mol), d_{H3-AC} (k_{pull} = 50 kcal/mol), and ψ_{4mer} (k_{pull} = 10 kcal/mol) during a 20 ns simulation. At this time, the constraint for ψ_{4mer} was removed, while the constraints for $d_{H2=BD}$ and d_{H3-AC} were maintained at a fixed target distance for an additonal 10 ns simulation.

amic information on the structural changes that attend C38-C38 hydrolysis in $^{C38/C81/C96}R1_4$. However, they do illustrate the intricacy and complexity of the structural rearrangement that occurs in $^{C38/C81/C96}R1_4$, and further underscore the global nature of this transition.

4.4 Molecular dynamics simulations of C81-C81 and C96-C96 breakage.

Molecular dynamics allows us to access protein states that are not experimentally observed to better understand those that do arise. In the case of $^{C38/C81/C96}R1_4$, we were interested in exploring the structural changes that would attend C81-C81 breakage and C96-C96 breakage, if such an event were to occur. Here, again, we used a Zn-free simulation of $^{C38/C81/C96}R1_4$ in the six-disulfide state as a starting point, and manually broke one copy of each disulfide bond, giving a Cys and Cso pair. In the case of C81-C81 breakage, we observed rapid and reversible opening of the tetrameric structure, which can be described by the Cα separation of the residues at site 81 ($d_{81\text{-}81}$) (**Figure 4.16a**). Assuming the observed transitions constituted a representative equilibrium description of the system, the free energy of the system with respect to a chosen reaction coordinate (ξ) can be determined. The relationship between the probability of adopting a particular microstate, P(ξ), and the free energy of the can be obtained from the partition function, and stated as follows:

$$\Delta G_{state} = -k_b T \ln P(\xi)$$

We binned the simulated structures with respect to $d_{81\text{-}81}$ in 0.5 Å increments (**Figure 4.16b**) and, using the above relationship, determine the potential of mean force (PMF) of structural opening with respect to the chosen reaction coordinate, in this case $d_{81\text{-}81}$ (**Figure 4.16c**). This analysis revealed a slight decrease in free energy upon open-

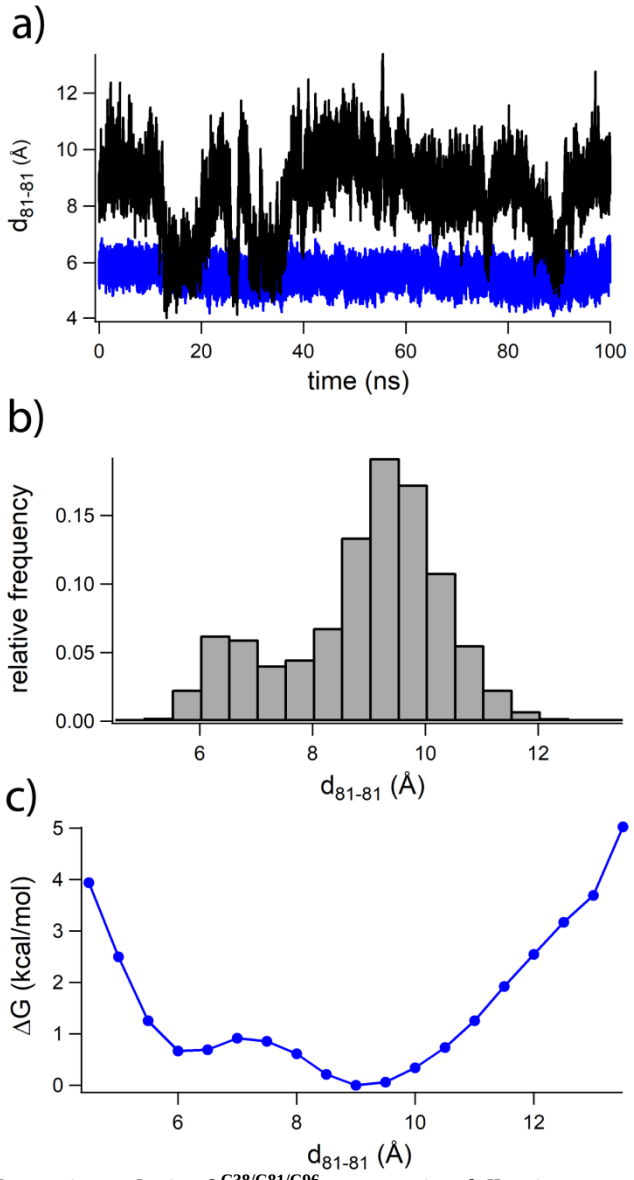

Figure 4.16 Thermodynamic analysis of $^{C38/C81/C96}R1_4$ opening following C81-C81 breakage.
(a) Cα separations of C81-C81 residue pairs forming an intact (blue) or broken (black) disulfide bond in $^{C38/C81/C96}R1_4$. (b) Probability distribution of $d_{81\text{-}81}$ for the broken disulfide bond, binned in 0.5 Å increments. (c) PMF of $^{C38/C81/C96}R1_4$ opening following C81-C81 breakage. Points correspond to the relative free energies estimated for each histogram bin, and the solid line shows interpolated values.

ing (ΔG_{open} = -0.13 kcal/mol) and a modest activation barrier ($\Delta G_{barr.}$ = 0.72 kcal/mol) to tetramer closure. Therefore, in the event of C81-C81 breakage, tetramer closure could readily occur due to thermal motion of the protein.

Breakage of a C96-C96 disulfide bond in metal-free $^{C38/C81/C96}R1_4$ also led to significant structural changes. Here, we monitored structural differences using the Cα separation of the 96-96 residue pair ($d_{96\text{-}96}$). Here, we observed tetramer opening, resulting in a significantly increased $d_{96\text{-}96}$ that continued to fluctuate, but never returned to a fully-closed state (**Figure 4.17a**). Generating a histogram by binning $d_{96\text{-}96}$ in 0.5 Å increments (**Figure 4.17b**) and determining the PMF with respect to $d_{96\text{-}96}$ (**Figure 4.17c**) revealed a single-well potential that favored an open state, with the free energy minimum occurring at $d_{96\text{-}96}$ = 11 Å (**Figure 4.17c**). If the free energy estimates for the most-closed state are taken as an estimate for the free energy of opening, we obtain $\Delta G_{open} \leq$ -5.9 kcal/mol. This is significantly larger than the value estimated for C81-C81 opening, and is unlikely to be surmounted by thermal energy. However, this value needs to be considered in the full energetic context of disulfide bond hydrolysis, which itself is not a thermodynamically favorable process. Therefore, we think it likely that C81-C81 bond breakage is not observed because the subsequent structural rearrangement provides insufficient energetic payoff to compensate for the penalty associated with bond breakage. Moreover, we anticipate that extending the simulation of $^{C38/C81/C96}R1_4$ bearing a broken C96-C96 bond should lead to population of states that permit reformation of the disulfide bond. An extension of the simulation is also desirable to further justify the underlying assumption of thermal equilibrium of the system that was made in carrying out this analysis.

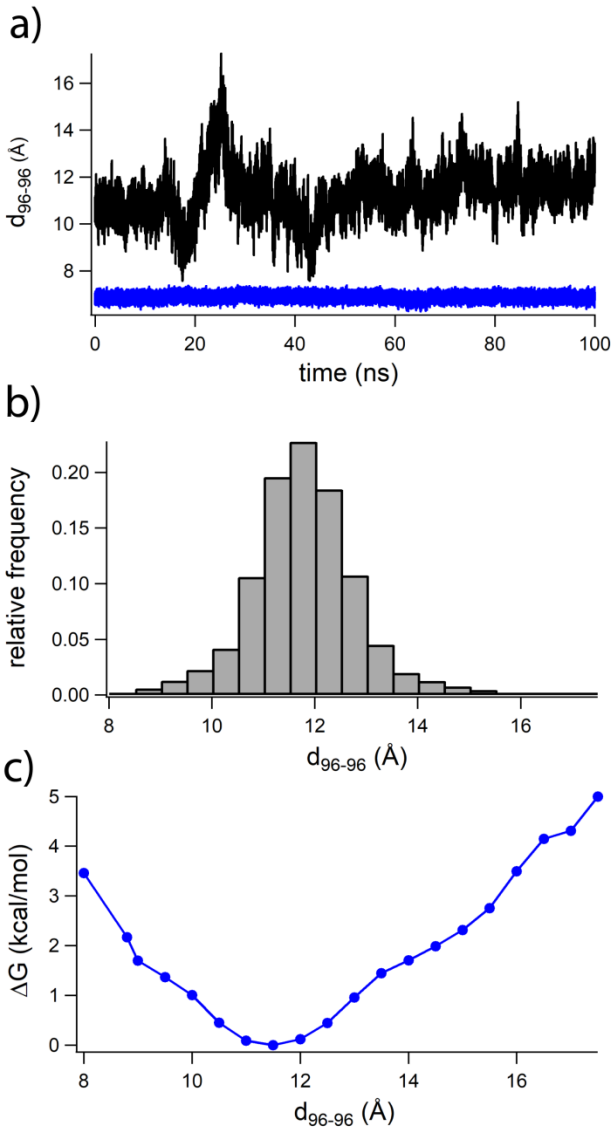

Figure 4.17 Thermodynamic analysis of $^{C38/C81/C96}R1_4$ opening following C96-C96 breakage.
(a) Cα separations of C96-C96 residue pairs forming an intact (blue) or broken (black) disulfide bond in $^{C38/C81/C96}R1_4$. (b) Probability distribution of $d_{96\text{-}96}$ for the broken disulfide bond, binned in 0.5 Å increments. (c) PMF of $^{C38/C81/C96}R1_4$ opening following C96-C96 breakage. Points correspond to the relative free energies estimated for each histogram bin, and the solid line shows interpolated values.

4.5 End-state thermodynamic analysis of C38-C38 disulfide breakage.

In light of the difficulties associated with efficiently simulating the transition from closed $^{C38/C81/C96}R1_4$ to the open state following C38-C38 disulfide bond breakage, we turned our attention to thermodynamic analysis of the end states of the structural transition. For this, we utilized the molecular dynamics simulations of metal-free $^{C38/C81/C96}R1_4$ in two states: a tetramer bearing a pair of intact C38-C38 crosslinks (a state not readily accessible by x-ray crystallography) with six disulfide bonds in total, and a tetramer bearing a single hydrolyzed C38-C38 disulfide bond with five remaining crosslinks (**Figure 4.8**). Simulating this pair of states elucidated the structural changes that occur following Zn^{II} removal, as well as the structural differences that can only follow disulfide bond hydrolysis. One aspect of structural relaxation following Zn^{II} is the dihedral strain of the disulfide bonds. The large difference in strain of the five disulfide bonds that persist between the Zn-bound and Zn-free crystal structures (ΔE_{strain} = -48 kJ/mol) is diminished when comparing the two simulated metal-free states of $^{C38/C81/C96}R1_4$ (ΔE_{strain} = -15 kJ/mol) (**Table 4.10**). However, focusing on this single energetic parameter does not fully capture the potential energy of the disulfide bonds, which includes many additional configurational and non-bonded energy parameters. Additionally, alleviation of disulfide bond strain is unlikely to be the sole driver behind C38-C38 disulfide bond hydrolysis following Zn^{II} removal. Therefore, we sought to compute energetic descriptors of the other key residues in $^{C38/C81/C96}R1_4$ as well.

Using the simulated structures of $^{C38/C81/C96}R1_4$, we computed the change in the potential energy of the engineered residues (the *i1* contacts, the disulfide bonds, and the metal binding ligands). This was accomplished by including configurational energetic erms (bonds, angles, etc.) and non-bonded interaction terms (electrostatics, van der Waals, etc.) of the entire protein structure, while excluding the C38-C38 disulfide bond modeled as undergoing hydrolysis. This excluded any systematic differences in the calculated potential energy that arose solely from the changes in type and number of atoms, and instead provided an estimate in potential energy difference arising from how a constant set of atoms was arranged ($E_{struct.}$). We computed $E_{struct.}$ for the disulfide bonds, *i1* interfacial residues, and metal ligands in the two states, and found $\Delta E_{struct.}$ for these engineered residues was well-converged during the 300 ns simulations (**Figure 4.18**).

We determined $\Delta E_{struct.}$ for individual sets of key residues, (the *i1* interfacial residues, the metal-binding ligands, and the disulfide bonds) to determine which residues experienced favorable changes in potential energy when undergoing the observed structural rearrangement that attends C38-C38 hydrolysis. Consistent with our analysis of disulfide bond strain, we found that rearrangement of the disulfide bonds to favored the structural rearrangement that attends C38-C38 hydrolysis ($\Delta E_{struct.}$ = -6.5 kJ/mol) (**Table 4.11**). By contrast, the change in potential energy of the metal-binding ligands disfavored structural rearrangement ($\Delta E_{struct.}$ = +43.4 kJ/mol), meaning they are unlikely to drive of disulfide bond hydrolysis (**Table 4.11**). Of the three sets of residues we interrogated, the most favorable potential energy difference attended the *i1* residues ($\Delta E_{struct.}$ = -164.5 kJ/mol), implicating the rearrangement of these interfaces as a major driver of C38-C38 breakage (**Table 4.11**). We further decomposed the change in potential energy of the *i1*

Table 4.10 Average disulfide strain in simulations of metal-free $^{C38/C81/C96}R1_4$.

Disulfide site	Average strain energy (kJ/mol)	
	$^{C38/C81/C96}R1_4$ (6 disulfides)	$^{C38/C81/C96}R1_4$ (5 disulfides)
C38-C38	21	15
C81-C81	17	16
C96-C96	13	10
5-bond sum	81	66

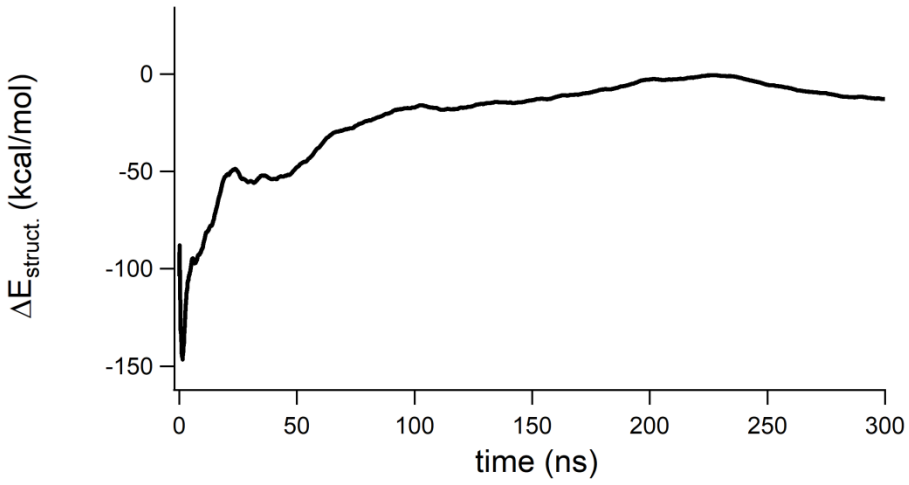

Figure 4.18 Convergence of structural potental energy for $^{C38/C81/C96}R1_4$ simulations.
Potential energy of engineered residues (intact disulfide bonds, *i1* interfacial residues, and metal ligands) in metal-free $^{C38/C81/C96}R1_4$ were calculated for simulations in the five-disulfide and the six-disulfide state. The running average of the difference in structural potential energy (ΔE_{struct}) is plotted for the transition from the six-disulfide to the five-disulfide state. This energy estimate appeared stable for the pair of 300 ns simulations.

Table 4.11 Potential energy differences of key residues in metal-free $^{C38/C81/C96}R1_4$ opening.

	$\Delta E_{struct.}$(kJ/mol)		
Protein feature	**Configurational**	**Non-bonded**	**Total**
Disulfide bonds	-24	17	-7
Metal ligands	2	42	43
i1 Residues	-41	-123	-165
Opposing *i1*	-9	-120	-129
Adjacent *i1*	-32	-8	-40
i1-*i1* interaction	-	-	-4

residues, and found large favorable contributions from both the interface containing the broken disulfide bond ($\Delta E_{struct.}$ = -39.6 kJ/mol) and the interface containing the intact C38-C38 disulfide bond ($\Delta E_{struct.}$ = -128.6 kJ/mol) (**Table 4.11**). We note that these terms do not sum to the total reported for both interfaces due to non-bonded interactions between the two interfaces, which account for the small discrepancy of ~4 kJ/mol (**Table 4.11**).

The potential energy calculations performed on the $^{C38/C81/C96}R1_4$ MD simulations provide insights into the energetic drivers of C38-C38 disulfide bond hydrolysis and subsequent structural rearrangement. Though modest relaxation occurs upon Zn^{II} removal, the disulfide bonds and the *i1* achieve more favorable arrangements, in terms of potential energy, following C38-C38 hydrolysis. The importance of the *i1* residues suggests why the C38-C38 disulfide bond, which is embedded in *i1*, hydrolyzes upon Zn^{II} removal rather than any of the other crosslinks. Additionally, the favorable changes in $\Delta E_{struct.}$ span both *i1* interfaces, consistent with the fact that C38-C38 hydrolysis was not observed in $^{C38}R1_2$, where rearrangement of both interfaces was no longer coupled due to removal of the *i2*-spanning disulfide crosslinks. Additionally, one would expect relaxation of both interfaces following hydrolysis of a single C38-C38 bond in $^{C38/C96}R1_4$ and $^{C38/C81/C96}R1_4$ to account for the fact that both bonds were never observed to hydrolyze: the requisite driving force arises from, and diminishes throughout, the tetramer as a whole. Rearrangement of a single protein interface is not sufficient to drive C38-C38 hydrolysis, and coupling of the two interfaces by C96-C96 disulfide bonds provides the requisite driving force for breakage of one, but only one, disulfide bond.

4.5 Methods.

The crystal structure of Zn-$^{C38/C81/C96}R1_4$ (PDB 5L32) was used as the starting model for simulating $^{C38/C81/C96}R1_4$ with six disulfide bonds, and the crystal structure of metal-free $^{C38/C81/C96}R1_4$ (PDB 5L31) was used as the starting model for $^{C38/C81/C96}R1_4$ with five disulfide bonds. In the latter case, an unmodeled loop (residues 46-50) in the 5L31 structure was repaired using PyMOL.[25] In the final treatment of the systems, heteroatoms and crystallographic waters from both structures were removed, the tetramer interiors were measured using a 1.4 Å rolling probe in VOIDOO, and the internal cavities were solvated using FLOOD.[17] The two structures were parameterized for the CHARMM27 forcefield,[8, 9] with additional parameters previously reported for an oxidized c-type heme and its axial ligands.[26] We also utilized custom CHARMM-compatible forcefield parameters for the sidechains of cysteine sulfenic acid (Cso) and cysteine sulfonate (Csa) (see **section 4.2**). Missing hydrogens were added and atom types, and their associated parameters, were assigned to the system in VMD using the PSFGEN,[27] and solvated in an 85 Å x 85 Å x 85 Å cube of TIP3P water using SOLVATE[10], leaving at least 10 Å of space on all sides in both states. Ions were placed using the Autoionize plugin[27] to simulate a 150 mM NaCl solute concentration and to neutralize the system. For the metal-free six-disulfide $^{C38/C81/C96}R1_4$ system, 71 sodium ions and 47 chloride ions were added, while 72 sodium ions and 47 chloride ions were added to the five-disulfide-bearing tetramer. Final builds of the solvated six- and five-disulfide bonded $^{C38/C81/C96}R1_4$ contained 56,796 atoms and 56,811 atoms, respectively.

All molecular dynamics simulations were carried out on a home-built computer using NAMD2.10 with multi-core and CUDA support.[28] Simulations were carried with a canonical NVT ensemble, with constant pressure (1.01325 bar) maintained by the Nosé-Hoover method (piston period = 100 fs; piston decay = 50 fs). Temperature was maintained at 300 K by Langevin dynamics with a damping coefficient of 1 ps^{-1}. Non-bonded interactions were evaluated out to a 14.0 Å cutoff, with a switching function applied for distances exceeding 10.0 Å, and with long-range electrostatic interactions calculated using the Particle mesh Ewald method. Hydrogen bond lengths were held fixed using the SHAKE algorithm.[29] Periodic boundary conditions were used for all simulations, and the dynamics trajectory was propagated using the velocity Verlet algorithm with a time step of 2 fs.

Initially, all protein and heme atoms were held fixed, and the systems were subjected to 1,000 steps of steepest-descent energy minimization, followed by a further 5,000 step minimization with no fixed atoms. The systems were equilibrated for 5 ns (2.5×10^6 steps), at which point the structures appeared to have relaxed. Subsequently, 300 ns simulations were carried out at a rate of about 20 ns simulated time per day, with atom coordinates saved every 500 steps (1 ps). Potential energies were computed using the NAMDEnergy plugin in VMD. Potential energies were calculated for the *i1* interfacial residues (residues A34, W41, S42, W66 and I69), the metal binding ligands (residues H63, H73, D74, and H77) and the disulfide bonds (residue C38, C81, and C96). The C38-C38 pair modeled as broken in the five-disulfide simulations, and its intact counterpart in the six-disulfide simulation, were excluded from all potential energy calculations, as were all solvent atoms. Differences in structural potential energy

($\Delta E_{struct.}$) calculated as the difference in potential energy for the internal and interaction energies of the residues of interest in the two simulated states. Convergence of ΔE_{struct} for all three sets of amino acids was monitored by computing the running average (*e.g.*, the n^{th} point is the average of the first n frames of the simulation) from the atom coordinates over the 300 ns trajectories, which was plotted in **Figure 4.18**. The computed $\Delta E_{struct.}$ values averaged over the entire 300 ns trajectories for each set of residues are reported in **Table 4.11**.

Chapter 4 was reproduced in part, with permission, from a manuscript currently being prepared for submission for publication: Churchfield, L.A.; Williamson, L.A.; Alberstein, R.G; Tezcan, F.A. Investigating the structural and mechanistic basis of allostery in a designed protein scaffold.

The dissertation author is primary author on all reprinted materials.

References

(1) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014**, *508*, 331.

(2) Klinman, J. P.; Kohen, A. *Journal of Biological Chemistry* **2014**, *289*, 30205.

(3) Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C. *Accounts of Chemical Research* **2016**, *49*, 809.

(4) Marsh, J. A.; Teichmann, S. A. *PLOS Biology* **2014**, *12*, e1001870.

(5) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433.

(6) Hertig, S.; Latorraca, N. R.; Dror, R. O. *PLoS Computational Biology* **2016**, *12*, e1004746.

(7) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014**, *508*, 331.

(8) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* **1998**, *102*, 3586.

(9) Mackerell, A. D.; Feig, M.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 1400.

(10) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. *Journal of Computational Chemistry* **2013**, *34*, 2757.

(11) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox *Journal* **2016**,

(12) Stote, R. H.; Karplus, M. *Proteins* **1995**, *23*, 12.

(13) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M., Jr. *Journal of Chemical Theory and Computation* **2013**, *9*, 2733.

(14) Pang, Y.-P. *Molecular Modeling Annual* **1999**, *5*, 196.

(15) Harding, M. M. *Acta Crystallographica Section D: Structural Biology* **2004**, *60*, 849.

(16) Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A. *Journal of the American Chemical Society* **2013**, *135*, 12013.

(17) Kleywegt, G. J.; Jones, T. A. *Acta Crystallograpghica D: Structural Biology* **1994**, *50*, 178.

(18) Petersen, M. T. N.; Jonson, P. H.; Petersen, S. B. *Protein Engineering* **1999**, *12*, 535.

(19) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341.

(20) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. *Current Opinion in Structural Biology* **2013**, *23*, 58.

(21) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *Journal of Computational Chemistry* **1992**, *13*, 1011.

(22) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141.

(23) Pan, A. C.; Sezer, D.; Roux, B. *The Journal of Physical Chemistry B* **2008**, *112*, 3432.

(24) Miao, Y.; Feher, V. A.; McCammon, J. A. *Journal of Chemical Theory and Computation* **2015**, *11*, 3584.

(25) Schrodinger, LLC *The PyMOL Molecular Graphics System, Version 1.3*; 2010.

(26) Autenrieth, F.; Tajkhorshid, E.; Baudry, J.; Luthey-Schulten, Z. *Journal of Computational Chemistry* **2004**, *25*, 1613.

(27) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33.

(28) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781.

(29) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.

Chapter 5: Conclusion and future directions.

5.1 Concluding remarks

Proteins are highly versatile building blocks that can be adapted to carry out an extraordinary range of functions. This complexity and diversity embodies both the potential of engineered proteins as biotechnological tools,[1] biological therapeutics,[2, 3] and in creating biomaterials.[4] Achieving these goals will undoubtedly benefit from a rich understanding of the chemical and biophysical principles that underlie the functional capabilities of proteins. To this end, creating designed protein systems that approach the complexity of their naturally occurring counterparts, both in terms of composition and

coordinated function, will further expand the protein engineering toolkit, and furnish model systems for study.

The allosteric $^{C38/C81/C96}R1_4$ system was generated by an iterative and synergistic design approach. The independently-tailorable interfaces of the protein complex can accommodate multiple distinct modalities of assembly: metal-binding groups, engineered protein-protein contacts, and redox-responsive disulfide crosslinks. Grafting these engineered features onto the cyt cb_{562} scaffold protein afforded an engineered variant, $^{C38/C81/C96}R1$, that was capable of self-assembling into a Zn-bound homotetramer, Zn-$^{C38/C81/C96}R1_4$. Removing of Zn^{II} from this templated scaffold led to the hydrolysis of a single C38-C38 disulfide crosslink, leaving five intact bonds remaining, including a second C38-C38 disulfide bond. The protein scaffold had remotely coupled the occupancy of its internalized Zn^{II} sites to a peripheral disulfide bond.

Incorporating multiple functional features in a single protein opens up the possibility of coordinated function, and properties that emerge from the collective action of multiple engineered features. Indeed, this was shown to be the case for the allosteric coupling of Zn^{II} release and C38-C38 disulfide hydrolysis in $^{C38/C81/C96}R1_4$. The modular and tailorable nature of the protein building blocks allowed us to interrogate scaffolds wherein disulfide crosslinks were selectively deleted. The observed allosteric coupling of Zn^{II} to C38-C38 hydrolysis was found to persist in Zn-$^{C38/C96}R1_4$, but this behavior was abolished in assemblies of disulfide-linked $^{C38}R1$. These observations imply that the local rearrangement that occurs about a single protein-protein interface and its embedded C38-C38 disulfide bond does not furnish sufficient driving force for bond scission. Instead,

covalent coupling of all four protomers within the tetrameric complex was necessary for Zn^{II} release to result in disulfide bond hydrolysis.

However, a purely phenomenological understanding of protein function, without any appreciation of the underlying chemistry and physics of such systems, remains a half-told story. MD simulations provide a computational tool for understanding the dynamics and energetics of a protein system in exquisite detail. Through the lens of the computational microscope, one can interrogate states of interest that cannot be accessed experimentally. In the case of $^{C38/C81/C96}R1_4$, this revealed that the structure only modestly rearranges upon Zn^{II} removal, and that additional favorable rearrangements are obstructed by the intact C38-C38 disulfide bonds. Breakage of a single C38-C38 disulfide bond permits tetramer opening that appears to be energetically driven by the relaxation of the remaining disulfide bonds, as well as favorable repacking of both *i1* interfaces within the protein. The intricate and multifarious structural and energetic changes that attend $^{C38/C81/C96}R1_4$ provide insight into the basis of C38-C38 hydrolysis, as well as the reason that a single bond is cleaved. The process is driven, and dissipates, across the tetramer as a whole. Additionally, *in silico* breakage of a C81-C81 or C96-C96 disulfide results in structural rearrangements that are attended by modest changes in free energy, revealing that the penalty associated with bond breakage is unlikely to be surmounted in these cases, an indication of why such breakage events were never observed for $^{C38/C81/C96}R1_4$ or related disulfide-linked constructs that lack C38-C38 crosslinks.[5, 6]

Collectively, these observations reveal that the interfacial fluidity of the *i1* contacts[7] is a crucial aspect of the observed allosteric behavior. These interfaces adopt

different packing configurations in the presence or absence of Zn^{II}, and this is at the heart of the Zn-responsiveness of the structure. Additionally, this responsiveness must strike a balance between specificity and fluidity. It unclear that a promiscuous set interfacial contacts, with little orientational or packing preference, could ever furnish sufficient driving force to effect structural change. Conversely, an interface that is rigidly locked into a single preferred structure would be trapped in the minimum-energy configuration, and therefore also be incapable dynamic function. Returning to the schematic example of a designed biological system presented in **Figure 1.2** provides a framework for appreciating the allosteric function on display in $^{C38/C81/C96}R1_4$. This rudimentary function encapsulates the essential properties of biological complexity. Therefore, it is hoped that this example of protein design provides some insight and guidance to future design efforts aimed at generating functional protein assemblies with functional properties that approach the level of sophistication found in Nature.

5.2 Toward a deeper understanding of $^{C38/C81/C96}R1_4$ allostery.

The allosteric behavior of $^{C38/C81/C96}R1_4$ is an intriguing phenomenon that has been probed through biochemical and biophysical methods in the present work. However, these investigations leave open productive avenues for further inquiry. Breakage of C81-C81 and C96-C96 crosslinks in $^{C38/C81/C96}R1_4$ afforded complexes that exhibited rapid structural transitions, which allowed for thermodynamic analyses of these events.

An appropriate simulation that allows for similar quantification of C38-C38 opening has remained elusive. A steered MD trajectory capturing the structural transition of interest would allow for the determination of the free energy changes that attend

structural opening following C38-C38 breakage. This analysis could be further extended toward quantification of a second, unobserved, C38-C38 disulfide bond, which can reasonably be expected to result in a $^{C81/C96}R1_4$-like structure. Simulating such structural transitions would allow for a determination and an examination of the intermediate states that are adopted. Preliminary simulations carried out in this work demonstrate that tetramer opening following C38-C38 breakage is non-linear with respect to local structural metrics. The complexity of this transition may undercut the predictive power of intuition in planning a steered MD experiment that appropriately captures the structural transition of interest. In light of this, unbiased sampling methods such as replica exchange[8] or replica exchange GaMD[9] are possible routes to effectively modeling additional structural transitions of interest in $^{C38/C81/C96}R1_4$.

The analysis of the end-states for $^{C38/C81/C96}R1_4$ opening that attends C38-C38 breakage described here discounted the possible (and likely considerable) importance of solvation effects. Hemoglobin, the classic example of an allosteric protein, undergoes structural changes upon binding O_2 that are thought to be accessible only due to the stabilizing effects of solvent molecules.[10, 11] However, access to structural snapshots that depict the structural transition of interest would permit the quantification of the overall energetic contribution of solvent on the observed structural change.[12] A more complete understanding of how the Zn^{II}-free and Zn^{II}-bound states differ would provide insight into the bulk measurements provided by ITC, which can difficult to interpret for ensembles of conformations.[13] In the case of Zn^{II} binding to $^{C38/C81/C96}R1_4$, the apparent enthalpy of binding is, counterintuitively, endothermic. A complete and detailed picture of the

structural and energetic changes that occur upon Zn^{II} binding would undoubtedly aid in resolving this puzzle.

5.3 Exploring new elements of design complexity in the R1 scaffold.

While the $^{C38/C81/C96}R1_4$ system illustrates engineering strategies by which challenges in protein design may be surmounted, there remain new opportunities for designing functional protein complexes with greater functional intricacy. Forming the $^{C38/C81/C96}R1_4$ is accomplished due to the underlying propensity of the involved R1 protomers to self-assemble, which drives the formation of the six disulfide crosslinks. Moreover, this is accomplished on top of the site-specific formation of the c-type heme linkages at Cys98 and Cys101 that occurs *in vivo*.[14] One underexplored functional role for Cys in the R1 system is that of a metal ligand. The robust templating of the R1 system should, in principle, permit the relegation of exposed Cys residues to distinct functional roles. Replacement of the Asp74 ligand with an additional cysteine residue in efficiently templating $^{C81/C96}R1_4$ scaffold gives the variant $^{C74/C81/C96}R1$, which is expected to bear a His_3Cys coordination environment similar to the engineered Type 1.5 Cu proteins.[15] Excitingly, when templated under an anaerobic atmosphere, the protein was found to preferentially for a tetramer in a Zn^{II}-templated fashion (**Figure 5.1**). This permitted isolation and subsequent crystallization of the protein (**Figure 5.2** and **Table 5.1**). The crystal structure revealed that, as designed, the cysteine residues adopted the distinct functional roles of metal-binding and disulfide bond formation. However, the templating of this sample was found to be poorly reproducible, frequently resulting in exclusive formation of aggregate, or failure to template following overnight incubation. Future

efforts involving this or other proteins should, therefore, include careful optimization of sample aeration and inclusion of chemical reductants/oxidants in the templating mixture.

The $^{C38/C81/C96}R1_4$ contains several distinct assembly features, but the achievable complexity is limited by the incorporation of a single protomer. Achieving selective heteromeric assembly in designed protein systems has been reported,[16] but remains challenging and difficult to generalize. One route to circumvent this challenge is to genetically fuse two protomers. A short peptide linker would hold two protein domains together, and would dictate the accessible relative orientations of the proteins. Constructing a homology model of the available Zn-$R1_4$ structure (PDB 3HNI) reveals that a linker of nine residues or more could connect the terminii of the proteins of a single *i1*-bridged dimer (**Figure 5.3**). However, when this construct expressed, purified, and analyzed by AUC, there was no discernable preference to adopt the desired 16-helix bundle dimeric species in the presence of Zn^{II}. Preparation of construct bearing linker lengths, including a 15-residue bridging peptide, also did not show preference toward the expected 16-helix bundle topology (**Figure 5.4**). These variants also did not afford diffractible crystals, and the observed propensity to aggregate was consistent with the formation of a heterogeneous mixture in solution.

Previous efforts to model suggest a preferred span length of ~ 10 Å,[17] which is far exceeded by the 25.7 Å separating the terminii in the Zn-$R1_4$ structure (**Figure 5.3**), even before accounting for the added distance from wrapping around the exterior of the protein surface. Future efforts in this area could employ metal-binding residues that enforce a topology more readily bridged by a peptide linker. Moreover, such an assembly need not employ the *i1* interfacial constructs of R1, which orient the terminii outward. Such a

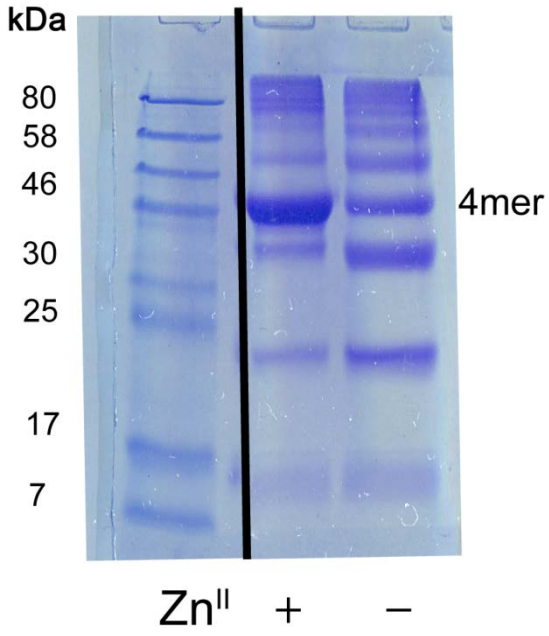

Figure 5.1 Non-reducing SDS-PAGE gels of $^{C74/C81/C96}$R1 templating reactions.
Templating reactions were prepared under an anaerobic atmosphere (Ar + 10% H_2) to contain 50 μM protein in the presence or absence of 1.5 equvalents $ZnCl_2$. Templating buffer consisted of 50 mM Tris, pH 7 and 150 mM NaCl. Samples were incubated under ambient atmosphere at 37 °C overnight.

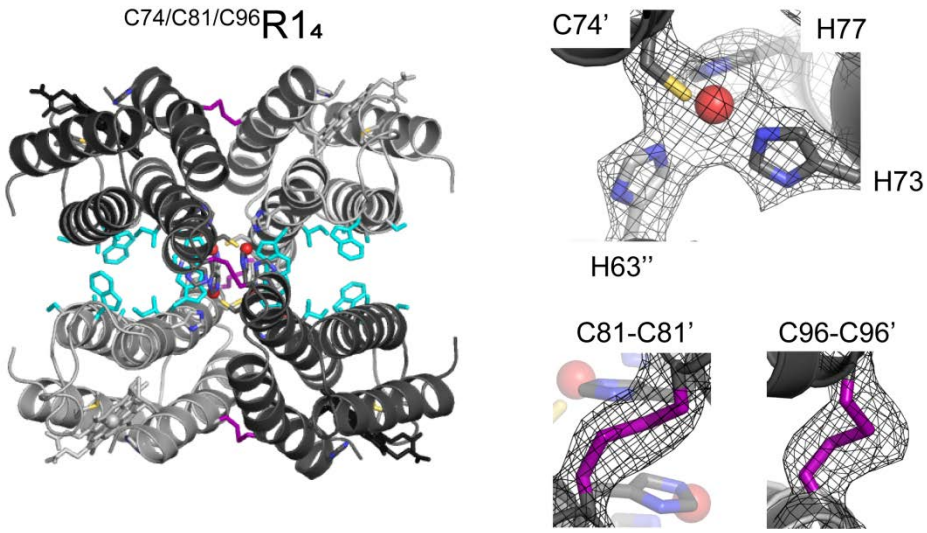

Figure 5.2 Crystal structure of $^{C74/C81/C96}R1_4$ bound to Zn^{II}.
Crystal structure of Zn-bound $^{C38/C81/C96}R1_4$ alongside close-up views of the Zn^{II} coordination sites and disulfide bonds. $2F_o$-F_c electron density maps are contoured at 1 σ (black). Metal ligands (gray), disulfide bonds (purple) and engineered *i1* contacts (cyan) are shown as sticks, while Zn^{II} ions (red) are shown as spheres.

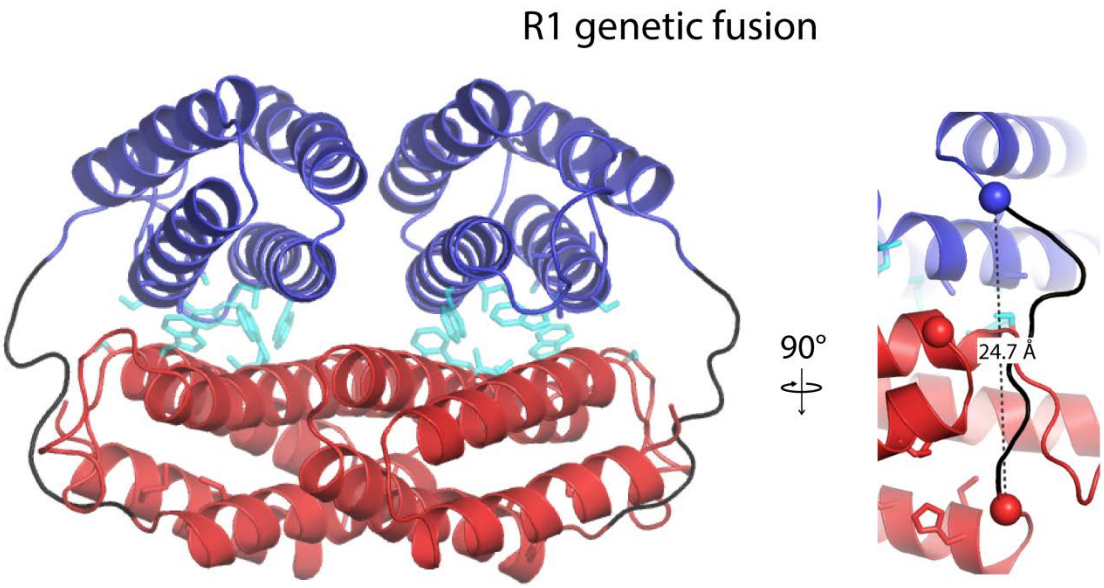

Figure 5.3 Homology model of an R1 fusion construct.
A model based on the Zn-$R1_4$ structure (PDB 3HNI) depecting the possible connection of *i1*-bridged protomers. The N-terminal domain (red) is connected to the C-terminal domain (blue) by a 9-residue linker (black). Right inset depicts the distance separating the C-α atoms of the terminal residues of each domain (analogous to residues 1 and 106 in unfused R1).

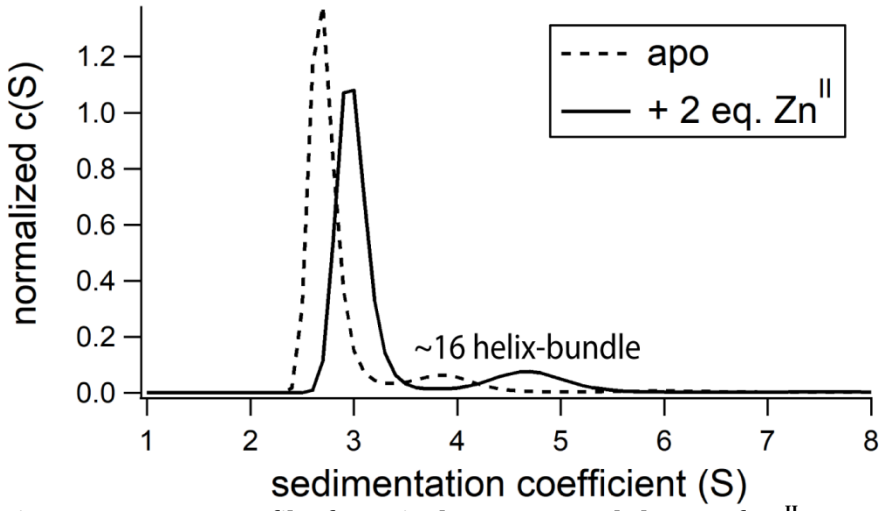

Figure 5.4 SV-AUC profile of RF_{15} in the presence and absence of Zn^{II}.

Table 5. 1 X-ray crystallography data collection and refinement statistics for $^{C74/C81/C96}$R1$_4$.
Values in parentheses denote statistics for the highest resolution shell.

‡Rsym = $\Sigma\Sigma_j|I_j - \langle I \rangle|/\Sigma\Sigma_j|Ij|$.
§R = $\Sigma||F_{obs}| - |F_{calc}||/\Sigma |F_{obs}|$.
[II]Free R calculated against 5% of the reflections removed at random for both structures.
¶Root mean square deviations from bond and angle restraints.
*Values in parentheses correspond to the highest resolution shell.

	Zn-$^{C74/81/C96}$R1$_4$
Data collection	
Space group	P 2_1
Cell dimensions	
a, *b*, *c* (Å)	49.0, 64.4, 74.9
α, β, γ (˚)	90, 106, 90
Resolution (Å)	1.7
R_{sym}‡* (%)	4.4 (49.4)
I/σ*	17.6 (2.9)
$CC_{1/2}$* (%)	99.9 (86.1)
Completeness* (%)	99.3 (98.3)
Redundancy*	3.7 (3.6)
Refinement	
No. unique reflections	35091
$R^{§}/R_{free}$ [II] (%)	24.7/19.3
R.m.s. deviations	
Bond lengths¶ (Å)	0.0211
Bond angles¶ (˚)	1.881

design strategy would afford greater control over the assembly features, and permit the facile design of scaffolds exhibiting greater complexity of composition.

5.4 Materials and Methods

Preparation and crystallization of $^{C74/C81/C96}R1_4$. The pET20b vector housing the $^{C81/C96}R1_4$ gene as used as a template for site-directed mutagenesis. The H74C mutation was installed with the mutagenic primer pair 5'-GGTCAGATTCACTGC-GCGCTGCACC-3' and 5'- GGTGCAGCGCGCAGTGAATCTGACC-3'. The sequence was confirmed, and the protein was expressed, purified, and templated as described in **2.3 Materials and Methods**. Purified $^{C74/C81/C96}R1_4$ was crystallized by sitting-drop vapor diffusion at room temperature. Crystallization screens were carried out using 500 µL reservoirs 2 µL of protein (1.5 mM monomer) and 1 µL of precipitant solution. Diffractible crystals were obtained with a precipitant solution consisting of 0.1 M Bis-Tris (pH 6.5) and 45% 2-methyl-2,4-pentanediol, and 0.1 M $NaCl_2$. Crystals were harvested, cryoprotected in perfluoropolyether cryo oil (Hampton Research), and stored in liquid nitrogen. Diffraction data was collected at 100K beamlines BL14-1 with 0.99 Å radiation at the Stanford Synchrotron Radiation Lightsource. Crystal data refinement and model building was carried out as described in **2.3 Materials and Methods.**

Preparation of R1 fusion constructs. A gene encoding an R1 fusion construct encoding two copies of the R1 gene linked head-to-tail by a nine-residue linker (RF9) was purchased from DNA 2.0 (see Appendix) in the pJ434 (low-copy) T7 expression vector featuring an AmpR marker gene. The gene of interest was flanked by *Nde*I and *EcoR*I restriction sites, and featured a *BamH*I restriction site within the linking peptide.

The codon usage of the second copy of R1 was manually altered to minimize sequence identity to the first copy in order to facilitate subsequent site-directed mutagenesis, and was therefore not codon-optimized. The purchased gene was successfully transformed into XL-1 Blue *E. coli*, but transformation into BL21(DE3) *E. coli* cells housing the ccm plasmid failed to yield viable transformants. The RF_9 gene was cloned into the pET20b vector by *Nde*I and *EcoR*I cloning, and confirmed by sequencing. The protein was expressed and purified essentially as described in Section 2.3, except that 2 mM DTT was excluded from the purification buffers due to the absence of surface-exposed cysteines. R1 fusion constructs were generated by iterative site-directed mutagenesis using the mutagenic primers indicated in Table 5.2. The resultant peptide linkers of the R1 fusion constructs are given in Table 5.3. Protein samples (2.5 μM polypeptide; 5 μM with respect to heme) in the presence or absence of four molar equivalents (with respect to polypeptide) of $ZnCl_2$ were prepared in a non-chelating buffer solution (20 mM MOPS, pH 7, and 150 mM NaCl). SV-AUC experiments were carried out as described in Section 2.3.

Table 5.2 Mutageneic primers for constructing R1 fusion genes.

Variant	Template	Mutagenic Primer (5' to 3')
RF_5	RF_7	GATCCGGTGGAGCGGCGGACC
RF_7	RF_9	CCTGCCACCAGAAGTATCGTGGATCCGGTGG
RF_9	-	-
RF_{11}	RF_9	GAAGTATCGTGGAGGTAGCGGTGGATCCGGTGGAG
RF_{13}	RF_{11}	GAAGTATCGTGGAGGTGGCGGTAGCGGTGGATCCGGTG
RF_{15}	RF_{13}	GGTGGCGGTAGCGGCGGAGGTGGATCCGGTG

Table 5.3 Linking petides encoded in R1 fusion genes.

Variant	Linking Peptide
RF_5	GSGGA
RF_7	GSGGGSA
RF_9	GGGSGGGSA
RF_{11}	GGSGGSGGGSA
RF_{13}	GGGGSGGSGGGSA
RF_{15}	GGGGSGGGGSGGGSA

References

(1) Glasscock, C. J.; Lucks, J. B.; DeLisa, M. P. *Cell Chemical Biology* **2016**, *23*, 45.

(2) Leader, B.; Baca, Q. J.; Golan, D. E. *Nature Reviews Drug Discovery* **2008**, *7*, 21.

(3) Vaughan, T.; Osbourn, J.; Jallal, B. *Protein Therapeutics*; John Wiley & Sons: Hoboken, NJ, 2007.

(4) Harden, J. L.; Li, H. *ACS Biomaterials Science & Engineering* **2017**, *3*, 658.

(5) Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A. *Journal of the American Chemical Society* **2010**, *132*, 8610–8617.

(6) Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A. *Journal of the American Chemical Society* **2013**, *135*, 12013.

(7) Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A. *Proceedings of the National Academy of Sciences U.S.A* **2010**, *107*, 1827–1832.

(8) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141.

(9) Huang, Y.-m. M.; McCammon, J. A.; Miao, Y. *Journal of Chemical Theory and Computation* **2018**,

(10) Colombo, M. F.; Rau, D. C.; Parsegian, V. A. *Science* **1992**, *256*, 655.

(11) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014**, *508*, 331.

(12) Nguyen, C. N.; Young, T. K.; Gilson, M. K. *Journal of Chemistry and Physics* **2012**, *137*, 044101.

(13) Fenley, A. T.; Muddana, H. S.; Gilson, M. K. *Proceedings of the National Academy of Sciences U.S.A* **2012**, *109*, 20006.

(14) Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R. *Biochemistry* **2006**, *45*, 10504.

(15) Kroes, S. J.; Hoitink, C. W. G.; Andrew, C. R.; Ai, J.; Sanders, L. J.; Messerschmidt, A.; Hagen, W. R.; Canters, G. W. *European Journal of Biochemistry* **2004**, *240*, 342.

(16) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R.; Yi, S.; Davis, T. N.; Gonen, T.; King, N. P.; Baker, D. *Nature* **2016**, *535*, 136.

(17) Choi, Y.; Agarwal, S.; Deane, C. M. *PeerJ* **2013**, *1*, e1.

Appendix

Gene encoding RF_9:

ATGTTACGCACTGTGATCGTTGCCGGAGCCCTGGTGTTGACGGCGAGCGCCG
TGATGGCGGCTGATCTTGAAGACAATATGGAAACCCTCAACGACAATTTAAA
AGTGATCGAAAAAGCGGATAACGCGGCGCAAGTCAAAGACGCGTTAACGAA
GATGGCGGCCGCAGCGGCGGATGCGTGGAGCGCAACGCCGCCGAAGCTCGA
AGATAAATCACCGGACAGCCCGGAAATGCACGATTTCCGCCACGGTTTCTGG
ATTCTGATTGGTCAGATTCACGACGCGCTGCACCTGGCAAATGAAGGTAAAG
TAAAAGAAGCGCAGGCTGCTGCAGAGCAACTGAAAACGACCTGCAACGCCT
GCCACCAGAAGTATCGTGGAGGTGGATCCGGTGGAGGTTCGGCGGCGGACCT
GGAGGACAACATGGAGACACTGAATGATAACCTGAAGGTTATTGAGAAGGC
CGACAATGCCGCCCAGGTGAAGGATGCCCTGACCAAAATGGCCGCGGCGGC
AGCGGACGCCTGGTCTGCGACCCCACCGAAATTAGAAGACAAAAGCCCGGA
TTCTCCGGAGATGCATGACTTTCGTCATGGCTTTTGGATTCTGATCGGCCAGA
TCCATGATGCCCTGCATCTGGCGAACGAGGGCAAAGTGAAAGAGGCCCAAG
CGGCCGCGGAACAGCTGAAGACAACATGTAATGCGTGTCATCAGAAATACCG
TTAA