Title
Metal-directed protein self-assembly

Author
Salgado. Eric N., Salgado. Eric N.

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Metal-Directed Protein Self-Assembly

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

Doctor of Philosophy

in

Chemistry

by

Eric N. Salgado

Committee in charge:
    Professor F. Akif Tezcan, Chair
    Professor Seth Cohen
    Professor Gourisankar Ghosh
    Professor Elizabeth Komives
    Professor Paul A. Price

2010

The Dissertation of Eric N. Salgado is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

Epigraph

"Science, like Nature, must also be tamed with a view towards its preservation."

- From *Natural Science* by Rush

Table of Contents

List of Abbreviations

AUC	Analytical Ultracentrifugation
Bis-His	Bis-Histidine clamp
BMB	Bis(maleimido)butane
BMH	Bis(maleimido)hexane
BMOE	Bis(maleimido)ethane
^{C82}RIDC-1	Gly82Cys variant of RIDC-1
^{C82}RIDC-1$_{BMB}$	BMB crosslinked ^{C82}RIDC-1
^{C82}RIDC-1$_{BMH}$	BMH crosslinked ^{C82}RIDC-1
^{C82}RIDC-1$_{BMOE}$	BMOE crosslinked ^{C82}RIDC-1
SO_4:^{C82}RIDC-1$_{2,BMOE}$	Sulfate-mediated ^{C82}RIDC-1$_{2,BMOE}$ tetramer
CD	Circular Dichroism
CPI	Crystal Packing Interactions
Cu_2:MBPC-1_2	Cu-mediated MBPC-1 dimer
Cu_2:RIDC-1_2	Cu-mediated RIDC-1 dimer
cyt cb_{562}	Cytochrome cb_{562}
cyt b_{562}	Cytochrome b_{562}
DLS	Dynamic Light Scattering
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ΔG_f	Equilibrium Free Energy of Folding
Guanidine hydrochloride	GuHCl
K_d	Equilibrium Dissociation Constant

MALDI	Matrix-Assisted Laser-Desorption Ionization
MBPC-1	Metal-Binding Cytochrome 1
MBPC-2	Metal-Binding-Cytochrome 2
MDPSA	Metal-Dependent Protein Self-Assembly
MeTIR	Metal-Templated Interface Redesign
MOPS	3-(N-morpholino)propanesulfonic acid
Ni_2:MBPC-1_3	Ni-mediated MBPC-1 trimer
PAGE	Poly-Acrylamide Gel Electrophoresis
Pd_2:MBPC-1_2	Pd-mediated MBPC-1 dimer
PEG	Polyethylene Glycol
PISA	Protein Interactions, Surfaces and Assemblies
PPI	Protein-Protein interaction
RIDC-1	Rosetta Interface Design Cytochrome 1
RIDC-1_2	Metal-Independent Dimer of RIDC-1
RIDC-2	Rosetta Interface Design Cytochrome 2
SDS	Sodium Dodecyl Sulfate
SE	Sedimentation Equilibrium
SV	Sedimentation Velocity
TRIS	Tris(hydroxymethyl)aminomethane
Zn_4:MBPC-1_4	Zn-mediated MBPC-1 tetramer
Zn_4:MBPC-2_4	Zn-mediated MBPC-2 tetramer
Zn_4:RIDC-1_4	Zn-mediated RIDC-1 tetramer
Zn_4:RIDC-2_4	Zn-mediated RIDC-2 tetramer

Zn$_4$: ^{C82}RIDC-1$_{2,BMB}$	Zn-mediated ^{C82}RIDC-1$_{2,BMB}$ tetramer
Zn$_4$: ^{C82}RIDC-1$_{2,BMH}$	Zn-mediated ^{C82}RIDC-1$_{2,BMH}$ tetramer
Zn$_4$:^{C82}RIDC-1$_{2,BMOE}$	Zn-mediated ^{C82}RIDC-1$_{2,BMOE}$ tetramer

List of Figures

List of Tables

Acknowledgements

First and foremost I must thank my P.I., Akif Tezcan for not only his intelligent and invaluable guidance through this project, but also for putting up with me for half a decade. This last sentiment also goes out to all of the members of our quickly expanding lab who have been invaluable for their thoughtful discussions and jovial spirits. Special thanks goes to Richard Lewis for not only keeping me entertained, but also starting work on the MeTIR derived proteins that have been so fruitful in the lab. Robert Radford was also instrumental in helping me with a detour in to synthetic organic chemistry that, although ultimately doomed, was worth trying.

I must also express my gratitude to Prof. Elizabeth Komives and her entire lab for help with AUC and ITC, allowing me to barrow various materials, and for inviting me to join in lab festivities. Of course these invitations would not have been extended to me without my beautiful girlfriend, Ingrid DeVries, who has put up with me for almost four years now, and for that I am truly indebted to her and owe her my love.

Without my family I would truly be lost, and so my love also goes out to my parents, Elias and Deborah Salgado, my sister, Candace, and the best niece in the world, Riley Paige, for supporting me in everything I've wanted to do, while at the same time keeping me grounded.

This dissertation was funded by both the UCSD Molecular Biophysics Training Program, as well a NIH Ruth L. Kirschstein National Research Service Award.

Chapters 1 and 7 are reproduced in part with permission from: Salgado, E. N., Radford, R. J., Tezcan, F. A. **2010**. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res*. 43, 661-672. Copyright 2010 American Chemical Society.

Chapter 2 is reproduced in part with permission from: Salgado, E. N., Faraone-Mennella, J., Tezcan, F. A. **2007**. Controlling Protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* 129, 13374-13375. Copyright 2007 American Chemical Society.

Chapter 3 is reproduced in part with permission from: Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., Tezcan, F. A. **2009**. Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical assemblies through Cu^{II} and Ni^{II} coordination. *Inorg. Chem*. 48, 2726-2728. Copyright 2009 American Chemical Society.

Chapter 4 is reproduced in part with permission from: Salgado, E. N., Lewis, R. A., Faraone-Mennella, J., Tezcan, F. A. **2008**. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* 130, 6082–6084. Copyright 2008 American Chemical Society.

Chapter 5 is reproduced in part from: Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B., Tezcan, F. A. **2010**. Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. USA*. 107, 1827-1832.

The dissertation author is primary author on all reprinted materials.

Vita

Education

2004	B.S. in Biochemistry, University of Massachusetts, Amherst
2007	M.S. in Chemistry, University of California, San Diego
2010	Ph.D. in Chemistry, University of California, San Diego

Awards and Honors

2009-2010	Ruth L. Kirschstein National Research Service Award Predoctoral Fellowship
2008	Molecular Biophysics Training Program Retreat Poster Award
2006-2009	Molecular Biophysics Training Program Fellowship

Publications

1. Radford, R. J., Brodin, J. D., Salgado, E. N., Tezcan, F. A. **2010**. Expanding the Utility of Proteins as Platforms for Coordination Chemistry. *Coord. Chem. Rev.* doi:10.1016/j.ccr.2010.10.010.

2. Brodin, J. D., Medina-Morales, A., Ni, T., Salgado, E. N., Ambroggio, X. I., Tezcan, F. A. **2010**. Evolution of metal selectivity in templated protein interfaces. *J. Am. Chem. Soc.* 132, 8610–8617.

3. Salgado, E. N., Radford, R. J., Tezcan, F. A. **2010**. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* 43, 661-672.

4. Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B., Tezcan, F. A. **2010**. Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. USA.* 107, 1827-1832.

5. Guce, A. I., Clark, N. E., Salgado, E. N., Ivanen, D. R., Kulminskaya, A. A., Brumer III, H., Garman, S. C. **2010**. Catalytic mechanism of human α-galactosidase. *J. Biol. Chem.* 285, 3625-3632.

6. Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., Tezcan, F. A. **2009**. Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical assemblies through Cu^{II} and Ni^{II} coordination. *Inorg. Chem.* 48, 2726-2728.

7. Salgado, E. N., Lewis, R. A., Faraone-Mennella, J., Tezcan, F. A. **2008**. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* 130, 6082–6084.

8. Salgado, E. N., Faraone-Mennella, J., Tezcan, F. A. **2007**. Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* 129, 13374-13375.

Fields of Study

Major Field: Biochemistry

Biochemistry: Bioinorganic Chemistry, Structural Biology, and Biophysics.

ABSTRACT OF THE DISSERTATION

Metal-Directed Protein Self-Assembly

by

Eric N. Salgado

Doctor of Philosophy in Chemistry

University of California, San Diego, 2010

Professor F. Akif Tezcan, Chair

The *de novo* design of protein-protein interactions (PPIs) has proven to be an immense challenge due to the difficulty in controlling and predicting the weak forces that govern them. In order to circumvent this challenge, we set out to induce new PPIs between monomeric proteins through the coordination of metal ions. In our strategy, the strength of the metal bonding interactions should be sufficient to drive the formation of PPIs without initial consideration of the weak non-covalent interactions along the newly formed interfaces. Additionally, the distinct geometric preferences of metal ions should only allow for a limited number of organizations of proteins around a given metal center, thus permitting a degree of foresight into the oligomeric architectures that will be formed.

As a model system, we chose a stable, four-helix bundle protein, cytochrome *cb*$_{562}$, onto which two *i* and *i+4* bis-histidine "clamp" coordination motifs were installed. The resulting metal-binding construct, MBPC-1, is shown to be capable of forming discrete oligomeric species whose supramolecular architectures depend on the stereochemical preferences of the added metal.

Further studies to exert control over the morphology of a Zn-mediated tetrameric assembly of MBPC-1 (Zn_4:MBPC-1_4) highlighted the fact that, although we initially ignored secondary, non-covalent interactions along our newly induced interfaces, these interactions do play a role in determining the overall structure of the complex. This idea led us to computationally "evolve" our protein by re-designing residues along interfaces of the Zn_4:MBPC-1_4 tetramer to include more favorable hydrophobic packing interactions. The resulting constructs were shown to not only form a significantly stabilized tetramer, but also to have the ability to form oligomers in the absence of metals. Further stabilization of this tetrameric architecture by way of chemical crosslinking has allowed for the formation of a potentially metal-selective complex that can serve as a stable platform onto which functionalized metal centers, and in turn enzymatic activity, may be engineered in the future.

Chapter 1

Introduction

Due in part to the diversity of functional groups found within the canonical twenty amino acids, proteins have proven to be the most versatile of biological building blocks. A large fraction of proteins exist in a multimeric form, which affords increased stability over the monomeric species, enables allostery as well as the construction of interfacial active sites, and provides genetic coding efficiency.[1] The known multimeric protein complexes display a wide range of internal symmetries that allow the construction of large assemblies out of simple monomers (Figure 1.1). Despite the presence of large, high-symmetry structures as maxi-ferritin,[2] viral capsids,[3] and bacterial flagella,[4] the most prevalent protein oligomers are in fact smaller C_2 dimers and D_2 tetramers.[5]

As any given cell contains thousands of different proteins, it is highly important that the protein-protein interactions (PPIs) that guide the formation of oligomeric complexes be specific between protein partners in order to maintain properly operating cellular functions. Two factors that contribute to this specificity are: 1) shape complementarity between large (typically ≥ 1000 Å^2) surfaces of interacting proteins; and 2) a finely tuned network of favorable interactions along protein interfaces.[6] Although the burial of hydrophobic residues is the major driving force for PPIs, protein interfaces often contain a relatively large number of polar/charged interactions, especially compared to the interior of a folded protein.[7, 8]

The *de novo* design of non-natural proteins and protein assemblies is a highly sought-after goal, not only due to the better fundamental understanding of protein structure and function that this pursuit would bring, but also due to the fact that it could lead to the creation of novel protein-based materials. Along these lines, a combination of experimental and computational approaches have been employed toward engineering

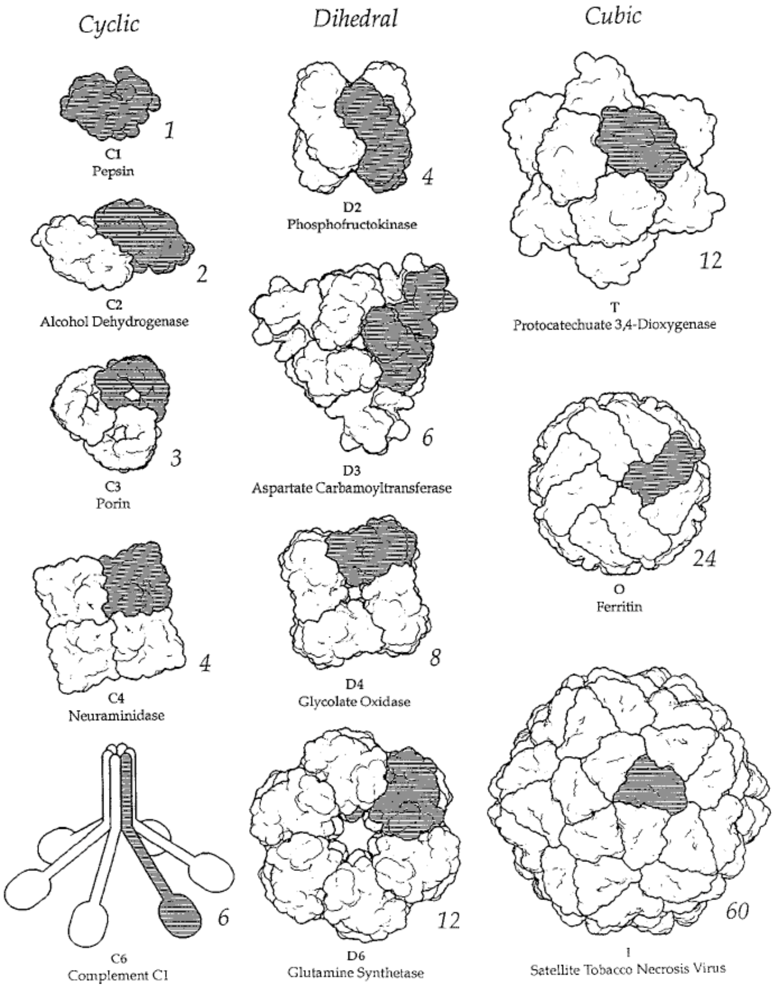


Figure 1.1. Point group symmetries in various multimeric protein complexes. Point group symbols are included below each protein structure *(e.g., C1* and *D2*), and the number of monomeric subunits in each group is included below and to the right of the structure *(e.g., 24* in octahedral group *O*). One monomeric subunit is shaded in each example. Adapted from reference 1.

four-helix bundle assemblies,[9, 10] novel proteins[1] and protein active sites,[12] and *de novo* PPIs.[13] Despite these significant advances, there still exist a number of limitations in designing proteins and PPIs. One major problem lies in the inability to accurately predict the effects of water-based hydrogen bonding networks known to exist within naturally occurring PPI interfaces.[14] Similarly, the previously mentioned presence of charged residues along these interfaces serve as another roadblock. The sheer number of individual charged amino acid side chains throughout even a moderately sized protein makes necessary the explicit modeling of all possible electrostatic interactions between neighboring residues to ensure accuracy of the calculation. Unfortunately, at the present, this remains computationally infeasible.[15] Adding to these complications is the presence of dynamic protein motions that cannot always be sufficiently modeled, which poses a challenge when trying to predict surface shape complementarity.

A large fraction of natural proteins (30-50%) contain metal ions, which are central to their structure and function. Not surprisingly, in many of the well-characterized *designed* protein systems, metal ions or cofactors have played a central role in providing a stable fold and/or acting as functional sites.[16, 17] Possibly the most ubiquitous examples for a structural metal site are zinc finger domains such as those found in Zif268.[18] These domains generally employ a His_2Cys_2 coordination motif to bind a Zn^{2+} ion (Figure 1.2 A) that serves to stabilize a $\beta\beta\alpha$ tertiary fold capable of binding DNA in a sequence-specific manner. The utility of this domain, in fact, has been exploited for the design of numerous engineered DNA binding proteins.[19]

Besides imparting structural integrity, metal centers have been adopted through evolution to serve a multitude of functions including oxygen transport, redox chemistry, and dinitrogen reduction. A wonderfully intricate use of metal centers can be seen within Photosystem II, the first protein complex of the light-dependent pathway in photosynthesis. Here, an elaborate pathway of Mg- and Fe-porphyrin cofactors, a Mn_4Ca cluster, and an iron center (Figure 1.2 B)[20] allow for the conversion of light energy into redox equivalents, their use in the evolution of dioxygen through the catalytic splitting of water, formation of a proton gradient required for production of ATP, and the transport of electrons necessary for subsequent steps in photosynthesis. As a watermark in protein design, engineering of a stable, water-soluble analogue of this complex would not only prove a significant achievement, but also have tremendous implications to the field of renewable energy.

While the study and design of metalloproteins and their functions has thrived, parallel work employing small molecules, seeking to understand and mimic these functions and developing new activities, has rapidly moved forward. Synthesis of metal-organic complexes has the advantages of an expanded repertoire of functional groups not contained within the canonical amino acids. This, coupled with a generally intuitive design strategy based on the coordination of metal ions by ligands that are, compared to the complexity of protein structures, relatively simple, allows for the rational design of supramolecular structures capable of self-assembly. Such work has evolved into the burgeoning fields of supramolecular coordination chemistry and metal-organic frameworks (MOFs), which offer potential applications ranging from gas storage and separation to heterogeneous catalysis (Figure 1.3 A),[21-24] and have allowed the design of closed, self-assembling structures whose sizes can be modulated, through application of chemical intuition, to the magnitude of small viral particles (Figure 1.3 B).[25, 27]

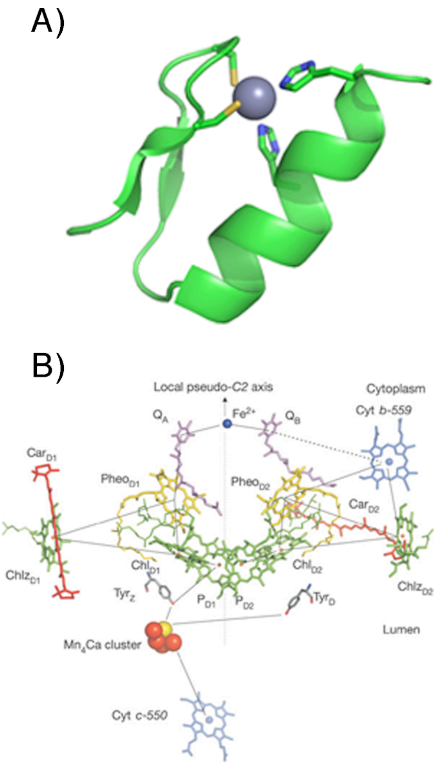


Figure 1.2. Diversity of metal ions in protein scaffolds. A) Structure of a representative zinc finger domain from the Zif286 DNA binding protein displaying the His_2Cys_2 structural Zn^{2+} binding site (represented in sticks and spheres, respectively. Adapted from PDB ID 1ZAA[21]). B) Redox active cofactors and metal centers involved in the electron transport chain of photosystem II, adapted from reference 23.

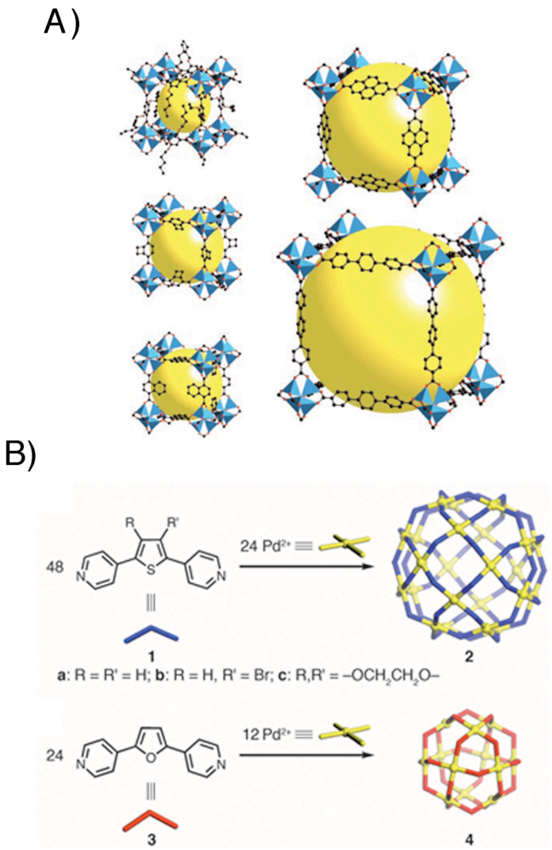


Figure 1.3. Metal-directed assembly of small molecule complexes. A) Single crystal x-ray structures of various IRMOF compounds. The large yellow spheres represent the largest van der Waals spheres that would fit in the cavities without touching the various frameworks (adapted from reference 30). B) Self-assembly of $M_{24}L_{48}$ spheres **2** and $M_{12}L_{24}$ sphere **4** driven by the square-planar coordination of Pd^{2+} template ions (adapted from reference 27).

Goals of Dissertation

In our work, we have sought to combine the seemingly disparate fields of metalloprotein design with metal-directed assembly to develop a methodology for the formation of novel PPIs through metal coordination. This strategy, Metal-Directed Protein Self-Assembly (MDPSA), offers several advantages. First, metal-ligand bonds are stronger than the non-covalent bonds that make up protein interfaces, obviating the need to engineer large surfaces to produce favorable protein-protein docking. Second, metal-ligand bonds are highly directional, and thus, the stereochemical preference and symmetry of metal coordination may be imposed onto PPIs. Third, metal-ligand bonds can be kinetically labile, allowing PPIs to proceed under thermodynamic, rather than kinetic control. Fourth, metal coordination can be formed or broken through pH changes or external ligands, rendering PPIs responsive to external stimuli. And fifth, metal ions bring along intrinsic reactivity (Lewis acidity, redox reactivity), which may be incorporated into protein interfaces.

In order to engineer functional metal coordination sites within MDPSA derived interfaces, we found it useful to think about how functional metal centers and the surrounding protein environments may have evolved in natural systems. It is probable that some metals were incorporated into pre-existing protein scaffolds that presented the right coordination spheres, which subsequently were optimized for metal-based functions through cycles of natural selection. It is also quite likely that, early in the evolution of proteins, metals could initially have nucleated the assembly of random peptides or proteins around them, followed by rigidification and optimization of the surrounding peptide chain(s) for metal reactivities beneficial to a particular organism. With this hypothetical evolutionary time course in mind, we developed a rational engineering approach, metal-templated interface redesign (MeTIR). MeTIR involves the computational re-design of PPI interfaces with the intended goal of introducing energetically more favorable interactions that can stabilize metal-mediated oligomers. An additional, evolution-inspired step involving the formation of chemically crosslinked two-domain protein constructs ultimately results in a designed protein system that, while not demonstrating the ability to self-assemble in the absence of any templating metal, does appear to have imparted upon it a degree of metal binding specificity. This work, while uncovering new obstacles that must be overcome to make MDPSA and MeTIR more effective, nonetheless sets the stage for functionalization of future designed metalloenzymes.

Chapter 1 is reproduced in part with permission from: Salgado, E. N., Radford, R. J., Tezcan, F. A. **2010**. Metal-directed protein self-assembly. *Acc. Chem. Res*. 43, 661-672. Copyright 2010 American Chemical Society.

References

1. Goodsell, D. S., Olson, A. J. **2000**. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* 29,

2. Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., Thomas, C. D., Shaw, W.

V., Harrison, P. M. **1991**. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature*. 349, 541-544.

3. Johnson, J. E., Speir, J. A. **1997**. Quasi-equivalent viruses: A paradigm for protein assemblies. *J. Mol. Biol.* 269, 665-675.

4. Afzelius, B. A., Dallai, R., Lanzavecchia, S., Bellon, P. L. **1995**. Flagellar structure in normal human spermatozoa and in spermatozoa that lack dynein arms. *Tissue & Cell*. 27, 241-247.

5. Levy, E. D., Erba, E. B., Robinson, C. V., Teichmann, S. A. **2008**. Assembly reflects evolution of protein complexes. *Nature*. 453, 1262-1265.

6. Jones, S., Thornton, J. M. **1996**. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA*. 93, 13-20.

7. Larsen, T. A., Olson, A. J., Goodsell, D. S. **1998**. Morphology of protein–protein interfaces. *Structure*. 6, 421-427.

8. Bogan, A. A., Thorn, K. S. **1998**. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 1-9.

9. Cochran, F. V., Wu, S. P., Wang, W., Nanda, V., Saven, J. G., Therien, M. J., DeGrado, W. F. **2005**. Computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. *J. Am. Chem. Soc.* 127, 1346-1347.

10. Hecht, M. H., Richardson, J. S., Richardson, D. C., Ogden, R. C. **1990**. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science*. 249, 884-891.

11. Lu, Y., Berry, S. M., Pfister, T. D. **2001**. Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds. *Chem. Rev.* 101, 3047-3080.

12. Dahiyat, B. I., A Mayo, S. L. **1997**. De novo protein design: fully automated sequence selection. *Science*. 278, 82-87.

13. Havranek, J. J., Harbury, P. B. **2003**. Automated design of specificity in molecular recognition. *Nat. Struct. Mol. Biol.* 10, 45-52.

14. Janin, J. **1999**. Wet and dry interfaces: The role of solvent in protein-protein and protein-DNA recognition. *Structure*. 7, R277-R279.

15. Morozov, A. V., Kortemme, T., Baker, D. **2003**. Evaluation of models of electrostatic interactions in proteins. *J. Phys. Chem. B*. 107, 2075-2090.

16. Kaplan, J., DeGrado, W.F. **2004**. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11566–11570.

17. Shifman, J. M., Moser, C., C., Kalsbeck, W. A., Bocian, D. F., Dutton, P. L. **1998**. Functionalized de novo designed proteins: mechanism of proton coupling to oxidation/reduction in heme protein maquettes. 37, 16815-16827.

18. Pavletich, N. P., Pabo, C. O. **1991**. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*. 10, 809-817.

19. Choob, Y., Isalan, M. **2000**. Advances in zinc finger engineering. *Curr. Opin. Struct. Biol.* 10, 411-416.

20. Loll, B., Kern, J., Saenger, W., Zouni, A., Biesiadka, J. **2005**. Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature*. 438, 1040-1044.

21. Rowsell, J. L. C., Millward, A. R., Park, K. S., Yaghi, O. M. **2004**. Hydrogen sorption in functionalized Metal−Organic Frameworks. *J. Am. Chem. Soc.* 126, 5666-5667.

22. James, S. L. **2003**. Metal-organic frameworks. *Chem. Soc. Rev.* 32, 276–288.

23. Millward, A. R., Yaghi, O. M. **2005**. Metal−Organic Frameworks with exceptionally high capacity for storage of carbon dioxide at room temperature. 127, 17998-17999.

24. Tanabe, K. K., Wang, Z., Cohen, S. M. **2008**. Systematic functionalization of a Metal−Organic Framework via a postsynthetic modification approach. *J. Am. Chem. Soc.* 130, 8508-8517.

25. Tominaga, M., Suzuki, K., Kawano, M., Kusukawa, T., Ozeki, T., Sakamoto, S., Yamaguchi, K., Fujita, M. **2004**. Finite, Spherical Coordination Networks that Self-Organize from 36 Small Components. *Angew. Chem. Int. Ed. Engl.* 43, 5621-5625.

26. Sun, Q.-F., Iwasa, J., Ogawa, D., Ishido, Y., Sato, S., Ozeki, T., Sei, Y., Yamaguchi, K., Fujita, M. **2010**. Self-Assembled M24L48 Polyhedra and Their Sharp Structural Switch upon Subtle Ligand Variation. *Science*. 328, 1144-1147.

27. Eddaoudi, M., Kim, J., Rosi, N., Vodak, D., Wachter, J., O'Keeffe, M., Yaghi, O. M. **2002**. Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science*. 295, 469-472.

Chapter 2

First Test of MDPSA: Generation of MBPC-1 and Characterization of its Zn-Mediated Self-Assembly

Introduction

For our initial studies to probe the feasibility and scope of Metal-Directed Protein Self-Assembly (MDPSA), we looked to employ a relatively rigid protein building block of uniform shape and size. Needless to say, it is also important that the protein in question be monomeric so as to not predispose the formation of oligomers of any kind. Towards this end, we chose to study cytochrome cb_{562} (cyt cb_{562}), a variant of the four-helix bundle protein, cytochrome b_{562} (cyt b_{562}).[1] In addition to its cylindrical shape, and the fact that it is monomeric up to millimolar concentrations, there are several factors that make cyt cb_{562} a good model system for investigating metal-mediated PPIs. For one, its helical-bundle fold is highly stable, a feature that is further strengthened through the engineering of covalent heme-polypeptide linkages into the parent cyt b_{562}. As a result, its structure is not perturbed by modifications on its surface. Secondly, the all-α-helical makeup of the protein leads to uniform surface features that facilitate the introduction of metal-binding motifs. Finally, it is very easy to express and purify in large quantities in the reasonably short period of a few days.

Even when proteins possess a simple shape and a uniform topology like cyt cb_{562}, it can be challenging to visualize protein-protein docking geometries that would be amenable to metal-mediated crosslinking: surface composition and topography of proteins are highly heterogeneous. For inspiration, we turned to crystal packing interactions (CPIs), which provide a source of feasible protein-protein docking geometries. A given CPI typically is not extensive (<1000 Å^2) and only stable in combination with other CPIs under crystallization conditions, yet it provides a metastable arrangement of proteins in which there are no steric clashes between them. Importantly, CPIs may contain twofold or higher symmetry, which minimizes the number of metal coordination motifs that need to be engineered onto protein surfaces and provides a starting point to design high-order protein oligomers and superstructures. An examination of CPIs in cyt cb_{562} crystals reveals that each of the four monomers in the asymmetric unit are paired in an anti-parallel fashion with another monomer along their third helices (Figure 2.1 A). This relatively large (775 Å^2), C_2 symmetrical interface is clearly non-physiological, since cyt cb_{562} is monomeric in solution.

At the same time, this close-packed arrangement of monomers shows us a route for the self-assembly of cyt cb_{562}, whereby two metal binding motifs can be incorporated near each end of Helix3 for metal-mediated crosslinking. To address the challenge of selective metal localization, we have exploited the simple principle of the chelate effect. All metal coordinating groups on a protein's surface are either formally or effectively monodentate and are considered as weak ligands at neutral pH except histidine (His) and cysteine (Cys). Therefore, if a bi- or tridentate metal chelating motif can be installed on a protein surface, it should theoretically out-compete other functionalities for metal binding while leaving coordination sites free to accommodate other protein monomers. The *i*, *i*+4 bis-His motif on α-helices, in particular, is a high-affinity bidentate motif frequently utilized for the assembly of natural and engineered metalloproteins.[2] We thus constructed a model system, Metal Binding Protein Cytochrome 1 (MBPC-1) which is a cyt cb_{562} variant containing two such bis-His motifs (His59/His63 and His73/His77) near the ends of Helix3 (Figure 2.1 B). Initial studies probed the ability of MBPC-1 to form Zn^{2+}-dependent oligomers as the lability of Zn likely permits the exploration of different

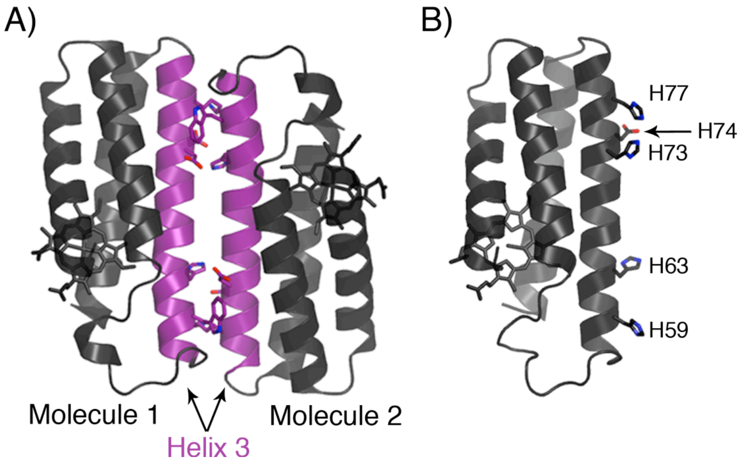


Figure 2.1. CPI interface of cyt *cb*$_{562}$ inspires design of MBPC-1. A) Anti-parallel arrangement of cyt *cb*$_{562}$ molecules in the crystal lattice along their Helix3′s (magenta). B) Model of MBPC-1, where key residues involved in metal binding are depicted as sticks.

coordination geometries, resulting in the formation of the thermodynamically most stable quaternary structure. These experiments, in both the solution and crystalline states, revealed this stable form to be a Zn-mediated MBPC-1 tetramer, Zn_4:MBPC-1_4.[3]

Materials and Methods

Site-Directed Mutagenesis and Protein Expression/Purification/Characterization. Site-directed mutagenesis was performed on the pETc-b562 plasmid (denoted as wildtype), constructed as previously described[4] by ligation of the N-terminal leader sequence of *Rhodopseudomonas palustris* cyt *c*556[5] in-frame to the gene encoding *E. coli* cyt b_{562} (pNS207b562),[6] the latter containing the R98C/Y101C mutations necessary for *c*-type linkage of the heme group to he polypeptide chain. Mutagenesis was performed using the QuikChange kit (Stratagene) and employing primers obtained from Integrated DNA Technologies. The mutant plasmids were then transformed into XL-1 blue *E. coli* cells and purified using the QIAprep Spin Miniprep kit (Qiagen). Sequencing of the mutant plasmids was performed by UCSD Moores Cancer Center. Mutant plasmids were transformed into BL21(DE3) *E. coli* cells along with the *ccm* heme maturation gene cassette plasmid, pEC86. Cells were plated on LB agar, containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol, and grown overnight.

5 ml of LB medium was then inoculated from these colonies, 125 µl of which was used to inoculate 1 L cultures after reaching an O.D. of 0.6. 1 L cultures were allowed to incubate for 16 hours at 37°C, with rotary shaking at 250 rpm with induction not being necessary. Mutant-expressing cells were then sonicated for 10 min (30 sec pulse at 45%

amplitude with 1 min rest between pulses), brought to pH 5 with the addition of HCl, and centrifuged at 16,000 g, 4^{o}C, for 1 hr. The protein was then purified by ion-exchange chromatography on a CM-Sepharose matrix (Amersham Biosciences) using a NaCl gradient in 5 mM sodium acetate, pH 5. After exchange into 10 mM sodium phosphate, pH 8, the protein was further purified using an Uno-Q (BioRad) anion exchange column on a DuoFlow chromatography workstation (BioRad) using a NaCl gradient. Purity was determined by SDS-PAGE gel electrophoresis. Verification of mutations was made through MALDI mass spectrometry.

Guanidine hydrochloride denaturation. In order to determine the stability of MPBC-1, and whether or not coordination of Zn^{2+} increased its stability, guanidine hydrochloride (GuHCl) denaturation was employed. All samples contained 5 μM protein with varying concentrations of GuHCl in 20 mM sodium acetate (pH 5), or 20 mM TRIS (pH 7). Samples at pH 7 were analyzed in the presence or absence of 300 μM $ZnCl_2$. The unfolding transition was monitored by UV-Vis spectroscopy at 415 nm on an HP 8452A diode array spectrophotometer. The refractive indices of the solutions were determined at the sodium D line with a Bausch and Lomb Abbe-3L refractometer. GuHCl concentrations were calculated from these refractive indices using the equation:[7]

$$[\text{GuHCl}] = 57.147(\Delta N) + 38.68(\Delta N)^2 - 91.60(\Delta N)^3$$

where ΔN is the difference in refractive index of the sample containing GuHCl and buffer without added denaturant.

Dynamic Light Scattering (DLS). DLS measurements and analyses were made on a DynaPro (Wyatt Technologies) instrument. All MBPC-1 samples were analyzed at a concentration of 649 μM in 20 mM TRIS (pH 7) with 0 μM, 325 μM (2:1 protein:Zn^{2+}), 649 μM (1:1 protein:Zn^{2+}), or 1.5 mM (1:2 protein:Zn^{2+}) $ZnCl_2$. The results are summarized in Table 2.2. It must be noted that oligomers may not be directly inferred from DLS data as molecular weight and hydrodynamic radii were calculated using a mono-modal spherical model.

Sedimentation Velocity (SV). All SV experiments containing Zn^{2+} were performed in 20 mM TRIS (pH 7) using appropriate volumes of $ZnCl_2$ stock solution (5 mM in 20 mM TRIS, pH 7) to yield a 1:1 Zn^{2+}:protein concentration ratio (5, 50, 200, and 600 μM). Metal-free samples contained 1 mM EDTA in 20 mM TRIS (pH 7). Measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor at 47,000 rpm for a total of 179 scans per sample. The following wavelengths were used for detection: 570 nm (50 μM protein), 582 nm (200 μM protein), 620 nm (400 μM protein), and 670 nm (600 μM protein).

All data were processed using a continuous c(s) model in the program SEDFIT.[8] Buffer viscosity and density were calculated at 25°C with SEDNTERP (http://www.jphilo.mailway.com). Vbar was calculated as in SEDNTERP:

$$\mathrm{Vbar} = \Sigma(g_i \times \mathrm{Vbar}_{i.}) / \Sigma(g_i)$$

where g_i is the number of grams of the *i*th amino acid residue and Vbar_i is the partial specific volume of the residue. The heme was added to this calculation assuming a partial specific volume of 0.82 mg/ml and a mass of 619 g/mol. Data were processed by fixing the following parameters: Vbar = 0.7316 ml/g; buffer density (ρ) = 0.99764 g/ml; buffer viscosity = 0.0089485 poise; bottom of the cell; and sample meniscus. Optimization of a baseline correction constant over all scans was performed using the "Run" command in SEDFIT with a confidence level of 0.95. This was then followed by similar optimization of time invariant (TI) and radial invariant (RI) systematic noise. The weight averaged frictional ratio (f/f_o) of the sample was then calculated using the Fit command with a confidence level of 0. A final distribution was obtained after increasing the confidence level back to 0.95 and executing the "Run" command.

Nuclear Magnetic Resonance (NMR) Spectroscopy. NMR samples contained 1.35 mM MBPC-1 in 5 mM TRIS (pH 7), with or without 1.35 mM Zn^{2+}, 90% H_2O/10% D_2O. Experiments were carried out at 293 K on a Bruker DMX 500 MHz spectrometer, delivering a field gradient (G) along the *z-axis*. A Bi-Polar LED sequence containing a Watergate sequence was used for water suppression.[9] For the pulse-field gradient (PFG) experiments, rectangular shaped pulse up to 65.70 Gauss/cm and sine shaped pulse up 42.17 Gauss/cm were used for the sample with and without Zn, respectively. The diffusion delay (Δ) was 100 ms with 2 ms gradient duration (δ/2) for the sample containing Zn, and 130 ms with 2.5 ms gradient duration for the sample without Zn. 31 values of gradient intensity were used, from 5-20% in 1% increments and from 25-95 % in 5% increments. Peak intensities (I) were plotted according to: $-\ln(I/I_0)=G_2\gamma_2\ \delta\ _2(\Delta-\delta/3-\tau/2)$, where γ is the gyromagnetic ratio and τ is 0.3 ms. The slope of the curve yielded the translational diffusion coefficient. For calculating the Stokes radius the TRIS peak at 3.62

ppm was used as internal standard for normalizing the solvent viscosity.[10]

Crystallography. MBPC-1 was crystallized by sitting drop vapor diffusion with a mother liquor consisting of 100 mM TRIS (pH 7) and 1.25 mM $ZnCl_2$, with 25% PEG 3350. The drop consisted of 2 μL of the protein (1.52 mM in 20 mM HEPES pH 7.5) and 1 μL mother liquor. Crystals typically appeared after a week and grew to ~ 200 μm × 100 μm × 50 μm. The crystals to be used for diffraction experiments were exchanged into a solution containing 20% glycerol as a cryoprotectant and frozen in liquid nitrogen or directly in the cryostream.

X-ray diffraction data were collected at 100 K using a Bruker Apex II CCD detector and monochromatized Cu-Kα radiation (1.54 Å) produced by a Siemens sealed source. The data were processed using SAINT and Bruker SADABS. The structure of Zn_4:MBPC-1_4 was determined at 2.9 Å resolution by molecular replacement with MOLREP[11] using the cyt cb_{562} structure (PDB ID 2BC5) as the search model. Rigid-body, simulated-annealing, positional and thermal refinement with CNS,[12] along with manual rebuilding, and water placement with XFIT,[13] produced the final model. The Ramachandran plot was calculated with PROCHECK.[14] All figures were created using Pymol.[15]

Results and Discussion

Evidence for Zn^{2+} binding by MBPC-1. It has been shown that the coordination of

Table 2.1. X-ray data collection and refinement statistics for Zn_4:MBPC-1_4.
[†] For reflections above 2σ.
[‡] $R_{sym} = \Sigma\Sigma_j |I_j - \langle I \rangle| / \Sigma\Sigma_j |I_j|$.
[§] $R = \Sigma||F_{obs}| - |F_{calc}|| / \Sigma|F_{obs}|$ (2 σ cutoff).
[II] Free R calculated against 7% of the reflections removed at random.
[¶] Root mean square deviations from bond and angle restraints.

Residues in complex	4 x (106 + 1 Heme + 1 Zn)
No. of complexes / asymmetric unit	1
Metal ions in asymmetric unit	4 Zn
Waters in asymmetric unit	20
Unit cell dimensions (Å)	88.984 x 66.659 x 68.901
	$\alpha = \gamma = \beta = 90^{o}$
Symmetry group	$P2_12_12_1$
Resolution (Å)	50 - 2.9
X-ray wavelength (Å)	1.54
Number of Unique Reflections	8145
Completeness (%)*	85.1 (66.2)
$\langle I / \sigma I \rangle$*	5.8 (1.5)
Rsym[‡] (%)*	16.0 (42.0)
R[§] (%)*	24.8 (31.2)
Free R[II] (%)*	29.5 (42.7)
Rms Bnd[¶] (Å)	0.018
Rms Ang[¶] (°)	1.33
Ramachandran plot (%)	
Residues in most favored regions	93
Residues in add.l allowed regions	7
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0

Table 2.2. Interfacial hydrogen bonding interactions in Zn$_4$:MBPC-1$_4$, determined using the PISA server (Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html))[16] for the analysis of the protein-protein interfaces formed in the complex.

Zn$_4$:MBPC1$_4$

Interface between chains	Residue/Atom	Dist (Å)	Residue/Atom
A-B	A:ASP 74[OD1]	3.03	B:HIS 63[NE2]
	A:THR 96[OG1]	3.23	B:THR 96[OG1]
B-C	B:THR 31[OG1]	3.17	C:HIS 59[NE2]
	B:GLN 41[OE1]	3.14	C:ARG 34[NH2]
	B:ASP 66[OD1]	2.87	C:ARG 34[NH2]
	B:ASP 66[OD2]	3.10	C:HIS 73[ND1]
	B:ARG 34[NH2]	2.85	C:GLN 41[OE1]
	B:ARG 34[NH2]	2.49	C:ASP 66[OD1]
A-D	A:ASP 66[OD1]	2.74	D:ARG 34[NH2]
	A:HIS 59[NE2]	3.56	D:THR 31[OG1]
	A:ARG 34[NH2]	3.35	D:GLN 41[OE1]
	A:ARG 34[NH2]	2.89	D:ASP 66[OD1]
	A:HIS 73[ND1]	2.78	D:ASP 66[OD2]
C-D	C:ASP 74[OD1]	2.83	D:HIS 63[NE2]
	C:THR 96[OG1]	2.59	D:THR 96[OG1]

Table 2.3. Crystallographically unique bond distances and angles for metal coordination in Zn_4:MBPC-1_4.

Zn1				**Zn2**			
Distance (Å)		**Angle (°)**		**Distance (Å)**		**Angle (°)**	
H63	2.18	63-Zn-73	110.3	H63	2.04	63-Zn-73	86.8
H73	2.10	63-Zn-74OD1	74.62	H73	2.2	63-Zn-74OD1	100.5
D74 OD1	2.47	63-Zn-74OD2	128.2	D74 OD1	2.57	63-Zn-74OD2	149.9
D74 OD2	2.25	63-Zn-77	131.8	D74 OD2	2.09	63-Zn-77	105.27
H77	2.13	73-Zn-74OD1	141.4	H77	2.00	73-Zn-74OD1	157.18
		73-Zn-74OD2	102.9			73-Zn-74OD2	110.0
		73-Zn-77	98.5			73-Zn-77	85.6
		74OD1-Zn-77	105.7			74OD1-Zn-77	112.7
		74OD2-Zn-77	78.2			74OD2-Zn-77	101.0

Zn3				**Zn4**			
Distance (Å)		**Angle (°)**		**Distance (Å)**		**Angle (°)**	
H63	2.22	63-Zn-73	101.3	H63	2.05	63-Zn-73	92.8
H73	2.02	63-Zn-74OD1	80.4	H73	2.29	63-Zn-74OD1	97.7
D74 OD1	2.46	63-Zn-74OD2	130.4	D74 OD1	2.65	63-Zn-74OD2	150.3
D74 OD2	2.26	63-Zn-77	125.7	D74 OD2	2.11	63-Zn-77	91.0
H77	2.07	73-Zn-74OD1	142.5	H77	2.03	73-Zn-74OD1	159.8
		73-Zn-74OD2	102.0			73-Zn-74OD2	112.8
		73-Zn-77	93.3			73-Zn-77	78.6
		74OD1-Zn-77	116.3			74OD1-Zn-77	118.3
		74OD2-Zn-77	95.8			74OD2-Zn-77	108.2

divalent metal ions by a bis-His clamp placed along an α-helix can result in a stabilization of a given protein. This stability can be observed as an increased equilibrium free energy of folding (ΔG_f), and consequent shift in mid-point denaturant concentration, of the protein as determined by chemical denaturation.[17] Guanidine hydrochloride (GuHCl) denaturation of MBPC-1 at pH 5.0 (Figure 2.2, green triangles), where cyt *cb*$_{562}$ is most stable,[4]veals that the new construct is, as expected, fairly stable itself ($\Delta G_f = -30.1$ kJ × mol^{-1} , mid-point GuHCl concentration ($[\text{GuHCl}]_{1/2}$) = 2.6 M). At pH 7.0, (Figure 2.2, blue circles), the stability is seen to decrease dramatically as evidenced by a ΔG_f of -9.9 kJ × mol^{-1} and a $[\text{GuHCl}]_{1/2}$ at 1.7 M. However, on the addition of 300 μM $ZnCl_2$ to the protein at pH 7.0 (Figure 2.2, red squares), some of this stability is regained, as reflected by an increased ΔG_f of -11.2 kJ × mol^{-1} and a $[\text{GuHCl}]_{1/2}$ at 2.0 M, suggesting that MBPC-1 is indeed binding the added metal.

Initial evidence for Zn^{2+} induced oligomerization of MBPC-1. While size exclusion chromatography (SEC) is a common technique for investigating the formation of oligomeric species, we found that only monomeric MBPC-1 species were observed in SEC experiments, regardless of whether the protein was incubated with Zn or the metal was included in the running buffer. As an alternative to SEC, we pursued DLS experiments. While preparing samples for DLS we observed that addition of greater than a 2 fold molar excess of Zn caused the protein to aggregate out of solution (Fig 2.3 B). On re-suspension of the pellet with excess EDTA (Fig. 2.3 C) or upon lowering the pH to 5, it was found that this aggregation could be reversed. This phenomenon was found to occur with MBPC-1 at much lower protein and Zn concentrations than wildtype cyt

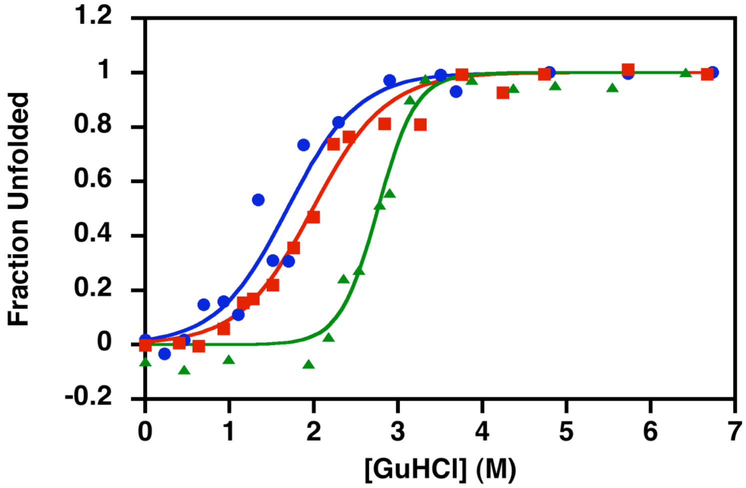


Figure 2.2. Chemical denaturation of MBPC-1. Solid lines correspond to fits to a two-state unfolding model of data collected by UV-Vis spectroscopy with 5 μM MBPC-1. The fits yield the following ΔG_f and $[GuHCl]_{1/2}$ values: pH 5 (green triangles, -30.1 kJ × mol^{-1}, 2.6 M); pH 7 (blue circles, -9.9 kJ × mol^{-1}, 1.7 M); and 300 μM Zn^{2+} in pH 7 (red squares, -11.2 kJ × mol^{-1}, 2.0 M).

cb$_{562}$, suggesting that the engineered bis-His clamps were not only able to bind the metal, but also capable of mediating new PPIs.

We then prepared fresh DLS samples with no more than 2 equivalents of Zn, as outlined in Table 2.2. What was observed was an increase of both the hydrodynamic radius and molecular weight as the ratio of Zn to protein was increased, leading to the conclusion that the coordination of the metal is mediating the formation of higher order oligomeric species of MBPC-1. Unfortunately, the DLS results rely on many assumptions about the hydrodynamic shape and monodispersity of the sample that ultimately are not very revealing about the details of the metal-dependent oligomeric state of MBPC-1 in solution and highlight the necessity for an alternative method for this purpose .

Formation of discrete oligomers of MBPC-1 on addition of Zn. As previously described, limitations of both SEC and DLS make the interpretation of the hydrodynamic properties of non-ideal samples difficult. Sedimentation velocity (SV), on the other hand, is not dependent on a flow rate greater than the rate of diffusion, as is the limiting case in SEC; nor does it assume a monodisperse sample of spherical shape, as does DLS. Thus, through its use, we sought to better understand the nature of MBPC-1 oligomers formed in solution upon binding Zn.

The SV results shown in Figure 2.4 indicate that MBPC-1 is monomeric in the presence of excess EDTA even up to 400 μM protein (red line). The sedimentation

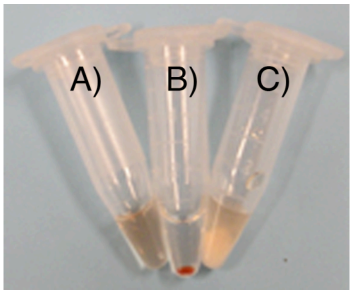


Figure 2.3. Zn^{2+} induced aggregation of MBPC-1. 30 μM MBPC-1 in a 20 mM TRIS (pH 7) solution (A), forms an aggregate (B, red pellet), in the presence of 5 mM $ZnCl_2$ in 20 mM TRIS pH 7. Upon re-suspension of the pellet with an excess of EDTA, the protein is re-solubilized (C), (white pellet is undissolved EDTA).

Table 2.4. Hydrodynamic Properties of MBPC-1 determined by DLS. Zn^{2+} was added to 649 μM MBPC-1 at ratios of 0, 2:1, 1:1 and 1:2 protein:Zn^{2+}, corresponding to 0 μM, 325 μM, 649 μM, and 1.3 mM Zn^{2+}, respectively.

$ZnCl_2$ Concentration	**Estimated Hydrodynamic Radius (nm)**	**Estimated Molecular Weight (kDa)**
0 μM	1.54726	9.34807
325 μM	2.02538	17.5526
649 μM	2.38625	25.7608
1.3 mM	3.47852	62.2204

coefficient of this monomer, at 1.8 S, correlates well to the theoretical value for monomeric MBPC-1 of 1.5 S calculated using the program HydroPro.[18] A range of samples containing a 1:1 Zn:protein ratio were next analyzed, with this ratio chosen so as to fully populate any possible Zn-mediated oligomeric species without inducing the previously described aggregation that occurs at higher ratios.

Addition of 50 μM $ZnCl_2$ to 50 μM MBPC-1 causes a broadening of the monomer peak (Figure 2.4, blue) and a shift to a higher sedimentation coefficient, consistent with the formation of a dimeric species. Increasing the concentration of metal and protein simultaneously to 200 μM (Figure 2.4, green) better resolves this dimeric species with a maximum at 2.6 S, while also revealing the formation of a tetrameric assembly at 4.4 S (theoretical value 4.5 S). This tetrameric species constitutes about 50% of all protein forms when the MBPC-1 and Zn concentrations are raised to 600 μM, which is the highest protein concentration measurable with the optics of the analytical ultracentrifuge.

To further corroborate the SV data, we sought to employ NMR spectroscopy to examine samples of MBPC-1 in the absence and presence Zn. Indicative of a well-folded, monomeric protein, the 1-D proton NMR spectrum of MBPC-1 at high concentrations (>1 mM) feature significant chemical shift dispersion (Figure 2.5 A). Upon addition of 1 equivalent of Zn, the peaks broaden considerably, as expected from the formation of a high-order oligomer (Figure 2.5 B).

In order to quantitatively determine the hydrodynamic properties of this oligomer, we utilized pulsed field gradient (PFG) diffusion NMR spectroscopy (Figure 2.5 C and D). These experiments yielded diffusion coefficients of 1.17×10^{-6} and 0.785×10^{-6} cm^2/s

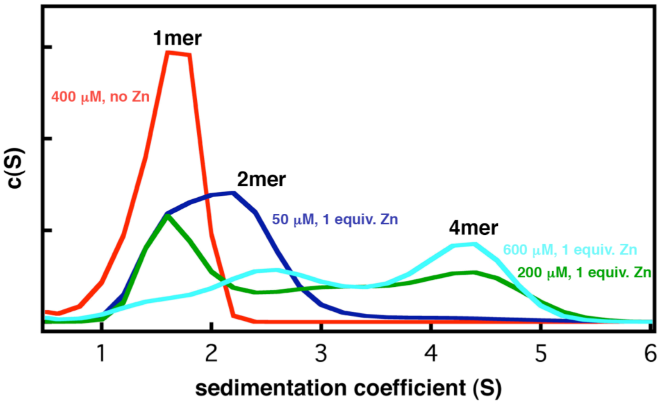


Figure 2.4. Sedimentation coefficient distributions of MBPC-1. Representative c(S) distributions of MBPC-1 at 50 μM (blue), 200 μM (green), and 600 μM (cyan) with equimolar Zn^{2+}, as well as 400 μM protein with 1 mM EDTA (red) are shown. The distributions are normalized with respect to the area covered under the curves.

for the protein in the absence and presence of 1 equivalent of Zn, respectively, with corresponding Stokes radii of 17.6 and 26 Å. The ~9-Å expansion is consistent with the formation of a MBPC-1 dimers and tetramers as observed in SV measurements..

Crystal structure of Zn_4:MBPC-1_4. In order to characterize the structure of the Zn-mediated tetrameric MBPC-1 species seen in solution, we crystallized MBPC-1 in the presence of low Zn^{2+} loadings. The resulting 2.9-Å data revealed a unique, tetrameric 16-helix complex, Zn_4:MBPC-1_4 (Figure 2.6, PDB ID:2QLA), stabilized by four Zn^{2+} ions. Here, a total of three non-crystallographic symmetry axes, which arise from the D_2 (222) symmetry of the complex, result in three distinct, inter-protomeric interfaces. These interfaces, denoted as i1, i2, and i3, bury 1080, 870, and 490 $Å^2$ of surface area, respectively. As shown in Table 2.3, the interactions along these surfaces are minimal and comprised almost entirely of polar side chains.

The overall organization of the tetramer can best be described as a pair of interlaced V-shaped dimers with a ~37° inter-protein angle. Each V is formed by the parallel alignment of two MBPC-1 molecules related by one of the three two-fold symmetry axes. The two V's are, in turn, wedged into one another in an anti-parallel fashion along a second two-fold symmetry axis. Both the formation of the V's and their interlacing are achieved entirely by inter-protein Zn coordination.

The ligands around each Zn center are supplied by three monomers (Figure 2.6 C): the 73/77 bis-His clamp from one molecule holds the Zn in a bidentate fashion; coordination of Asp74 from a second molecule stabilizes the V arrangement between the two; and coordination by His63 from a third molecule locks the two V's together. Interestingly, the 59/63 bis-His clamp is not utilized for bidentate coordination despite

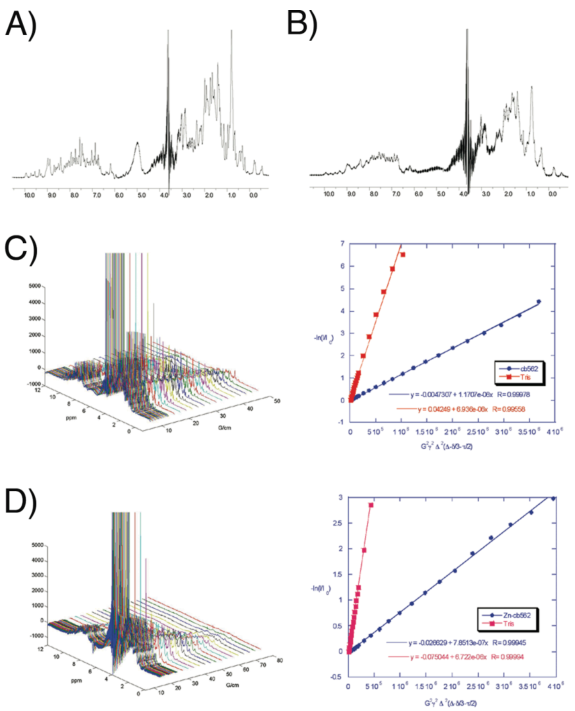


Figure 2.5. NMR spectra of MBPC-1 in the absence and presence of Zn^{2+}.1D-NMR spectra of MBPC-1 in the absence (A) and presence (B) of Zn^{2+} clearly demonstrate a peak broadening on addition of the metal indicating the formation of higher molecular weight species. Gradient-intensity-dependent NMR spectra of MBPC-1 in the absence (C) and presence (D) of Zn^{2+}, and corresponding fits to, yield the diffusional coefficients of the protein and TRIS which suggest the formation of a tetrameric species.

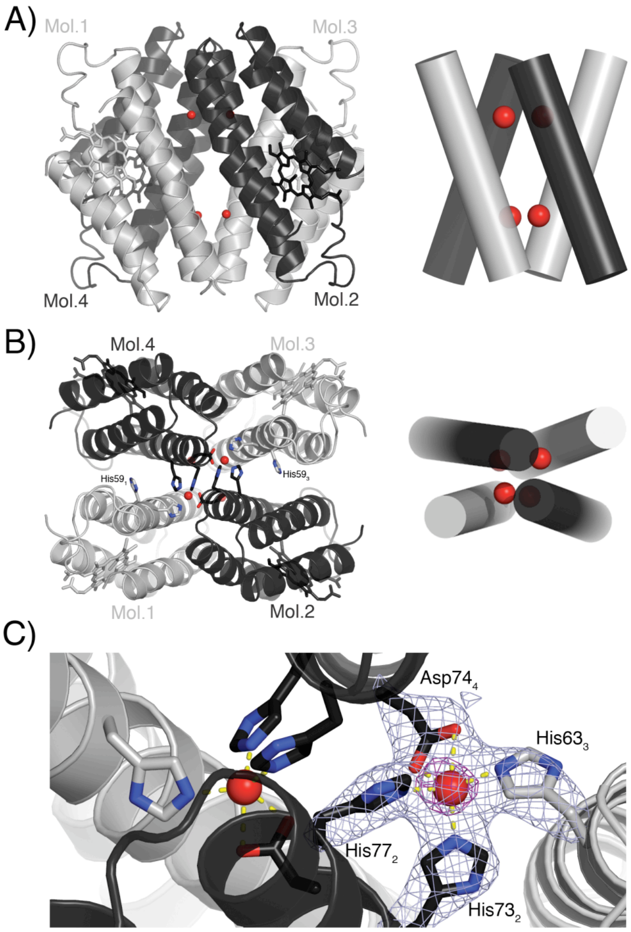


Figure 2.6. Crystal structure of the Zn$_4$:MBPC-1$_4$ tetramer. Pairs of protein molecules that form the V-shaped dimers are colored alike. Zn ions are shown as red spheres. (A) View of the assembly along the first of three non-crystallographic two-fold axes and the corresponding cylindrical representation of Helix 3 involved in Zn coordination; (B) view down the second non-crystallographic two-fold axis; (C) close-up view of the Zn coordination environment and simulated annealing *F*o - *F*c omit electron density map (gray, 4 σ; purple, 12 σ). The parent molecules for metal ligands are shown in subscripts.

the likelihood that it has a similar Zn affinity as the 73/77 couple. The His59 side chain is, in fact, H-bonded to Thr31 across the interface but is not involved in metal coordination.

The His^3Asp motif forms an identical distorted tetrahedral coordination environment around each Zn center with an average ligand-metal-ligand' angle of ~110^o(23) (Table 2.4), similar to the angles seen in the structure of tetrakis(1-methylimidazole-N^3)zinc(II) diperchlorate.[19] The metal-ligand distances in this similar small molecule complex are slightly shorter, at 1.991(2) Å, than those in Zn_4:MBPC-1_4 which has an average Zn-N_{his} distance of 2.11(9) Å, and Zn-$OD1_{asp}$ and Zn-$OD2_{asp}$ distances of 2.54(9) and 2.18(9) Å, respectively. Overall, this striking arrangement suggests that the formation of the protein assembly and metal coordination should be highly cooperative.

Conclusions

The fact that MBPC-1 is shown to be on par with wildtype cyt cb_{562} in terms of stability validates its use as a model protein onto which various metal binding groups can be installed without concern about disrupting the protein fold. As our solution-state studies show, it is clear that MBPC-1 forms Zn-dependent oligomers, specifically a dimeric and a tetrameric species when the metal is limited to a 1:1 molar ratio. The latter species, Zn_4:MBPC-1_4, appears to be the most thermodynamically favorable species at high protein concentrations as unambiguously validated by its X-ray crystal structure.

It is important to note that the interfaces between the four MBPC-1 monomers in Zn_4:MBPC-1_4 feature a large number of contacts (mostly polar) and bury a surface area exceeding 5000 Å^2, typical of a natural, stable protein oligomer. As the surfaces of cyt *cb*$_{562}$, a physiologically monomeric protein, are not optimized for self-association, these interfacial contacts are non-specific, and likely unfavorable. Thus, the formation of a discrete protein tetramer, Zn_4:MBPC-1_4, represents the first example for how a few, appropriately engineered metal ligand interactions can be used to direct the formation of discrete superprotein complexes which would otherwise be very challenging through other protein design approaches. This facile applicability of MDPSA paves the way for the generation of new biomaterials and manipulation of cellular processes.

Chapter 2 is reproduced in part with permission from: Salgado, E. N., Faraone-Mennella, J., Tezcan, F. A. **2007**. Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* 129, 13374-13375. Copyright 2007 American Chemical Society.

References

1. Faraone-Mennella, J., Tezcan, F. A., Gray, H. B., Winkler, J. R. **2006**. Stability and folding kinetics of structurally characterized cytochrome c-b562. *Biochemistry*. 45, 10504-10511.

2. Arnold, F. H., Haymore, B. L. **1991**. Engineered metal-binding proteins - Purification to protein folding. *Science*. 252, 1796–1797.

3. Salgado, E. N., Faraone-Mennella, J., Tezcan, F. A. **2007**. Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* 129, 13374-13375.

4. Faraone-Mennella, J., Tezcan, F. A., Gray, H. B., Winkler, J. R. **2006**. Stability and folding kinetics of structurally characterized cytochrome *c*-b_{562}. *Biochemistry.* 45, 10504-10511.

5. McGuirl, M. A., Lee, J. C., Lyubovitsky, J. G., Thanyakoop, C., Richards, J. H., Gray, H. B., and Winkler, J. R. **2003**. Cloning, heterologous expression, and characterization of recombinant class II cytochromes from *Rhodopseudomonas palustris. Biochim. Biophys. Acta* 1619. 23-28.

6. Nikkila, H., Gennis, R. B., and Sligar, S. G. **1991**. Cloning and expression of the gene encoding the soluble cytochrome b_{562} of *Escherichia coli. Eur. J. Biochem.* 202, 309-313.

7. Nozaki, Y. **1972**. The Preparation of guanidine hydrochloride. *Methods Enzymol.* 26, 43-50.

8. Schuck, P. **2000**. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* 78, 1606-1619.

9. Wu, D. H., Chen, A., Johnson, C. S. **1995**. Flow imaging by means of 1D pulsed-field-gradient NMR with application to electroosmotic flow. *J. Magn. Res. A* 115, 260-264.

10. Lee, H. W., Yang, W., Ye, Y. M., Liu, Z. R., Glushka, J., Yang, J. J. **2002**. Isolated EF-loop III of calmodulin in a scaffold protein remains unpaired in solution using pulsed-field-gradient NMR spectroscopy. *Biochim. Biophys. Acta.* 80-87.

11. Vagin, A., Teplyakov, A. **1997**. MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.

12. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S. K., J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M.,

Simonson, T., Warren, G. L. **1998**. Crystallography & NMR system: A new software system for macromolecular structure determination. *Acta Crystallogr. D.* 54, 905-921.

13. McRee, D. E. **1992**. A visual protein crystallographic software system for X11/XView. *J. Mol. Graphics*. 44-46.

14. Laskowski, R. A., Macarthur, M. W., Moss, D. S., Thornton, J. M. **1993**. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291.

15. DeLano, W. L. *The PYMOL molecular graphics system (http://www.pymol.org)*, 2003.

16. Krissinel, E., Henrick, K. **2007**. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 372, 774-797.

17. Krantz, B. A., Sosnick, T. R. **2001**. Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nat Struct Biol.* 8, 1042-1047.

18. Torre, J. G. d. l., Huertas, M. L., B. Carrasco. **2000**. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78, 719-730.

19. Chen, X.-M., Huang, X.-C., Xu, Z.-T., Huang, X.-Y. **1996**. Tetrakis(1-methylimidazole-N^3)zinc(II) Diperchlorate. *Acta Cryst. C.* 52, 2482-2484.

Chapter 3

Dictating the Symmetry of MBPC-1 Assemblies Through Metal Coordination Geometry

Introduction

The history of inorganic chemistry has yielded a vast collection of metal complexes whose optical, magnetic, catalytic, and structural properties have all been well characterized. These studies have led to an array of theories that attempt to explain these observed phenomena, such as crystal field theory (CFT). Application of this work, generated from years of studying small molecule complexes, to our protein-based systems should allow us to form some educated hypothesis concerning the nature of metal induced PPIs.

As an example, when one examines Ni^{2+} in terms of CFT, considering only the octahedral and tetrahedral geometries, one would expect a preference of the former geometry over the latter due to its increased crystal field stabilization energy. Comparison of tetrahedral and square-planar geometries demonstrates a similar preference for the latter when the electron configuration is low-spin and the complex consists of four strong-field ligands. The theoretical preference for Ni^{2+} to form complexes of these two geometries over those with tetrahedral arrangements is shown to be true through experimentation, as evidenced by the relatively low occurrence of tetrahedral complexes found within the Cambridge Structural Database (CSD).[1]

With these general trends in hand one can reasonably expect that a sample of MBPC-1, to which an equimolar amount of $NiSO_4$ has been added, will form a different structure from that derived through the tetrahedral coordination geometry around each Zn center in the Zn_4:MBPC-1_4 tetramer. Add to this the fact that imidazole, and by extension histidine, has been found to be a relatively weak-field ligand, lying between oxygen donors and ammonia on the spectrochemical series,[2] one would expect this Ni-mediated assembly to contain octahedrally coordinated metal centers. In fact, the tendency for divalent Zn and Ni to respectively form tetrahedral and octahedral sites within the context of a protein scaffold is another common observation born from various experimental data.[1]

With the coordination environment of MBPC-1 around differing metal ions expected to be distinct, it is reasonable to conclude that the overall symmetries of the metal-mediated assemblies should be different as well. For instance, the tetrahedral symmetry of the Zn binding sites in Zn_4:MBPC-1_4 should impart onto the overall protein assembly a tetrahedral symmetry, or at least a sub-group thereof: D_2, C_3, or C_2, given the limiting case that proteins are in fact chiral. As mentioned previously, Zn_4:MBPC-1_4 does indeed have a D_2 symmetry, and so adheres to this line of reasoning. It then follows that the octahedral coordination sphere in a Ni^{2+}-mediated assembly would result in a structure with octahedral symmetry or one of its sub-groups: tetrahedral; D_4; D_3; D_2; C_4; C_3; or C_2. Though this does suggest that a D_2 symmetric assembly may also form, it is only one of a number of potential final symmetries derived from whichever organization minimizes clashes in side chains along the induced PPI interfaces while also satisfying the coordination geometry at the metal center.

In order to investigate this notion, and further demonstrate that it is the coordination of the metal ions that are directing the oligomers we observe, we chose to study MBPC-1 in the presence of divalent Ni, Cu, and the special case of exchange-inert Pd. While the geometries of Ni^{2+} have been discussed, Pd^{2+} is known to preferentially form square-planar complexes. Cu^{2+} presents an interesting case as its preferred geometries are mainly derived from distortion of an octahedral geometry due to the Jahn-

Teller effect. Examination of these three distinctly different metals reveals that we can indeed create pleomorphic assemblies from a single metal-coordinating protein, whose morphologies are dependent on traditionally accepted metal coordination geometry preferences.

Materials and Methods

Site Directed Mutagenesis and Protein Expression/Purification/Characterization. Site directed mutagenesis was performed on the pETc-b562 plasmid (denoted as wildtype), as above, with sequencing performed by UCSD Moores Cancer Center. Mutant plasmids were transformed into BL21(DE3) *E. coli* cells along with the *ccm* heme maturation gene cassette plasmid, pEC86. Expression and purification of the protein was performed as previously described.

GuHCl denaturation. In order to determine if the ΔG_f of MPBC-1 increased in the presence of Cu^{2+} and Ni^{2+}, and so infer coordination of the metals by the protein, GuHCl titrations were performed. Samples contained 20 mM TRIS (pH 7) with 1 mM $NiSO_4$ or $CuSO_4$, 5 μM protein, and varying concentrations of GuHCl. The unfolding transition was monitored by circular dichroism (CD) at 222 nm on an AVIV 62DS spectrometer. The refractive indices of the solutions were determined with a Bausch and LombAbbe-3L refractometer.

Sedimentation Velocity of MBPC-1 with Ni^{2+} and Cu^{2+}. SV experiments were performed in order to determine the solution-state oligomerization behavior of MBPC-1 in the presence of Ni^{2+}, Cu^{2+}, and Pd^{2+}. All SV experiments were performed in 20 mM TRIS (pH 7) using appropriate volumes of 5 mM $NiSO_4$ or $CuSO_4$ in 20 mM TRIS (pH 7) to yield a 1:1 metal:protein ratio (50, 200, 400, and 600 μM protein). For Pd samples, 200 μM protein was incubated with 200 μM $Pd(NO_3)_2$ in 20 mM TRIS (pH 7) for 22 hrs. SV measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor at 47,000 rpm for a total of 180 scans per sample. The following wavelengths were used for detection: 570 nm (50 μM protein), 582 nm (200 μM protein with Cu^{2+} and Pd^{2+}), 600 nm (200 μM protein with Ni^{2+}), 620 nm (400 μM protein), and 670 nm (600 μM protein). Data was fit as previously described.

Sedimentation Equilibrium of MBPC-1 with Zn^{2+}, Ni^{2+}, and Cu^{2+}. Sedimentation equilibrium (SE) experiments were performed to determine the dissociation constants (K_d) of MBPC-1 oligomers formed in the presence of divalent Zn, Ni, and Cu. All SE experiments were performed in 20 mM TRIS (pH 7) using appropriate volumes of 5 mM $ZnCl_2$, $NiSO_4$, or $CuSO_4$ in 20 mM TRIS (pH 7) to yield a 1:1 metal:protein ratio (50, 200, 400, and 600 μM protein). Measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor at 25°C. The rotor speed was varied between 15,000 and 30,000 rpm, with scans taken at 14 and 16 hrs. Equilibrium was determined to be reached by visual inspection of the two overlaid scans. The following wavelengths were used for detection: 570 nm (50 μM protein with Ni^{2+} or Cu^{2+}), 572 nm (50 μM protein with Zn^{2+}), 600 nm (200 μM protein), 675 nm (400 μM protein with Zn^{2+} or Ni^{2+}), 685 nm (400 μM protein with Cu^{2+}) and 700 nm (600 μM protein with Zn^{2+}).

The 16 hr-scans were used for global analysis in SEDPHAT.[3] For all data sets, the menisci were assigned based on 47,000-rpm meniscus depletion scans and were fixed as constants. For samples containing Ni and Cu, data were initially fit to single-species monomer, dimer, and trimer models, as well as monomer-dimer, monomer-trimer, and monomer-dimer-trimer multi-species models. In each case, the molecular masses of the species were held constant, as was the position of the cell bottom. These masses were as follows: 12328 Da for monomeric MBPC-1; 24656 Da for the dimer; and 36984 Da for the trimer. SE scans of MBPC-1 in the presence of Ni^{2+} fit equally well to monomer-dimer, monomer-trimer, and monomer-dimer-trimer models, and were all found to have statistically indistinguishable chi-squared values. Allowing the larger species in both of the two-species models float resulted in an optimized fit wherein the second species converged to 24,000 Da, corresponding to the dimer (not shown). Floating of the trimeric species in the three-species model resulted in a third species with a mass in excess of 80,000 Da. Thus, a monomer-dimer association model was chosen to continue fitting these data to estimate thermodynamic properties of this assembly.

The monomer-dimer-trimer model was not employed to fit the SE scans for MBPC-1 in the presence of Cu^{2+}, as there was no expectation for such an association model based on SV and crystallographic experiments. However, the dimer, monomer-dimer, and monomer-trimer models were all found to fit well and have statistically indistinguishable chi-squared values. Allowing the larger species in both the two-species models to float resulted in an optimized fit wherein the second species converged to a mass of about 25,000 Da. Thus, a monomer-dimer association model was chosen to continue fitting these data to estimate thermodynamic properties of this assembly.

When fitting both the Ni and Cu data towards the monomer-dimer self-association model, the molecular mass of monomeric MBPC-1 (12,328 Da), the initial loading concentrations of protein, and the incompetent fraction (set to 0) were all held constant, while both the association constant and the bottom of the cell were treated as floating parameters.

For all data sets derived from samples containing Zn^{2+}, the menisci were assigned based on 45,000 rpm meniscus depletion scans and were fixed as constants, as was the molecular mass of the monomeric MBPC-1, 12,328 Da. In fitting the data to a monomer-dimer-tetramer model, the total amount of protein loaded for each experiment, the association constants, incompetent fraction, and the bottom of the cell were floating parameters.

For all data, a minimized fit was attained once rigorous F-statistics analyses were performed on individual floating parameters. Standard deviations for the resulting log K_a values were determined through Monte-Carlo analysis within SEDPHAT. Resulting fit statistics, log K_a values, incompetent fraction, and K_d values are shown in Table 3.1.

Crystallography. All crystals were obtained by sitting drop vapor diffusion. Cu_2:MBPC-1_2 was crystallized at 20°C with a precipitant solution consisting of 100 mM

Table 3.1. Results of global fitting of SE data for MBPC-1 in the presence of Ni^{2+}, Cu^{2+}, and Zn^{2+}. The chi-squared and rmsd values of the globally fit SE data for MBPC-1 with all metal ions are given, along with resulting association and dissociation constants and, in the case of the Zn^{2+} experiment incompetent fraction. These data correspond to the data shown in Figures 3.3, 3.4, and 3.5.

Metal		
Ni^{2+}	**χ^2 of Global Fit** 2.573789	**Fit rmsd** 0.010577
	Log$_{10}$(K_a), Monomer-Dimer 4.79480 (± 4.865530 x 10^{-1})	**Monomer-Dimer K_d** 1.6 x 10^{-5} M

Cu^{2+}	**χ^2 of Global Fit** 1.434284	**Fit rmsd** 0.004301
	Log$_{10}$(K_a), Monomer-Dimer 5.10336 (± 4.923074 x 10^{-1})	**Monomer-Dimer K_d** 7.8 x 10^{-6} M

Zn^{2+}	**χ^2 of Global Fit** 1.247513	**Fit rmsd** 0.003651
	Log$_{10}$(K_a), Monomer-Dimer 4.97945(± 8.573493 x 10^{-1})	**Monomer-Dimer K_d** 1.0 x 10^{-5} M
	Log$_{10}$(K), Dimer-Tetramer) 14.38736 (± 1.240897 x 10^{-1}	**Dimer-TetramerK_d** 1.6 x 10^{-5} M
	Incompetent Fraction 0.30427 (± 6.634254 x 10^{-3})	

HEPES (pH 7.5), 25% PEG 3350, 200 mM $CaCl_2$, 17.8 mM $CuSO_4$. The drop consisted of 2 μL protein (8.9 mM in 20 mM TRIS, pH 7) and 1 μL precipitant solution. Ni_2:MBPC-1_3 was crystallized at room temperature, with a precipitant solution that included 100 mM TRIS (pH 8.5), 23% PEG 4000, and 4.16 mM $NiSO_4$. The drop consisted of 2 μL protein (4.2 mM in 20 mM TRIS, pH 7) and 1 μL precipitant solution. Pd_2:MBPC-1_2 was crystallized at room temperature, with a precipitant solution that included 200 mM ammonium acetate, 100 mM HEPES (pH 7.5), 25% PEG 3350, and 4.7 mM $Pd(NO_3)_2$. The drop consisted of 2 μL protein (4.7 mM in 20 mM TRIS, pH 7) and 1 μL precipitant solution. Crystals for all complexes appeared within days and grew to maturation within a month. All crystals used for diffraction experiments were exchanged into a solution containing 20% glycerol as a cryoprotectant and frozen in liquid nitrogen or directly in the cryostream.

X-ray diffraction data for Cu_2:MBPC-1_2 was collected at 100 K at the Stanford Synchrotron Radiation Laboratory (BL 9-2) using 1.008-Å radiation. Another full data set was collected near the Cu K-edge (1.378 Å) to 2.3-Å resolution to calculate anomalous difference density maps to distinguish between Cu and Ca ions. The data were integrated and scaled using HKL2000.[4] X-ray diffraction data for Ni_2:MBPC-1_3 were collected at 100 K using an Bruker Apex II CCD detector and monochromatized Cu-Kα radiation (1.54 Å) produced by a Siemens sealed source. The data were processed using SAINT and Bruker SADABS. Another full data set was collected near the Ni K-edge (1.485 Å) to 2.5 Å resolution, processed with HKL2000 and used to calculate anomalous difference density maps to unambiguously identify Ni ions. X-ray diffraction data for Pd_2:MBPC-1_2 was collected at 100 K at the Stanford Synchrotron Radiation Laboratory (BL 9-2) using 0.98 Å radiation. The data were integrated with MOSFLM[5] and scaled with SCALA.[6]

The structures of Cu_2:MBPC-1_2, Ni_2:MBPC-1_3, and Pd_2:MBPC-1_2 were determined at 1.72, 2.04, and 1.9 Å resolutions, respectively, by molecular replacement with MOLREP[6,7] using the cyt cb_{562} structure (PDB ID 2BC5) as the search model. Rigid-body, positional and thermal refinement with REFMAC[6,8] along with manual rebuilding, and water placement with XFIT[9] produced the final model. The Ni_2:MBPC-1_3 diffraction data (R_3 spacegroup) were found to contain a twin fraction of ~15%. The data were detwinned using DETWIN[6] (twinning operator: k, h, -l), which resulted in the improvement of *R*free from 29.0% to 27.8%. The Ramachandran plot was calculated with PROCHECK.[10] All figures were produced with PYMOL.[11]

Electron Paramagnetic Resonance (EPR) spectroscopy. MBPC-1 EPR samples consisted of 150 μM protein in 20 mM MOPS buffer pH 7, with or without 100 μM $CuCl_2$ dihydrate. X-band spectra were obtained with a Bruker ELEXSYS E500 spectrometer equipped with a Bruker ER4131VT variable-temperature unit. The spectrum of the copper free sample was subtracted from that of the copper-containing sample to obtain the final adjusted spectrum. Spectra were recorded at 125 K with the following spectrometer conditions: Microwave frequency, 9.389688 GHz; power, 0.63 mW; modulation amplitude, 0.10 mT. The spectra were simulated using the W95EPR

Table 3.2. X-ray data collection and refinement statistics for Cu_2:MBPC-1_2 and Ni_2:MBPC-1_3.

‡Rsym= $\Sigma\Sigma j|\,I j-\langle I\rangle| / \Sigma\Sigma j|\,I j|$.

§R= $\Sigma|\,|Fobs| - |Fcalc|\,| / \Sigma|\,Fobs|$ (2 σ cutoff).

IIFree R calculated against 7% of the reflections removed at random.

¶Root mean square deviations from bond and angle restraints.

*Numbers in parentheses correspond to the highest resolution shell (1.76 – 1.72 Å for Cu_2:MBPC-1_2; 2.14 – 2.04 Å for Ni_2:MBPC-1_3)

	Cu_2:MBPC-1_2	**Ni_2:MBPC-1_3**
Residues in complex	2 x (106 + 1 Heme) + 2 Cu	3 x (106 + 1 Heme) + 2 Ni
No. of complexes / asymmetric unit	2	1/3
Metal ions in asymmetric unit	4 Cu, 6 Ca	2/3 Ni (intersubunit), 1 (intrasubunit)
Waters in asymmetric unit	511	57
Unit cell dimensions (Å)	47.57 x 89.50 x 52.02	a = b = 50.21, c = 121.97
	$\alpha = \gamma = 90^\circ$, $\beta = 111.0^\circ$	$\alpha = \gamma = 90^\circ$, $\beta = 120^\circ$
Symmetry group	$P2_1$	$R3$
Resolution (Å)	50 - 1.72	50 - 2.04
X-ray wavelength (Å)	1.01 1.378 (Cu K edge)	1.54
Number of Unique Reflections	42391	7499
Redundancy	5.7	4.6
Completeness (%)*	98.3 (100)	99.8 (99.2)
$\langle I / \sigma I\rangle$*	7.4 (4.2)	11.2 (2.0)
Rsym‡ (%)*	7.7 (20.9)	6.6 (49.2)
R§ (%)*	19.5 (22.0)	21.6 (26.1)
Free RII (%)*	24.2 (29.8)	27.8 (27.4)
Rms Bnd¶ (Å)	0.011	0.022
Rms Ang¶ (°)	1.22	1.95
Ramachandran plot (%)		
Residues in most favored regions	96.9	97.9
Residues in add.l allowed regions	3.1	2.1
Residues in generously allowed regions	0.0	0.0
Residues in disallowed regions	0.0	0.0

Table 3.3. X-ray data collection and refinement statistics for Pd$_2$:MBPC-1$_2$.
‡Rsym= $\Sigma\Sigma j|\,Ij - \langle I \rangle| / \Sigma\Sigma j|\,Ij|$.
§R= $\Sigma|\,|Fobs| - |Fcalc|\,| / \Sigma|Fobs|$ (2 σ cutoff).
IIFree R calculated against 7% of the reflections removed at random.
¶Root mean square deviations from bond and angle restraints.
*Numbers in parentheses correspond to the highest resolution shell (2.00 – 1.90)

	Pd$_2$:MBPC-1$_2$
Residues in complex	2 x (106 + 1 Heme) + 2 Pd
No. of complexes / asymmetric unit	1
Metal ions in asymmetric unit	2 Pd
Waters in asymmetric unit	73
Unit cell dimensions (Å)	41.52 x 64.78 x 89.04
	$\alpha = \beta = \gamma = 90^\circ$
Symmetry group	$P2_12_12_1$
Resolution (Å)	44.5 - 1.90
X-ray wavelength (Å)	0.98
Number of Unique Reflections	19546
Redundancy	4.3
Completeness (%)*	99.7 (99.5)
$\langle I / \sigma I \rangle$*	14.5 (4)
Rsym‡ (%)*	6.3 (34.4)
R§ (%)*	20.5 (28.8)
Free R^{II} (%)*	25.5 (29.1)
Rms Bnd¶ (Å)	0.016
Rms Ang¶ (°)	1.56
Ramachandran plot (%)	
Residues in most favored regions	97.4
Residues in add.l allowed regions	2.6
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0

Table 3.4. Interfacial hydrogen bonding interactions in Cu_2:MBPC-1_2, Ni_2:MBPC-1_3, and Pd_2:MBPC-1_2 complexes determined using the PISA server (Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)).[12] The interactions of residues involved in interfacial salt-bridges in each assembly are highlighted in red. The variation of H-bonding pattern between the two Cu_2:MBPC-1_2 interfaces, despite the near-identical overall structures, is a strong indication that these interactions do not play a role in complexation.

Cu_2:MBPC-1_2

Interface between chains	Residue/Atom	Dist (Å)	Residue/Atom
A-D	A:ASP 54[OD1]	2.68	D:LYS 27[NZ]
	A:ASP 54[OD2]	3.95	D:LYS 27[NZ]
	A:ASP 54[OD2]	3.81	A:LYS 27[NZ]
B-C	B:GLN 41[OE1]	3.57	C:ARG 34 [NH2]
	B:ASP 54[OD1]	3.26	C:ASN 80[ND2]
	B:ASP 66[OD1]	3.85	C:ARG 34[NH2]
	B:ARG 34[NH1]	2.82	C:GLN 41 [OE1]
	B:LYS 27[NZ]	2.66	C:ASP 54[OD2]

Ni_2:MBPC-1_3

Interface between chains	Residue/Atom	Dist (Å)	Residue/Atom
A-A	A:ARG 34[NE]	3.37	A:ASP 74[OD2]

Pd_2:MBPC-1_2

Interface between chains	Residue/Atom	Dist (Å)	Residue/Atom
A-B	A:GLN 41[OE1]	2.71	B:ARG 34[NH1]
	A:ASP 54[OD1]	3.08	B:LYS 27[NZ]
	A:ASP 66[OD1]	3.69	B:ARG 34[NH1]
	A:ARG 34[NH2]	3.25	B:GLN 41 [OE1]
	A:ASN 80[ND2]	3.05	B:ASP 54[OD1]
	A:ASN 80[ND2]	3.26	B:ASP 54[OD2]

Table 3.5. Crystallographically unique bond distances and angles for metal coordination in Cu_2:MBPC-1_2, Ni_2:MBPC-1_3, and Pd_2:MBPC-1_2. For brevity, the histidine residues involved in Cu, Ni, and Pd coordination were listed using only their residue number in describing angles. The His-Cu-OH_2 angles refer to the axially coordinated water (H_2O-1) that is positioned ~2.5 Å from the Cu ions. For Ni_2:MBPC-1_3, the subscripts (1, 2, 3) for His residues refer to the three symmetry-related MBPC-1 monomers.

Cu1				**Cu2**				**Cu3**				**Cu4**			
Distance (Å)		**Angle (°)**		**Distance (Å)**		**Angle (°)**		**Distance (Å)**		**Angle (°)**		**Distance (Å)**		**Angle (°)**	
H59	2.02	59-Cu-63	92.1	H59	2.11	59-Cu-63	84.9	H59	2.11	59-Cu-63	87.5	H59	2.05	59-Cu-63	90.8
H63	2.08	59-Cu-73	167.9	H63	2.12	59-Cu-73	180	H63	2.09	59-Cu-73	180	H63	2.12	59-Cu-73	180
H73	2.05	59-Cu-77	90	H73	2.06	59-Cu-77	93.6	H73	2.04	59-Cu-77	86.6	H73	2.05	59-Cu-77	90.5
H77	2.09	63-Cu-73	88.8	H77	2.10	63-Cu-73	90.2	H77	2.11	63-Cu-73	91.8	H77	2.14	63-Cu-73	90.1
H2O-1	2.59	63-Cu-77	180.0	H2O-1	2.58	63-Cu-77	180.0	H2O-1	2.53	63-Cu-77	180.0	H2O-1	2.51	63-Cu-77	180.0
H2O-2	3.64	73-Cu-77	89.2	H2O-2	3.47	73-Cu-77	91.3	H2O-2	3.54	73-Cu-77	94.1	H2O-2	3.37	73-Cu-77	88.7
		59-Cu-OH_2	97.4			59-Cu-OH_2	86.3			59-Cu-OH_2	95.6			59-Cu-OH_2	89.1
		63-Cu-OH_2	93.0			63-Cu-OH_2	89.9			63-Cu-OH_2	95.2			63-Cu-OH_2	94.8
		73-Cu-OH_2	94.8			73-Cu-OH_2	96.3			73-Cu-OH_2	95.4			73-Cu-OH_2	96.7
		77-Cu-OH_2	86.2			77-Cu-OH_2	87.5			77-Cu-OH_2	84.6			77-Cu-OH_2	83.7

Ni1				**Ni2**			
Distance (Å)		**Angle (°)**		**Distance (Å)**		**Angle (°)**	
H59	2.24	59_1-Ni-63_1	81.7	H73	2.21	73_1-Ni-77_1	91.1
H63	2.09	59_1-Ni-59_2	92.5	H77	2.18	73_1-Ni-73_2	92.8
		59_1-Ni-63_2	92.3			73_1-Ni-77_2	86.7
		59_1-Ni-59_3	92.5			73_1-Ni-73_3	92.7
		59_1-Ni-63_3	180.0			73_1-Ni-77_3	180.0
		63_1-Ni-59_2	180.0			77_1-Ni-73_2	180.0
		63_1-Ni-63_2	93.8			77_1-Ni-77_2	89.6
		63_1-Ni-59_3	92.3			77_1-Ni-73_3	86.2
		63_1-Ni-63_3	93.8			77_1-Ni-77_3	89.9

Pd1				**Pd2**			
Distance (Å)		**Angle (°)**		**Distance (Å)**		**Angle (°)**	
H59	2.02	59-Pd-63	89.2	H59	2.23	59-Pd-63	94.1
H63	2.01	59-Pd-73	173.0	H63	1.99	59-Pd-73	177.0
H73	1.95	59-Pd-77	89.1	H73	2.21	59-Pd-77	94.6
H77	1.98	63-Pd-73	92.7	H77	1.92	63-Pd-73	88.2
		63-Pd-77	173.7			63-Pd-77	170.6
		73-Pd-77	89.7			73-Pd-77	83.0

program for one copper(II) center, S = 1/2 and four equivalent nitrogens. Simulation parameters: $g_1 = 2.06(2)$; $g_2 = 2.06(1)$; $g_3 = 2.26(1)$, $W_1 = 1.05(8)$ mT, $W_2 = 0.59(8)$ mT, $W_3 = 0.90(8)$ mT, gaussian lineshapes. Hyperfine and superhyperfine coupling: Cu (I = 3/2): A_z = 540(20) MHz (0.018 cm^{-1}), $A_x = A_y$ = 60(20) MHz (0.0020 cm-1); N (I = 1) isotropic A_{iso} = 43(1) MHz (0.0014 cm^{-1}).[13]

Results and Discussion

Evidence for Ni^{2+}and Cu^{2}+ binding by MBPC-1. To determine if the addition of either Ni^{2+} or Cu^{2+} to a sample of MBPC-1 results in a stabilization of the protein similar to that seen in the presence of Zn^{2+}, we again looked to GuHCl denaturation, this time monitoring the transition by CD at 222 nm. At pH 5.0 (Figure 3.1, green triangles), a ΔG_f of -33.2 kJ x mol^{-1} was determined with a [GuHCl]$_{1/2}$ at 2.7 M. These values are fairly consistent with those determined by UV-Vis spectroscopy and again demonstrate MBPC-1 to be stable. At pH 7.0, (Figure 3.1, blue circles), the stability is again seen to decrease, as indicated by a ΔG_f of -21.2 kJ x mol^{-1} with a [GuHCl]$_{1/2}$ at 1.8 M. Although this decrease in ΔG_f is less dramatic than that measured by UV-Vis, possibly due to differences in noise between the two types of data collection, it nonetheless again reveals a decreased stability of the MBPC-1 construct on increasing the pH.

As no metal induced aggregation was witnessed at any Ni or Cu loads, the amount of these metals added to CD samples was increased to 1 mM from the 300 μM used in Zn^{2+} samples. As seen in Figure 3.1 A and B, the addition of both Ni and Cu restore the

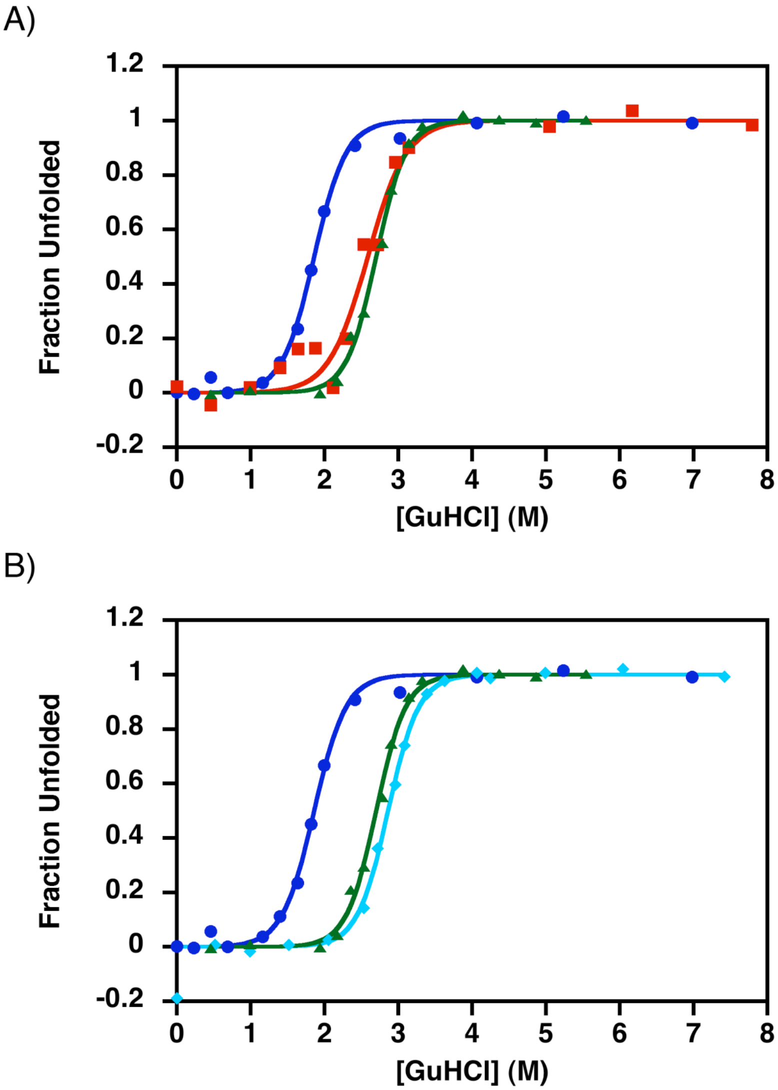


Figure 3.1. Chemical denaturation of MBPC-1 in the presence of Ni^{2+} and Cu^{2+}. Solid lines correspond to fits to a two-state unfolding model of data collected by CD spectroscopy with 5 μM MBPC-1. Resulting ΔG_f and $[GuHCl]_{1/2}$ for MBPC-1 at pH 5 (green triangles, -33.2 kJ x mol^{-1}, 2.7 M) and pH 7 (blue circles, -21.2 kJ x mol^{-1}, 1.8 M) are shown in both A and B. A titration in the presence of 1 mM Cu^{2+} (A, red squares) yields a ΔG_f = -24.3 kJ x mol^{-1} and [GuHCl]1$_{/2}$ = 2.6 M. In the presence of 1 mM Ni^{2+} (B, cyan diamonds), ΔG_f = -33.9 kJ x mol^{-1} and $[GuHCl]_{1/2}$ = 2.9 M.

stability of the protein to that at pH 5 (Cu: ΔG_f = -24.3 kJ x mol^{-1} and $[GuHCl]_{1/2}$ = 2.6 M; Ni: ΔG_f = -33.9 kJ x mol^{-1} and $[GuHCl]_{1/2}$ = 2.9 M). We interpret these data as an indication that MPBC-1 is capable of binding both Ni and Cu, possibly to a far greater degree than the protein could bind Zn.

Formation of dimeric MBPC-1 assemblies based on Ni^{2+} and Cu^{2+} coordination. As discussed, coordination of Ni and Cu by MBPC-1 will most likely result in assemblies that differ from that induced by the coordination of Zn. Assuming the coordination indicated in the GuHCl titrations with both of these metals is localized at the installed bis-His clamps, and given that no Cu- or Ni-mediated aggregation was observed at any metal:protein ratio, one would expect to see the formation of discreet metal-mediated oligomers when samples are examined by SV.

As observed in Figure 3.2 A, samples containing a 1:1 Cu^{2+}:MBPC-1 ratio form a fully populated peak with a sedimentation coefficient of ~2.6 S, close to the theoretical value of a dimer of MBPC-1. The fact that this dimer is fully populated, even at concentrations as low as 50 μM, is consistent with the higher degree of Cu binding by MBPC-1 indicated in the GuHCl titrations. A similar trend, seen in Figure 3.2 B, is observed for MBPC-1 with 1:1 Ni^{2+}, with a dimeric species being readily formed even at concentrations that are relatively low compared to those required to form Zn_4:MBPC-1_4. Nonetheless, more monomeric MBPC-1 is seen in the presence of Ni than Cu, suggesting that the latter forms more readily.

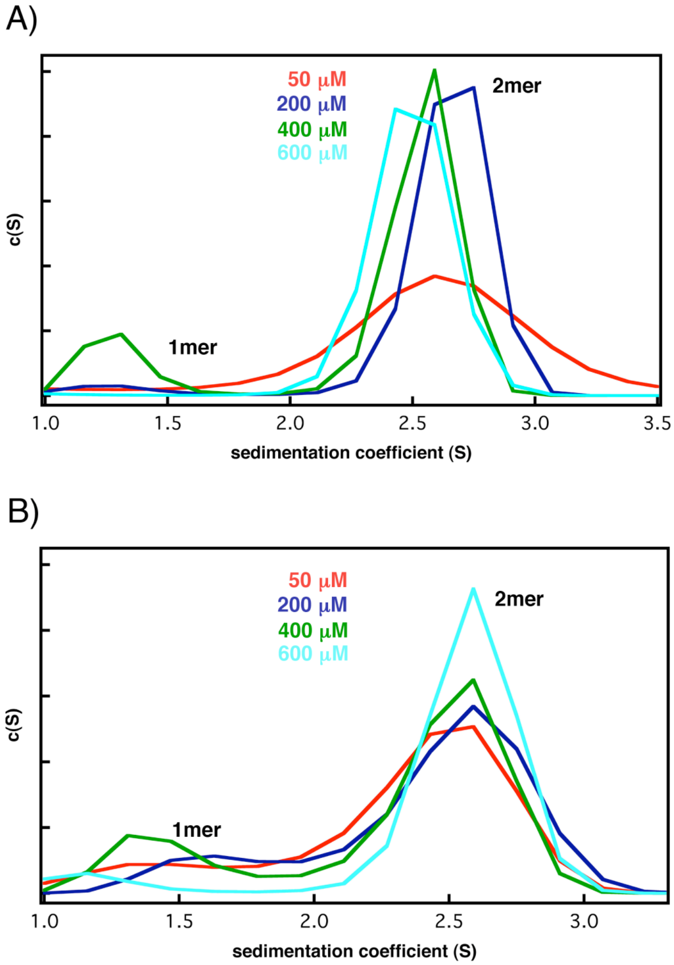


Figure 3.2. Sedimentation coefficient distributions of MBPC-1 in the presence of Ni^{2+} and Cu^{2+}. Representative c(s) distributions of MBPC-1 at 50 μM (red), 200 μM (blue), 400 μM (green), and 600 μM (cyan) with equimolar Cu^{2+}(A) or Ni^{2+} (B). The distributions are normalized with respect to the area covered under the curves.

Differences in apparent K_d of metal-mediated MBPC-1 oligomers. We next employed sedimentation equilibrium (SE) to determine K_d values for the formation of the different metal-mediated oligomeric species. Initial experiments began with samples containing MBPC-1 with a 1:1 molar ratio of added $ZnCl_2$ at varying concentration between 50 and 600 μM, similar to the concentrations used in Figure 2.4. Absorbance scans of these samples collected at varying rotor speeds results in data that can be globally fit to a monomer-dimer-tetramer (M-D-T) model (Figure 3.3, Table 3.1). This model, however, was found to fit only after the inclusion of an incompetent fraction of the protein.

The use of this added component generally describes some portion of the sample that is unavailable to take part in the equilibrium being modeled. In this particular system, this fraction would be expected to include a population of MBPC-1 that has not bound any of the added Zn^{2+}. Accordingly, the fits of the SE data reveal that ~30% of MBPC-1 in the sample is not bound to Zn, thus does not take part in the equilibrium. This is seemingly consistent with the relatively small increase in stability seen in the GuHCl titrations with added Zn and the idea that MBPC-1 does not readily bind the metal. Keeping this in mind, a predicted M-D K_d, derived through visual inspection of the Zn:MBPC-1 SV distributions, could reasonably be estimated at ~50 μM. Inclusion of this incompetent fraction, however, results in much lower K_d values from the SE fits. Thus, a M-D K_d of 10 μM, as calculated from these SE experiments, is actually reasonable, as is the D-T K_d of 16 μM.

Similar analysis of samples ranging from 50 to 400 μM MBPC-1 with equimolar $CuSO_4$ are found to fit best to a M-D self association model (Figure 3.4 A). Fitting these

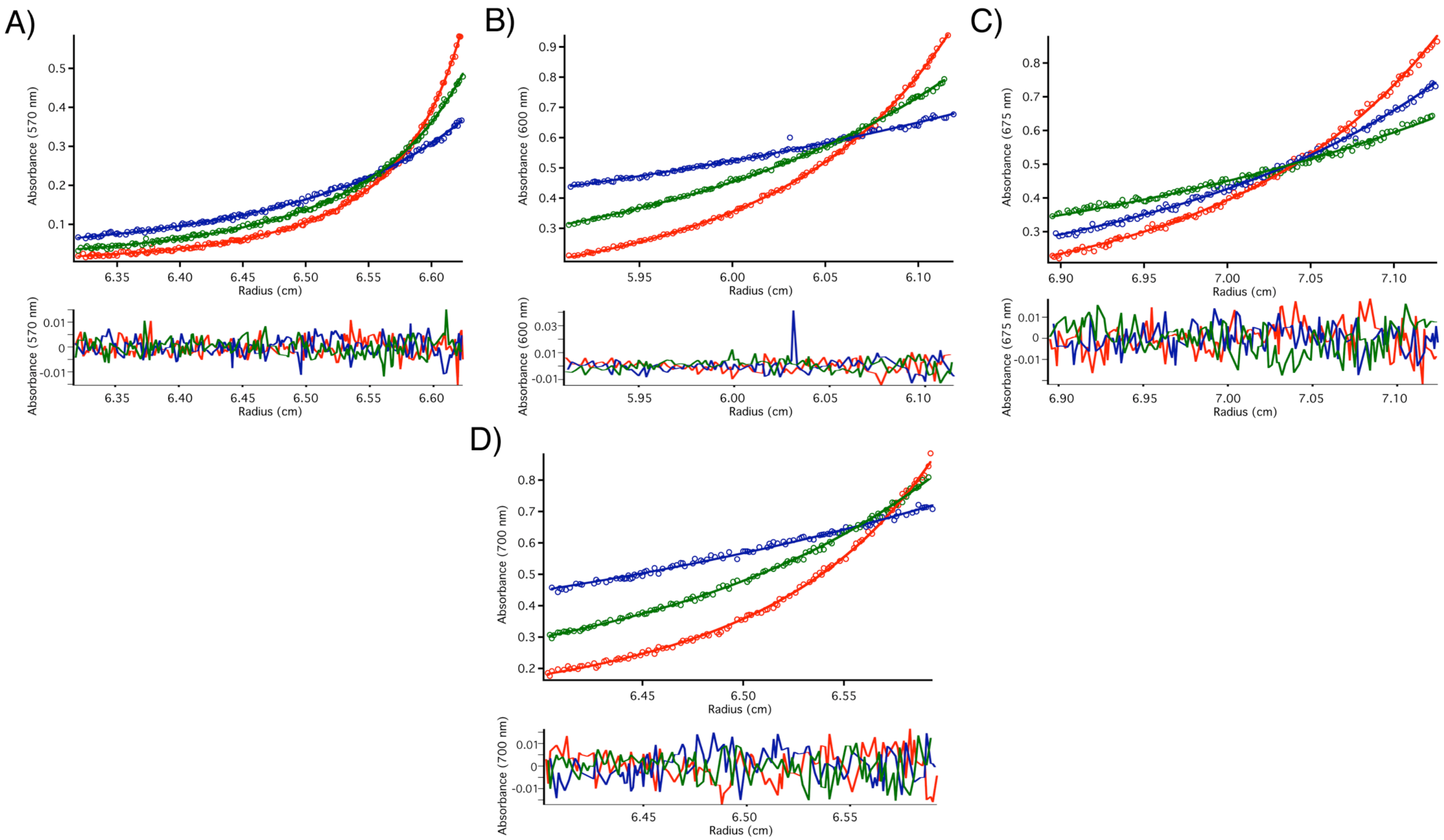


Figure 3.3. Sedimentation equilibrium data for MBPC-1 in the presence of Zn^{2+}. A) 50 μM protein at 20,000, 25,000, and 30,000 rpm. B) 200 μM protein at 10,000, 15,000, and 20,000 rpm. C) 400 μM protein at 10,000, 15,000, and 20,000 rpm. D) 600 μM protein at 10,000, 12,5000, and 15,000 rpm. All samples were run in 20 μM TRIS (pH 7), at 25°C, with equimolar $ZnCl_2$. Scans were globally fit to a monomer-dimer-tetramer model yielding a minimized fit whose results and statistics are shown in Table 3.1.

data to this model (Figure 3.4 B-D, Table 3.1), can be accomplished without the inclusion of an incompetent fraction. This would suggest that all of the MBPC-1 present is bound to Cu in some way, correlating well with the previously described denaturation titrations and thus explaining the accelerated stability of MBPC-1 in the presence of Cu.

The resulting fit of the SE data reveals a M-D K_d of ~7.8 μM. Comparing this value to the 10 μM K_d of the Zn-MBPC-1 M-D association would suggest that these two interactions should be on par with each other in terms of their favorability. Again, looking back to the previously described SV experiments, this is clearly not the case. Thus it can be concluded that it is the difference in the ability of MBPC-1 to coordinate Cu over Zn, as described by incompetent Zn-free fraction and lesser metal induced stabilization, that differentiates the formation of these metal induced oligomers.

SE experiments containing 50 to 400 μM MBPC-1 with equimolar $NiSO_4$ describe a similar phenomenon as seen in the Cu containing samples. Here the data is again best described by a M-D association (Figure 3.5 A) without the inclusion of an incompetent fraction, explaining both the high population of dimer observed in the SVs and the high level of stability seen in the GuHCl titration. From the SE fits (Figure 3.5 B-D, Table 3.1), a 16 μM M-D K_d for the Ni^{2+} mediated assembly was found. This value is about two-fold greater than the Cu-mediated K_d, which explains the more pronounced monomer presence in Ni containing SV samples. Again, absence of an incompetent fraction in these samples explains the reason for the clear increased favorability of formation of the Ni complex over the Zn tetramer despite seemingly equivalent K_d values.

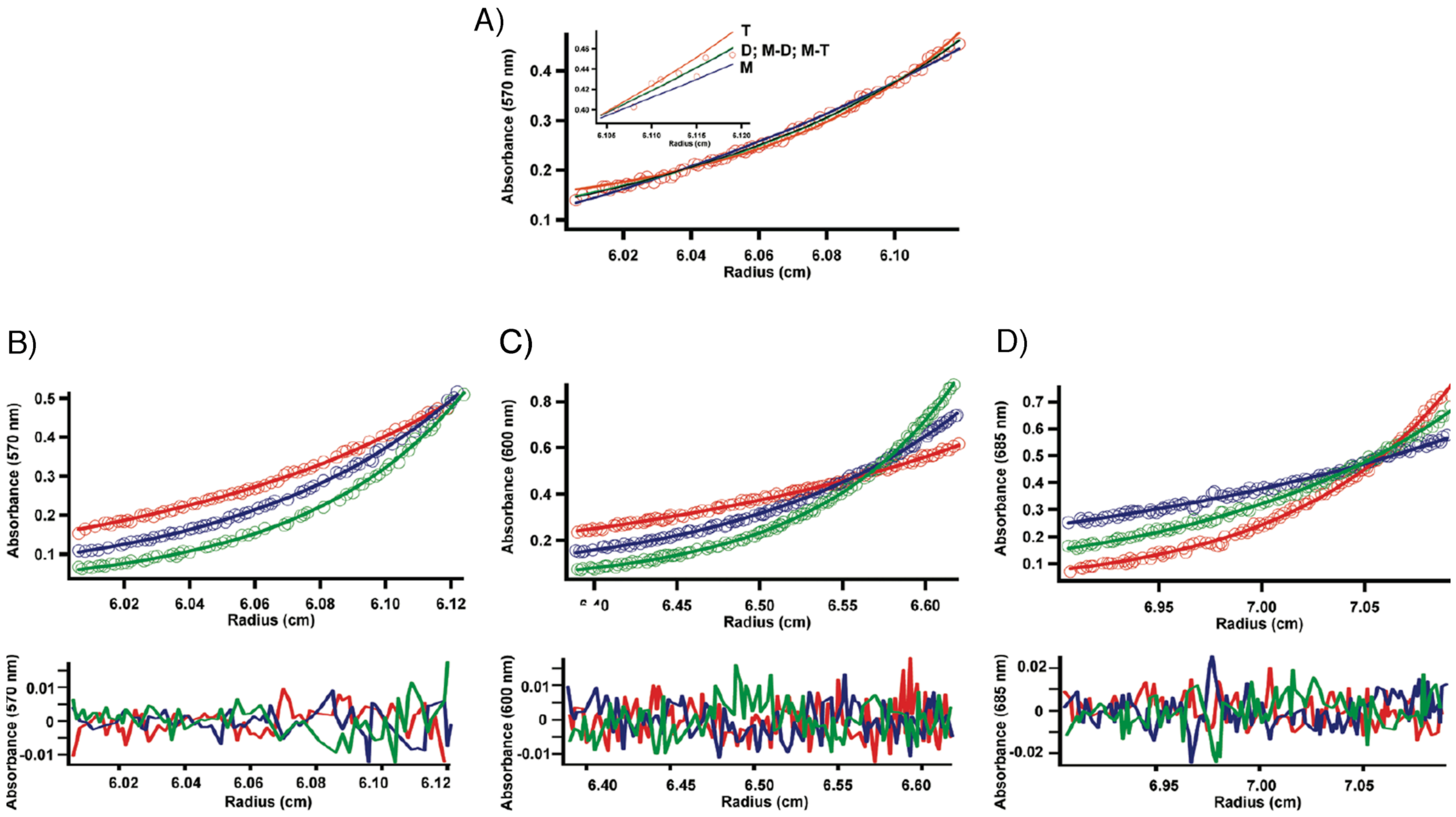


Figure 3.4. Sedimentation equilibrium data for MBPC-1 in the presence of Cu^{2+}. A) Comparison of species analysis models used to globally fit SE data. Only data from the 20,000 rpm scan from the 50 µM protein sample are shown (red circles) with single species monomer (blue), dimer (green), and trimer (orange) models, and multi-species monomer-dimer (cyan) and monomer-trimer (black) models overlaid. (inset) Close-up of the fits of single species monomer (M), dimer (D), and trimer (T) models, and multi-species monomer-dimer (M-D) and monomer-trimer (M-T) demonstrating the favorable fitting of the single-species dimer model, as well as the two-species monomer-dimer and monomer-trimer models. B) 50 µM protein at 25,000, 30,000, and 35,000 rpm. C) 200 µM protein at 15,00, 20,000, and 25,000 rpm. D) 400 µM protein at 15,000, 20,000, and 25,000 rpm. All samples were run in 20 mM TRIS buffer (pH 7), at 25° C, with equimolar amounts of $CuSO_4$. Scans shown in B), C) and D) were globally fit to a monomer-dimer model yielding a minimized fit whose results and statistics are shown in Table 3.1.

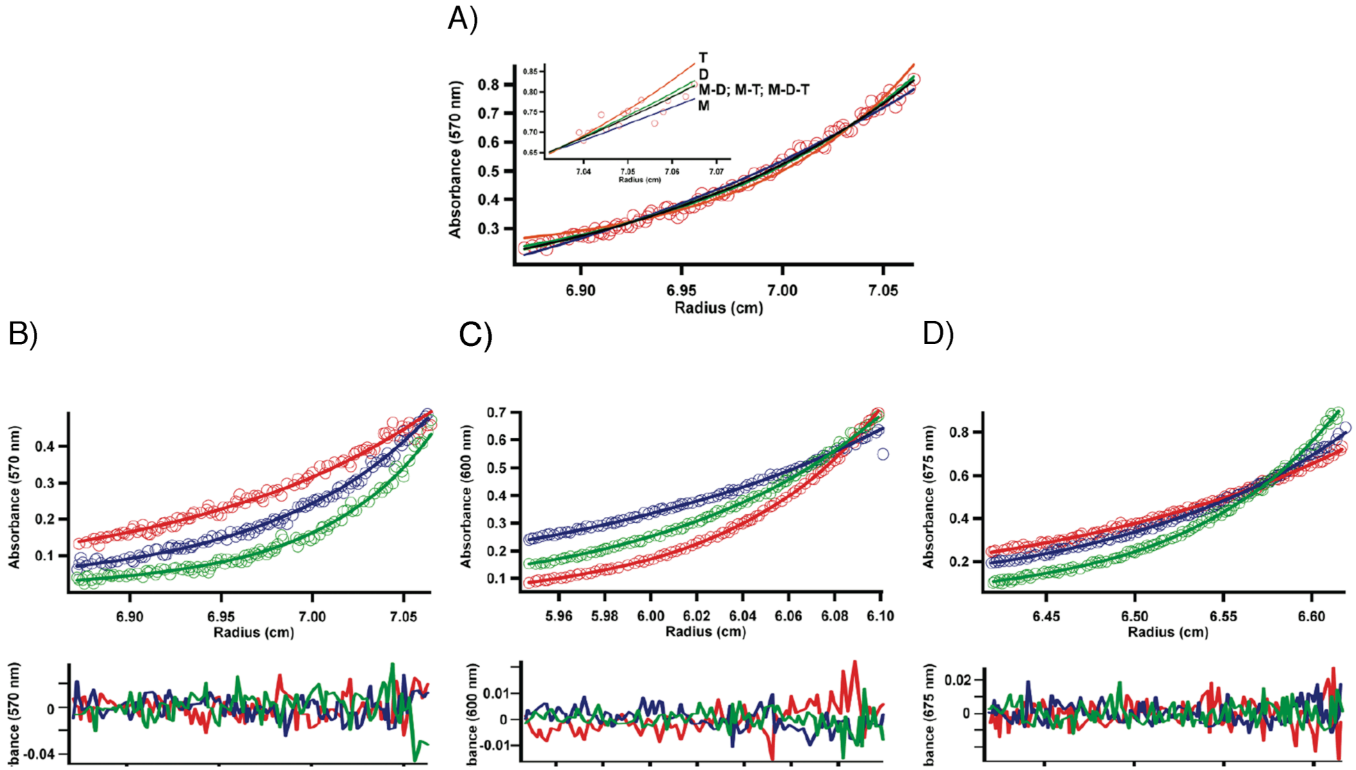


Figure 3.5. Sedimentation equilibrium data for MBPC-1 in the presence of Ni^{2+}. A) Comparison of species analysis models used to globally fit SE data. Only data from the 20,000 rpm scan from the 50 μM protein sample are shown (red circles) with single species monomer (blue), dimer (green), and trimer (orange) models, and multi-species monomer-dimer (cyan), monomer-trimer (light green), and monomer-dimer-trimer (black) models superimposed. (inset) Close-up of the fits of single species monomer (M), dimer (D), and trimer (T) models, and multi-species monomer-dimer (M-D), monomer-trimer (M-T), and monomer-dimer-trimer (M-D-T) models, demonstrating the favorable fitting of the latter models. B) 50 μM protein at 20,000, 25,000, and 30,000 rpm. C) 200 μM protein at 17,500, 20,000, and 25,000 rpm. D) 400 μM protein at 20,000, 25,000, and 30,000 rpm. All samples were run in 20 mM TRIS buffer (pH 7), at 25° C, with equimolar amounts of $NiSO_4$. Scans shown in B), C) and D) were globally fit to a monomer-dimer model yielding a minimized fit whose results and statistics are shown in Table 3.1.

Crystallization of the Cu_2:MBPC-1_2 dimer. With sufficient solution state data describing a Cu^{2+} mediated dimer of MBPC-1, we looked to crystallize the assembly in order to compare it to Zn_4:MBPC-1_4. The resulting structure (PDB ID: 3DE8)[14] reveals four MBPC-1 monomers in the asymmetric unit, which, much like in the structure of cyt cb_{562}, are aligned in pairs in an anti-parallel fashion along helix3 (Figure 3.6 A). Unlike the wildtype structure in which these contacts were mediated by CPIs, these two dimeric units are actually held together through the coordination of two interfacial Cu ions per dimeric pair. These four distinct Cu ions are in turn differentiated from other metals in the structure by distinct anomalous difference peaks calculated from data collected at the Cu K edge.

Consideration of each of these dimeric pairs presents us with the Cu_2:MBPC-1_2 assembly, consistent with the previously described solution state studies. This dimer possesses an overall C_2 symmetry, with the 2-fold symmetry axis bisecting the Cu-Cu axis (Figure 3.6 B and C). Unlike the interface of Zn_4:MBPC-1_4, the buried surface area between the monomers in both of the crystallographically distinct Cu_2:MBPC-1_2 dimers is small (800 $Å^2$ including the metal coordination sphere). As in Zn_4:MBPC-1_4, this interface is devoid of favorable side-chain interactions that would typically be expected to drive protein oligomerization (Table 3.4).

The Cu ions are found in a square-pyramidal coordination sphere, comprised of one bis-His motif from each monomer to form the equatorial coordination plane with an axial aquo ligand (Figure 3.6 D). An inspection of the Cu_2:MBPC-1_2 structure invokes the possibility that there could be rotational freedom about the Cu-Cu axis to some extent. A superposition of the two Cu_2:MBPC-1_2 dimers observed in the asymmetric unit

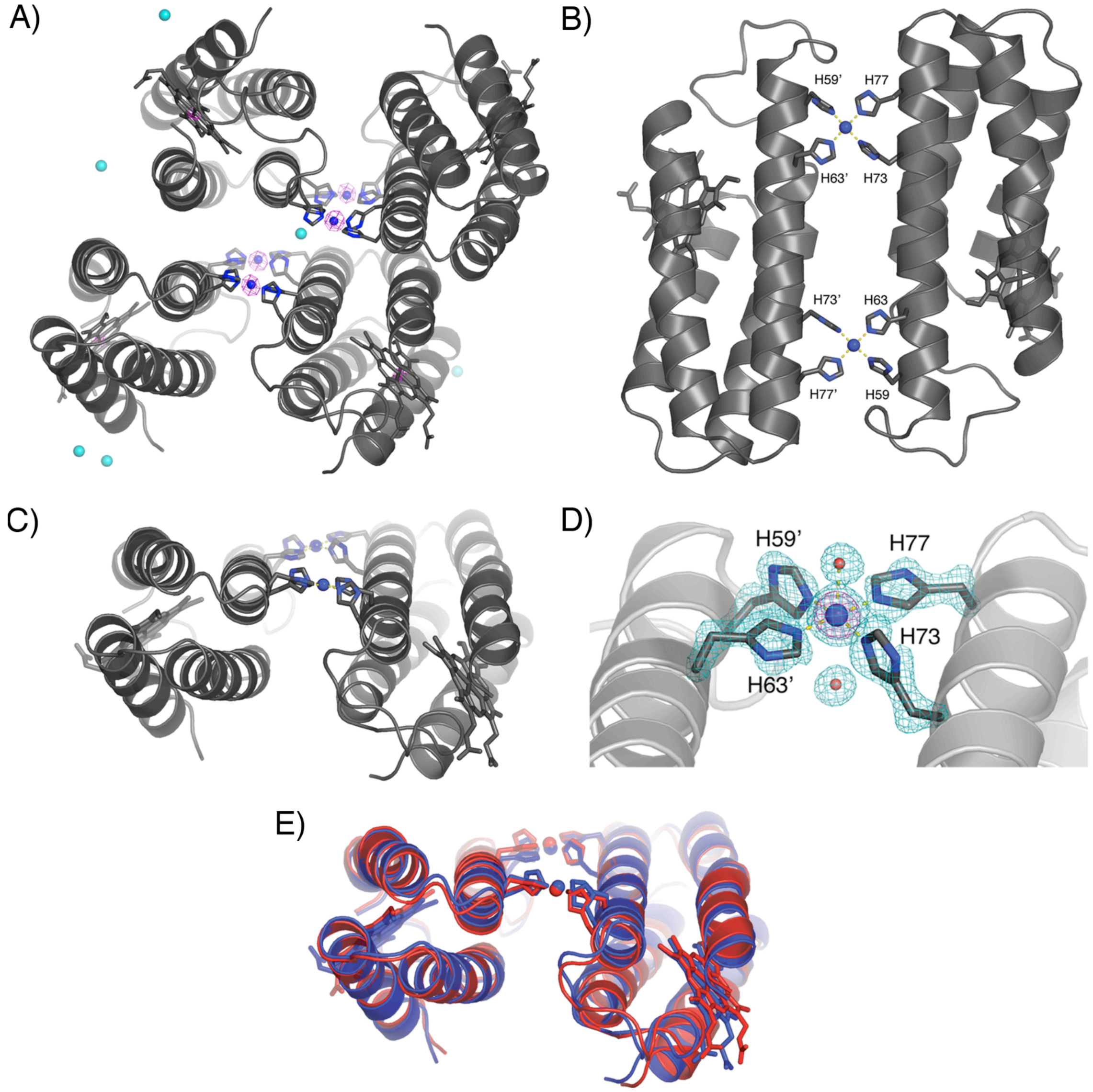


Figure 3.6. Crystal structure of Cu_2:MBPC-1_2. A) Full asymmetric unit of MBPC-1 crystallized in the presence of Cu^{2+}. Cu ions are shown in blue, and the six calcium ions involved in lattice packing interactions are shown in cyan. The anomalous difference density map calculated using 1.378-Å radiation (Cu K-edge) is shown in magenta (15σ), very clearly highlighting the position of the Cu ions. In (B) and (C) one of the two dimers seen in the asymmetric unit are shown from the side and top, respectively. D) A representative Cu coordination environment in Cu_2:MBPC-1_2 and the corresponding simulated-annealing *F*o - *F*c omit electron density maps (cyan, 3.2*σ*; magenta, 8*σ*). Water molecules shown as red spheres are positioned at 2.6 and 3.5 Å, respectively, from Cu. Axial coordination to the second water molecule is likely overcome by local electrostatic effects. E) Overlay of the two Cu_2:MBPC-1_2 assemblies in the asymmetric unit.

reveals that the relative orientations of the monomers in these molecules are nearly identical (rmsd over all Cα's = 0.885 Å; Figure 3.6 E), suggesting that the "flat" conformation of Cu_2:MBPC-1_2 likely is the preferred geometry.

The average Cu-N His distance in Cu_2:MBPC-1_2 is 2.08(3) Å (Table 3.5), which compares well with the distances observed in the copper(II) tetrakis(*N*-methylimidazole)·$2H_2O$ [2.02(3) Å].[15] As expected from a Cu^{2+} center, the axial aquo ligand in Cu_2:MBPC-1_2 is subject to Jahn-Teller distortion and positioned at a distance of 2.55(4) Å. The N-metal-N angles formed between His residues at cis positions are 90(2)° and those between the His residues at trans positions are 178(4)° and 180(1)°, indicating near-ideal square-planar protein-metal coordination environments. The square-planar copper coordination environment is further corroborated by its axial electron paramagnetic resonance (EPR) spectrum (S = 1/2), which is modeled well with four equivalent coordinating N ions (Figure 3.7). Similar to the model complexes, the histidine imidazole groups in this structure adopt a staggered arrangement, where the imidazole planes of *cis*-His residues are ~90° to one another and those of *trans*-His groups are nearly coplanar.

Crystallization of a Ni_2:MBPC-1_3 trimer. As both Ni^{2+} and Cu^{2+} induce the formation of a species in solution consistent with a MBPC-1 dimer, and given the fact that the crystal structure of the Cu-mediated assembly was shown to verify this fact, it is reasonable to assume that the resulting structure formed on crystallization of MBPC-1 with $NiSO_4$ would yield a similar C_2 assembly. The resulting structure (PDB ID: 3DE9)[13], is actually very different. Here, we find a parallel trimer of MBPC-1 molecules held together by two Ni ions (Figure 3.8 A and B), each coordinated octahedrally by

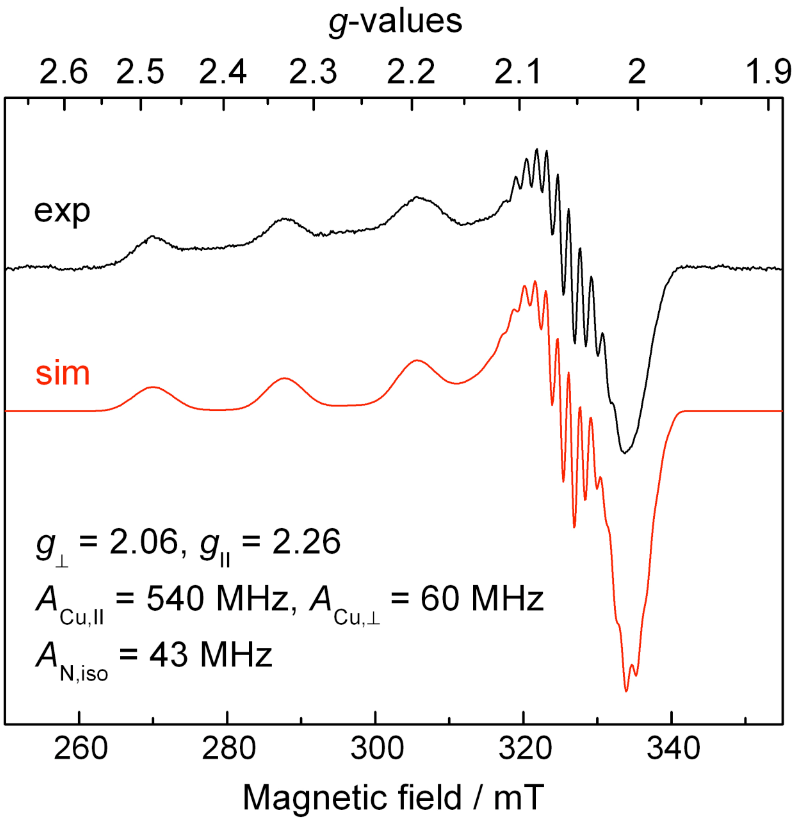


Figure 3.7. EPR spectra for Cu_2:MBPC-1_2. Experimental (black) and simulated (red) X-band EPR spectra for Cu_2:MBPC-1_2 and the parameters used for the simulation. The sample contained 1.5-fold molar excess of MBPC-1 over copper (150 vs 100 μM) to ensure that there was no free copper in solution. The data were collected at 125 K.

three di-His motifs (Figure 3.8 C). Both Ni ions are located on a crystallographic 3-fold symmetry axis in the rhombohedral crystal lattice (R_3 space group), whereby the monomeric components of Ni_2:MBPC-1_3 are interrelated by perfect 3-fold rotational symmetry (C_3; Figure 3.8 B). As in the copper induced dimer, protein surface interactions within Ni_2:MBPC-1_3 are minimal and non-specific (Table 3.4), burying only ~650 Å^2 between monomers.

Ni-N_{His} distances 2.18(6) compare well with the distances observed in the nickel(II) hexaimidazole [2.13(3) Å] complex (Table 3.5).[16] The N-metal-N angles formed between His residues at cis positions are 91(3)°, and those between the His residues at trans positions are 180(1)°, indicating near-ideal octahedral protein-metal coordination environments. Similar to the model complexes, the histidine imidazole groups in this structure adopt a staggered arrangement, where the imidazole planes of *cis*-His residues are ~90° to one another and those of *trans*-His groups are nearly coplanar.

An overlay of the Cu_2:MBPC-1_2 and Ni_2:MBPC-1_3 structures indicates that there is minimal difference between the hinge angles formed between individual MBPC-1 molecules in these structures (119.4° vs 120.0°; Figure 3.8 D). On the basis of this similarity, we suggest that the dimeric species formed at intermediate nickel-protein concentrations is also a flat structure resembling Cu_2:MBPC-1_2, bearing two octahedral Ni-$His_4(H_2O)_2$ centers. Upon an increase in the MBPC-1 concentration to millimolar levels under crystallization conditions, a third MBPC-1 molecule presumably coordinates the Ni ions to yield the observed trimeric structure, which would not be favored in the case of Cu^{2+} given its preference for four-coordinate geometry. It is important to note,

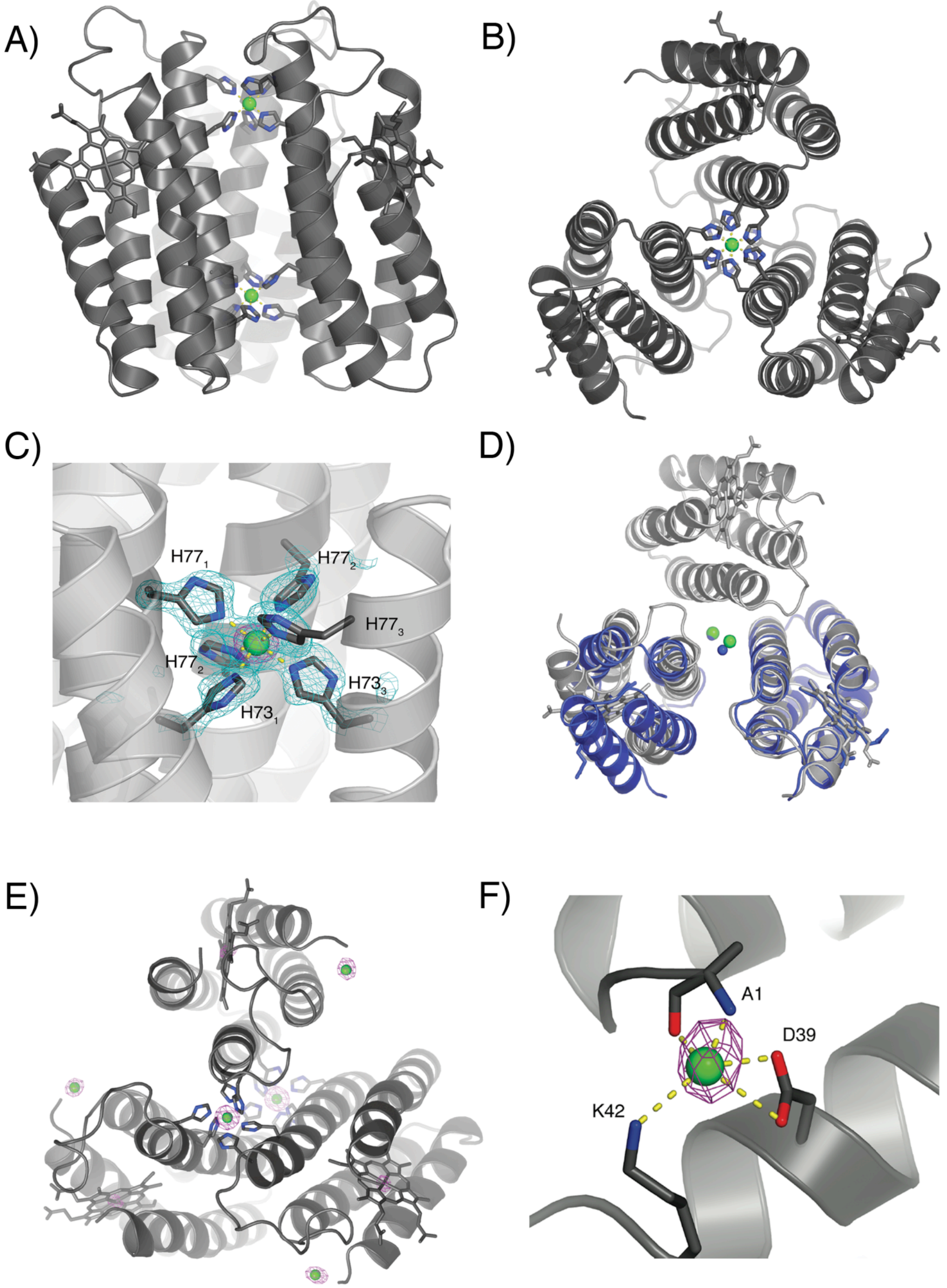


Figure 3.8. Crystal structure of Ni_2:MBPC-1_3. Side (A) and top (B) view of the C_3 symmetric Ni-mediated trimer of MBPC-1. C) Ni coordination environment in Ni_2:MBPC-1_3 and the corresponding simulated-annealing Fo - Fc omit electron density maps (cyan, 3.2σ; magenta, 8σ). D) Overlay of Cu_2:MBPC-1_2 (blue) and Ni_2:MBPC-1_3 (gray) based on Cα's of a single monomeric unit. Cu and Ni ions are shown as blue and green spheres. The hinge angles between the monomeric units in both complexes were calculated as $\angle$[center of mass (COM) of monomer A-COM of metal ions-COM of monomer B]. E) Anomalous difference density map (12σ) calculated using 1.485-Å radiation (Ni K edge) identifying the locations of Ni ions associated with Ni_2:MBPC-1_3. F) The coordination environment of the Ni ion associated with the N-terminus of MBPC-1. The bond distances are 2.03 Å (Ni-NAla1), 2.00 Å (Ni-OAla1), 2.64 Å (Ni-OD1Asp39), 3.08 Å (Ni-OD2Asp39), and 3.10 Å (Ni-NZLys42).

however, that Ni ions are immediately involved in crystal packing (Figure 3.8 E and F), and therefore it is likely that formation of Ni_2:MBPC-1_3 is also favored through lattice interactions. In any case, both the C_3 trimer observed in the solid state and the dimeric, and by inference, C_2 symmetrical form observed in solution are fully compatible with the octahedral coordination preference of Ni^{2+}.

Crystallization of Pd_2:MBPC-1_2. Pd^{2+} serves as a special case in the study of metal-mediated PPIs as its ability to form exchange inert complexes means any assembly formed will most likely be a kinetic, rather than thermodynamic, product. As seen in Figure 3.9 A, the addition of Pd^{2+} to MBPC-1 is no exception. Here, incubation of the protein with various ratios of $Pd(NO_3)_2$ results in full conversion to higher molecular weight species as resolved by native poly-acrylamide gel electrophoresis (PAGE). As native gels are run with fully folded protein that may not proceed through the crosslinked acrylamide in an ideal fashion, deciphering the higher order bands can be somewhat difficult. Nonetheless, these bands likely correspond to either dimeric and trimeric species, or two morphologically different dimers that travel at different rates through the gel. Subsequent examination of a sample of MBPC-1 incubated with 1:1 Pd^{2+} for 22 hrs by SV (Figure 3.9 B) shows that the main species does in fact appear to be dimeric, although a shoulder corresponding to a trimer and a faint peak corresponding to a tetramer can be seen. The ability of the Pd^{2+}:MBPC-1 complex to remain intact while running through a PAGE gel, and similarly through an SEC column (not shown), is unique from all three of the assemblies examined thus far as it is the only one kinetically stable enough to be detected by these methods.

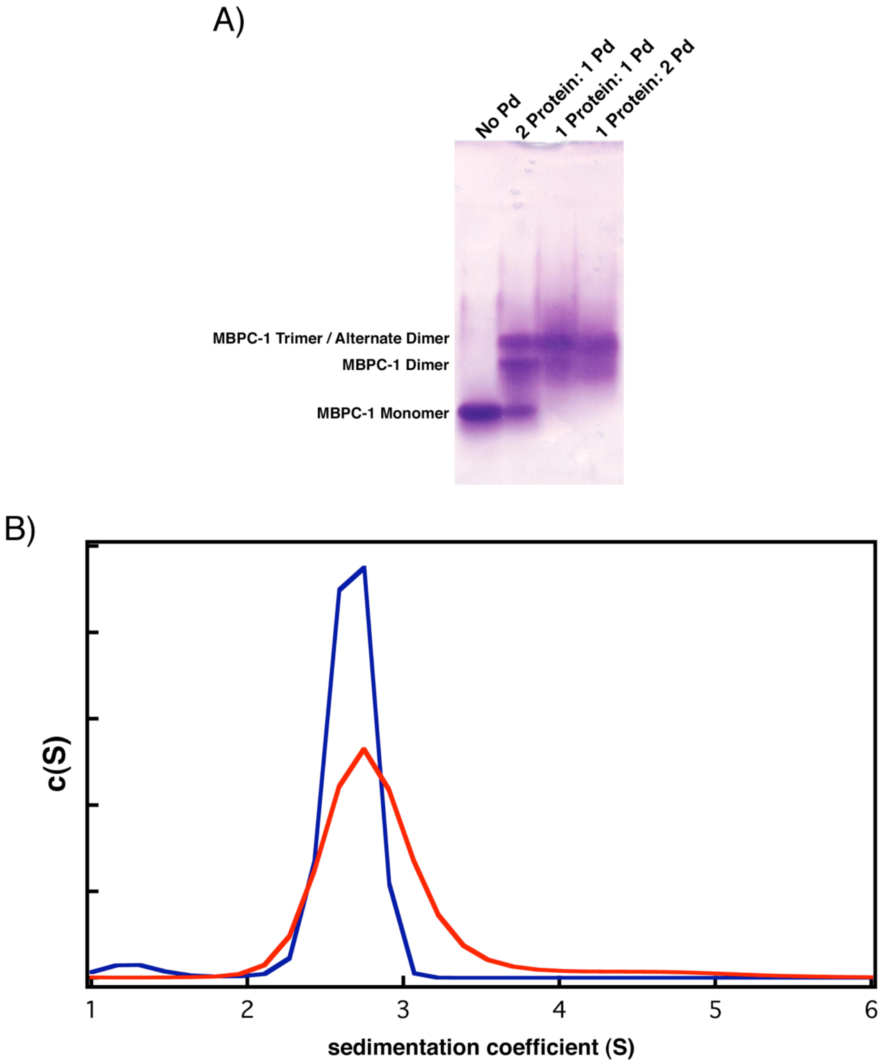


Figure 3.9. Solution studies of MBPC-1 incubated with Pd^{2+}. A) Samples of 200 μM MBPC-1 incubated with Pd^{2+} at 0, 2:1, 1:1, and 1:2 protein: metal ratios for 22 hrs. are run through a native 15% polyacrylamide gel, separating into at least three different species: monomeric; dimeric; trimeric/ alternate dimer conformation. B) An SV experiment run on a sample of 200 μM MBPC-1 incubated with 1:1 Pd^{2+} for 22 hrs. (red) shows the formation of mostly dimeric species similar to an analogous sample containing 1:1 Cu^{2+} (blue), though a possible trimeric shoulder and small tetrameric peak can be seen as well.

Despite the heterogeneity witnessed in solution, we set out to crystallize a Pd-mediated assembly of MBPC-1. Here we assume the structure elucidated will be the thermodynamically favored form, as multiple kinetic products would not result in a crystalline lattice. Thus, we believed that, due to the fact that Pd^{2+} almost exclusively forms four-coordinate square-planar complexes, we would uncover a dimer identical in structure to Cu_2:MBPC-1_2, but lacking the fifth axial water ligand. As shown in Figure 3.10 A and B, we see that a Pd_2:MBPC-1_2 dimer is indeed what we find. Here, the assembly has a C_2 symmetry axis that bisects the Pd-Pd axis, akin to that seen in the Cu complex, resulting in an anti-parallel arrangement of two monomers along helix3. This arrangement again buries a small interface (~700 Å^2) with limited favorable side chain interactions (Table 3.4). A superposition of the Pd_2:MBPC-1_2 and Cu_2:MBPC-1_2 dimers reveals that the relative orientations of the monomers in these oligomers are nearly identical (root-mean-square deviation over all Cα's = 0.733 Å; Figure 3.10 D), suggesting this organization is the most favorable, as was the case for the Cu-mediated dimer.

The Pd ions themselves are found in a square-planar coordination sphere, comprised of one bis-His motif from each monomer (Figure 3.10 C). The average Pd-N His distance in Pd_2:MBPC-1_2 is 2.04(11) Å, while The N-metal-N angles formed between His residues at cis positions are 90(4)° and those between the His residues at trans positions are 175(3)° and 172(2)°, indicating near-ideal square-planar protein-metal coordination environments (Table 3.5). Although a Pd(II)tetrakis(imidazole) does not appear in the CSD, a similar Pd(II) bis-histimino dichloride complex does contain

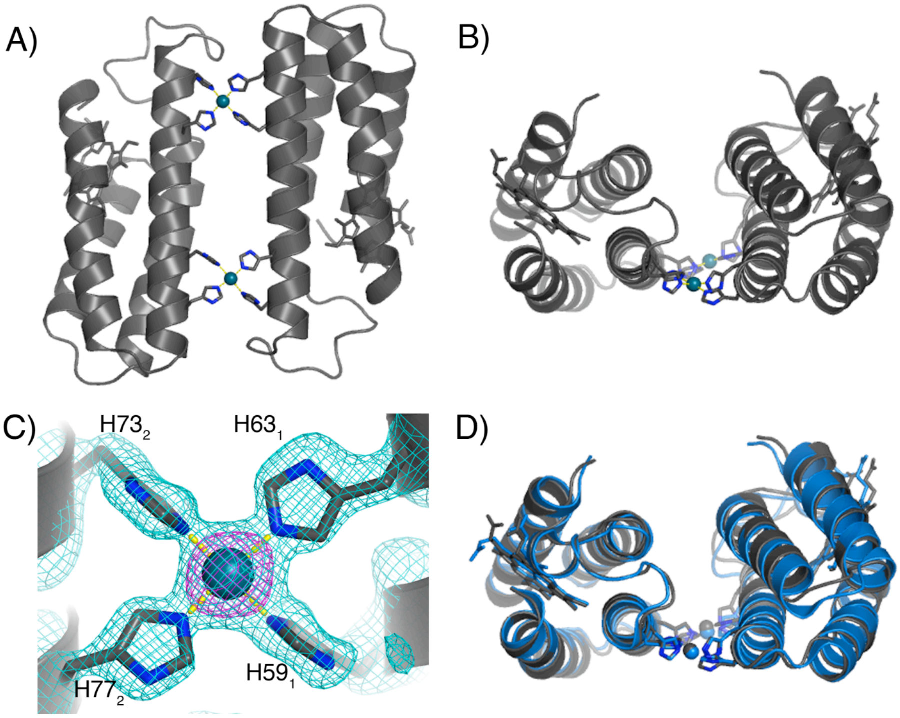


Figure 3.10. Crystal structure of Pd$_2$:MBPC-1$_2$. The full asymmetric unit contains one anti-parallel aligned dimer, shown from the side (A) and top (B). C) A representative Pd coordination environment in Pd$_2$:MBPC-1$_2$ and the corresponding simulated-annealing Fo - Fc omit electron density maps (cyan, 2σ; magenta, 8σ). E) Overlay of the Pd$_2$:MBPC-1$_2$ and Cu$_2$:MBPC-1$_2$ assemblies.

equivalent Pd-N His bond distances, square-planar arrangement, and a planar orientation of the trans imidazole rings observed in Pd_2:MBPC-1_2.[17]

Conclusions

From solution state studies there appears to be a large difference in the ability of MBPC-1 to coordinate divalent Ni and Cu compared to Zn. Nevertheless, these studies, in conjunction with the crystal structures elucidated, make it clear that the distinct MBPC-1 oligomerization geometries obtained with Cu (C_2), Ni (C_3 and/or C_2), Pd (C_2) and Zn (D_2) indicate that the supramolecular arrangement of this non-self-associating protein can be controlled by the metal coordination geometry, using principles commonly applied for the self-assembly of small molecules. Importantly, the facile access to different symmetries through metal coordination without the need to engineer large molecular surfaces may open up the path for the construction of multi-dimensional protein architectures, which require building blocks that simultaneously utilize a combination of these symmetry elements.

On the basis of the observations for MBPC-1, it is tempting to suggest that any protein with metal-chelating motifs on the surface can, in principle, be treated as a large polydentate ligand, whose supramolecular arrangement can be predicted by simple coordination chemistry rules. Yet, proteins possess large, topologically complex surfaces with many functional groups, which can not only coordinate metals but also interact with one another attractively or repulsively. Thus, while these concepts can aid in future assembly designs, the exclusive population of desired super-protein architectures through metal coordination will undoubtedly require a thorough consideration of non-covalent interactions as well as the precise localization of metal coordination.

Chapter 3 is reproduced in part with permission from: Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., Tezcan, F. A. **2009**. Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical assemblies through Cu^{II} and Ni^{II} coordination. *Inorg. Chem.* 48, 2726-2728. Copyright 2009 American Chemical Society.

References

1. Rulisek, L., Vondrasek, J. **1998**. Coordination geometries of selected transition metal ions (Co^{2+}, Ni^{2+},Cu^{2+}, Zn^{2+}, Cd^{2+}, and Hg^{2+}) in metalloproteins. *J. Inorg. Biochem.* 71, 115-127.

2. Sundberg, R. J., Martin, R. B. **1974**. Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. *Chem. Rev.* 74, 471-517.

3. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A., Schuck, P. **2004**. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Analytical Biochemistry* 326, 234-256.

4. Otwinowski, Z., Minor, W., *Processing of x-ray diffraction data collected in oscillation mode*. 1997, Vol. 276: Macromolecular Crystallography, part A, p 307-326.

5. Leslie, A. G. W. **1992**. Recent changes to the MOSFLM package for processing film and image plate data *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*. 26,

6. Murshudov, G. N., Vagin, A. A., Dodson, E. J. **1994**. The CCP4 Suite: Programs for protein crystallography. *Acta Cryst. D*. 50, 760-763.

7. Vagin, A., Teplyakov, A. **1997**. MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.

8. Murshudov, G. N., Vagin, A. A., Dodson, E. J. **1997**. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.D.* 53, 240-255.

9. McRee, D. E. **1992**. A visual protein crystallographic software system for X11/XView. *J. Mol. Graphics*. 44-46.

10. Laskowski, R. A., Macarthur, M. W., Moss, D. S., Thornton, J. M. **1993**. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291.

11. DeLano, W. L. *The PYMOL molecular graphics system (http://www.pymol.org)*, 2003.

12. Krissinel, E., Henrick, K. **2007**. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 372, 774-797.

13. Neese, F. **1995**. *QCPE Bull.* 15, 5.

14. Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., Tezcan, F. A. **2009**. Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical assemblies through Cu^{II} and Ni^{II} coordination. *Inorganic Chemistry*. 48, 2726-2728.

15. Su, C.-C., Hwang, K.-Y., Jan Hua, C., Wang, S.-L., Liao, F.-L., Horng, J.-C. **1995**. Bonding properties of copper(II)-imidazole chromophores: structures and electronic properties of tetrakis imidazole copper(II) complexes. Molecular structures of Cu(N-methylimidazole)$_4$(ClO_4)$_2$ and [Cu(N-methylimidazole)$_4$(H_2O)$_2$]Cl_2(H_2O). *Polyhedron*. 14, 3011-3021.

16. Konopelski, J. P., Reimann, C. W., Hubbard, C. R., Mighell, A. D., Santoro, A. **1976**. Hexakis(imidazole)nickel(II) Chloride Tetrahydrate. 2911-2913

17. Dahan, F. **1976**. The crystal structure of bis(histamino)palladium dichloride, ($N_3C_5H_92PdCl_2$. *Acta. Cryst. B*. 32, 2472-2475.

Chapter 4

The Influence of Secondary Interactions on Metal-Directed Protein Self-Assembly

Introduction

The exact nature of the forces that control naturally occurring PPI interfaces are not fully understood even today. A basic tenet has been that these interfaces are driven by the burial of a high number of hydrophobic interactions along extended surfaces with a high degree of surface complementarity.[1,2] However, surveys of large numbers of oligomeric assemblies deposited in the Protein Data Bank (PDB) reveal that these hydrophobic residues can just as often form small, localized patches instead of a singular surface,[3] often-times appearing closer in similarity to the rest of the protein surface than a hydrophobic core. In fact, charged residues have been found to be fairly well represented along PPI interfaces and are proposed to help in the formation of hydrogen bonding networks that strengthen the interaction,[4] as well as provide directional guidance of proper complexation.[5]

While rational design and control of PPIs would provide access to novel protein assemblies[6] as well as the manipulation of cellular processes,[7,8] such efforts are hampered by the these extensive and poorly understood non-covalent bonds. With MDPSA, we sought to circumvent this complication by controlling the self-assembly of proteins by metal binding, akin to work by several groups that have shown that small organic building blocks with acceptor groups can self-assemble into discrete super-structures and frameworks through metal coordination.[9,10] We were aware of the fact, however, that these protein based ligands came with the added complications that (a) protein surfaces are replete with polar side chains capable of coordinating metals, and (b) the interactions between individual proteins—in contrast to those between organic building blocks—may not be negligible.

MBPC-1, which was engineered with two di-His motifs (59/63 and 73/77) located near each terminus of a helix3, is capable of forming a Zn^{2+} mediated tetrameric assembly (Zn_4:MBPC-1_4).[11] The crystal structure of Zn_4:MBPC-1_4 revealed a unique quaternary architecture, in which two V-shaped dimers wedged into one another, held together by four Zn ions with identical His_3(63/73/77)-Asp_1(74) coordination environments. The key to this supra-molecular arrangement was the Asp74 residue located within the His73/77 clamp, which allowed the V-shaped dimers and, ultimately, the observed tetramer to be formed with an extensive PPI surface area of nearly 5000 $Å^2$. This raises the possibility that secondary interactions between proteins may still influence the formation of the observed tetramer. If, on the other hand, these interactions had negligible effect and metal coordination was the sole determinant of the supra-molecular geometry, then the whole oligomeric assembly could be "inverted" simply by moving the coordinating Asp residue from within the His73/77 motif at the C-terminal end of helix3 to an analogous position at position 62 inside the N-terminal His59/63 bis-His clamp (Figure 4.1).

The resulting mutant, MBPC-2, was indeed found to form an "inverted" V, wherein the vertex is at the N-terminal end of helix3 rather than the C-terminal end. However, the presence of an Arg-Asp salt bridging pair along the interaction interface appears to force this variant to adopt a much different Zn coordination sphere. On further study, this salt bridge and an analogous pair in Zn_4:MBPC-1_4 were found to be crucial for specific formation of the Zn-mediated tetramers over non-specific aggregate assemblies, clearly pointing to the importance of secondary interactions to metal-mediated PPIs.

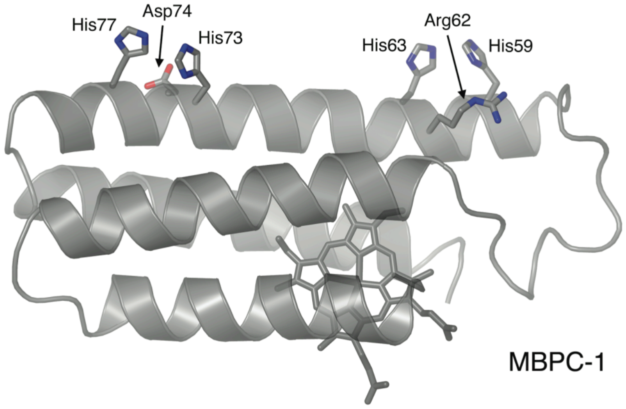


Figure 4.1. Structure of a MBPC-1 monomer. Based on the monomeric building blocks of the Zn_4:MBPC-1_4 assembly; PDB ID: 2QLA), demonstrating the analogous positioning of Arg62 at the N-terminal end of helix3 to Asp74 at the C-terminal end.

Materials and Methods

Site Directed Mutagenesis and Protein Expression/Purification/Characterization. Site directed mutagenesis was performed on the pETc-b562 plasmid (denoted as wildtype), as above, with sequencing performed by UCSD Moores Cancer Center. Mutant plasmids were transformed into BL21(DE3) *E. coli* cells along with the *ccm* heme maturation gene cassette plasmid, pEC86. Expression and purification of the protein was performed as previously described.

Sedimentation Velocity. SV experiments were performed in order to determine the solution-state oligomerization behavior of MBPC-2. All SV experiments were performed in 20 mM TRIS (pH 7) using appropriate volumes of 5 mM $ZnCl_2$ in 20 mM TRIS (pH 7) to yield a 1:1 Zn^{2+}: protein concentration ratio (50, 200, 400, and 600 μM protein). Measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor at 47,000 rpm for a total of 180 scans per sample. The following wavelengths were used for detection: 570 nm (50 μM protein), 582 nm (200 μM protein), 620 nm (400 μM protein), and 670 nm (600 μM protein). All data were processed as previously described with the following parameters fixed: buffer density (ρ) = 0.99764 g/ml; buffer viscosity = 0.0089485 poise; Vbar-MBPC-2 = 0.732 g/ml; Vbar- (R34K)MBPC-2 = 0.7329 g/ml; Vbar-(R34D)MBPC-2 = 0.7311 g/ml; Vbar-(R34K)MBPC-1 = 0.732 g/ml; Vbar-(R34D)MBPC-2 = 0.7329 g/ml.

Sedimentation Equilibrium. In order to determine the apparent K_d values of the oligomeric states of MBPC-2, SE experiments were undertaken. All SE experiments were carried out in 20 mM TRIS (pH 7) using appropriate volumes of 5 mM $ZnCl_2$ in 20 mM

TRIS (pH 7) to yield a 1:1 Zn^{2+}: protein concentration ratio (50, 200, 400, and 600 μM). Measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor at speeds between 10,000 and 30,000 rpm. Scans were taken at 14 and 16 hrs, by which sedimentation equilibrium was determined to be reached by visual inspection. The following wavelengths were used for the detection: 572 nm (50 μM protein), 600 nm (200 μM protein), 675 nm (400 μM protein), and 700nm (600 μM protein).

The 16-hr scans were used for global analysis as described previously with the menisci assigned based on 45,000 rpm meniscus depletion scans and were fixed as constants, as was the 12242 Da molecular mass of the monomeric protein. In fitting the data to a monomer-dimer-tetramer model, we used the total amount of protein loaded for each experiment, the association constants, and the bottom of the cell as floating parameters. A minimized fit was attained once rigorous F-statistics analyses were performed on individual floating parameters. Standard deviation for the resulting $\log_{10}(K)$ values and incompetent fractions were determined through Monte-Carlo analysis within SEDPHAT.[12] Resulting fit statistics, log K_a values, incompetent fraction, and K_d values are shown in table 4.1.

Crystallography. All crystals were obtained by sitting drop vapor diffusion. Zn_4:MBPC-2_4 was crystallized at 4°C with a precipitant solution consisting of 100 mM TRIS (pH 7.5), 200 mM $CaCl_2$, 12% PEG 400, and 1.25 mM $ZnCl_2$. The drop consisted of 2 μL protein (1.595 mM in 20 mM TRIS pH 7) and 1 μL precipitation solution. Crystals appeared after about 2 hrs, reaching a maximum size of ~ 400 μm x 100 μm x 50 μm after a week. Zn_4:(R34K)MBPC-2_4 was crystallized in a manner similar to

Table 4.1. Results of global fitting of SE data for MBPC-2 in the presence of Zn^{2+}. The chi-squared and rmsd values of the globally fit SE data for MBPC-2 with the metal ion are given, along with resulting association and dissociation constants, as well as incompetent fraction. These data correspond to the data shown in Figures 4.3.

Zn^{2+}	**χ^2 of Global Fit** 1.825776	**Fit rmsd** 0.010153
	Log$_{10}$(K_a), Monomer-Dimer 11.87674 (±2.647087 x 10^{-1})	**Monomer-Dimer K_d** 1.328 x 10^{-12} M
	Log$_{10}$(K_a), Dimer-Tetramer 27.44355 (±5.271404 x 10^{-1})	**Dimer-TetramerK_d** 7.115x 10^{-10} M
	Incompetent Fraction 0.04121 (± 1.217359 x 10^{-3})	

Zn_4:MBPC-2_4 at room temperature. The precipitation solution was 100 mM TRIS (pH7.5), 100 mM $CaCl_2$, 11% PEG 400, and 2 mM $ZnCl_2$. Crystals appeared within two days and grew to ~ 400 μm × 400 μm × 200 μm. The crystals to be used for diffraction experiments were exchanged into a solution containing 20% glycerol as a cryoprotectant and frozen in liquid nitrogen or directly in the cryostream. X-ray diffraction data were collected at 100 K using a Bruker Apex II CCD detector and monochromatized Cu-Kα radiation (1.54 Å) produced by a Siemens sealed source. The data were processed using SAINT and Bruker SADABS. The structures of Zn_4:MBPC-2_4 and Zn_4:(R34K)MBPC-2_4 were determined at 1.87 and 1.75-Å resolution, respectively, by molecular replacement (MOLREP[13,14]) using the cyt c*b*$_{562}$ structure (PDB ID 2BC5) as the search model. Rigid-body, simulated-annealing, positional and thermal refinement with CNS,[15] along with manual rebuilding, and water placement with XFIT,[16] produced the final model. The Ramachandran plot was calculated with PROCHECK.[17] All figures were produced with PYMOL.[18]

In addition to four zinc ions in the asymmetric unit for both assemblies, we found five calcium ions that appear to stabilize CPIs. The identities of calcium ions were deduced from their coordination environments, which are entirely composed of acidic side chains and water molecules. The presence of calcium did not alter the Zn-dependent sedimentation behavior of any variants described in this study, indicating that the calcium ions are solely involved in crystal packing and not protein oligomerization.

Table 4.2. X-ray data collection and refinement statistics for Zn$_4$:MBPC-2$_4$ and Zn$_4$:(R34K)MBPC-2$_4$.
[‡]Rsym= $\Sigma\Sigma_j|I_j - \langle I \rangle| / \Sigma\Sigma_j|I_j|$.
[§]R= $\Sigma||F_{obs}| - |F_{calc}|| / \Sigma|F_{obs}|$.
[II]Free R calculated against 7 and 6.6% of the reflections removed at random for the MBPC-2 and R34KMBPC-2 structures, respectively.
[¶]Root mean square deviations from bond and angle restraints.

	Zn$_4$:MBPC-2$_4$	**Zn$_4$:(R34K)MBPC-2$_4$**
Residues in complex	4 x (106 + 1 Heme + 1 Zn)	4 x (106 + 1 Heme + 1 Zn)
No. of tetramers / asymmetric unit	1	1
Waters in asymmetric unit	506	545
Calcium ions in asymmetric unit	5	5
Unit cell dimensions (Å)	47.97 x 90.13 x 49.36	47.97 x 89.90 x 49.67
	$\alpha = \gamma = 90^\circ$, b = 109.6°	$\alpha = \gamma = 90^\circ$, b = 110.3°
Symmetry group	$P2_1$	$P2_1$
Resolution (Å)	25 - 1.87	20 - 1.75
X-ray wavelength (Å)	1.54	1,54
Number of Unique Reflections	36614	37193
Redundancy	3.5	2.8
Completeness (%)*	94.6 (99.3)	93.3 (82.1)
⟨I / sI⟩*	6.95 (1.6)	13.5 (2.0)
Rsym[‡] (%)*	16.8 (63)	7.6 (51)
R[§] (%)*	23.1 (23.1)	19.9 (27.4)
Free R[II] (%)*	27.3 (31.2)	24.7 (32.6)
Rms Bnd[¶] (Å)	0.011	0.012
Rms Ang[¶] (°)	1.16	1.23
Ramachandran plot (%)		
Residues in most favored regions	94.8	97.9
Residues in add.l allowed regions	5.2	2.1
Residues in generously allowed regions	0.0	0.0
Residues in disallowed regions	0.0	0.0

Table 4.3. Interfacial hydrogen bonding interactions in Zn$_4$:MBPC-1$_4$, Zn$_4$:MBPC-2$_4$ and Zn$_4$:(R34K)MBPC-1$_4$ complexes, determined using the PISA server (Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)).[19] The interactions of residues involved in interfacial salt-bridges in each individual assembly are highlighted in red.

Zn$_4$:MBPC2$_4$

Interface between chains	Residue/Atom	Dist (Å)	Residue/Atom
A-B	A:ASN 99[OD1]	3.80	B:LYS 83[NZ]
	A:GLN 103[OE1]	2.65	B:LYS 83[NZ]
B-C	B:GLN 41[O]	2.95	C:ARG 34[NH1]
	B:ASP 62[OD1}	3.15	C:ARG 34[NH1}
	B:ARG 34[NH1]	2.92	C:GLN 41[O]
A-D	A:ASP 62[OD2]	3.19	D:ARG 34[NH1]
	A:GLN 41[O]	3.43	D:ARG 34[NH1]
	A:GLN 41[O]	3.48	D:ARG 34[NH2]
	A:ASP 54[OD2}	2.54	D:ASN 80[ND2]
	A:ARG 34[NH1]	2.90	D:GLN 41[O]
C-D	None		

Zn$_4$:(R34K)MBPC1$_4$

Interface between chains	Residue/Atom	Dist (Å)	Residue/Atom
A-B	A:ASN 99[OD1]	3.80	B:LYS 83[NZ]
	A:GLN 103[OE1]	2.65	B:LYS 83[NZ]
B-C	B:GLN 41[O]	2.95	C:ARG 34[NH1]
	B:ASP 62[OD1}	3.15	C:ARG 34[NH1}
	B:ARG 34[NH1]	2.92	C:GLN 41[O]
A-D	A:ASP 62[OD2]	3.19	D:ARG 34[NH1]
	A:GLN 41[O]	3.43	D:ARG 34[NH1]
	A:GLN 41[O]	3.48	D:ARG 34[NH2]
	A:ASP 54[OD2}	2.54	D:ASN 80[ND2]
	A:ARG 34[NH1]	2.90	D:GLN 41[O]
C-D	None		

Results and Discussion

Formation of a Zn-mediated MBPC-2 tetramer. In order to generate the MBPC-2 variant protein, we made two simple mutations (Asp74Ala, Arg62Asp) to remove the coordinating Asp between the 73/77H bis-His clamp and place it between the 59/63H clamp (Figure 4.1). Analysis of the resulting protein with the addition of equimolar Zn via SV demonstrates that MBPC-2 does indeed still form both metal-mediated dimeric and tetrameric assemblies (Figure 4.2). What is more, this mutant appears to form these oligomeric species more readily than MBPC-1. This is apparent when one considers that, at 50 μM protein and Zn, roughly 50% of MBPC-2 forms the tetramer, whereas a similar population size was only reached with MBPC-1 when concentrations were increased to 600 μM (Figure 4.2 inset).

Given that there is a clear increased favorability in the formation of the Zn:MBPC-2 oligomers, it then follows that the K_d of formation for both the dimeric and tetrameric species will have decreased. In order to test this idea SE was once again employed (Figure 4.3). When fitting these data the inclusion of an incompetent fraction was again found to be necessary. Unlike MBPC-1, which had a 30% incompetent fraction, MBPC-2 was found to only have ~4% of the protein unavailable to the M-D-T equilibrium (Table 4.1), suggesting a difference in the ability for this protein to bind Zn.

Seemingly due to this increased ability to coordinate Zn, the resulting K_d for the formation of the MBPC-2 dimer is found to be 1.3 pM, while the D-T K_d is found to be 711 pM. If these values were exactly true, then MPBC-2 would essentially form an

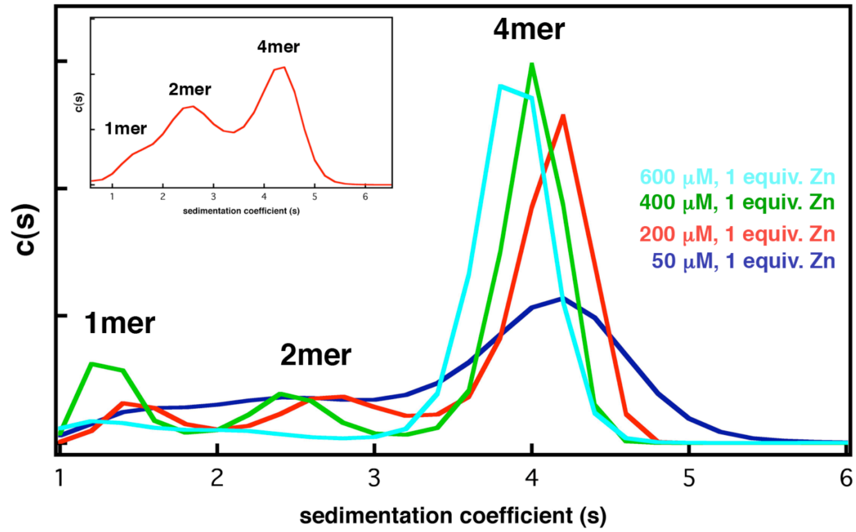


Figure 4.2. Sedimentation coefficient distributions of MBPC-2. Sedimentation coefficient distributions of 50 µM (blue), 200 µM (red), 400 µM (green), and 600 µM (cyan) MBPC-2 with equimolar Zn^{2+} are shown. The distributions are normalized with respect to the area covered under the curves. (Inset) 600µM MBPC-1 with equimolar Zn^{2+} is shown for comparison of the two mutant proteins.

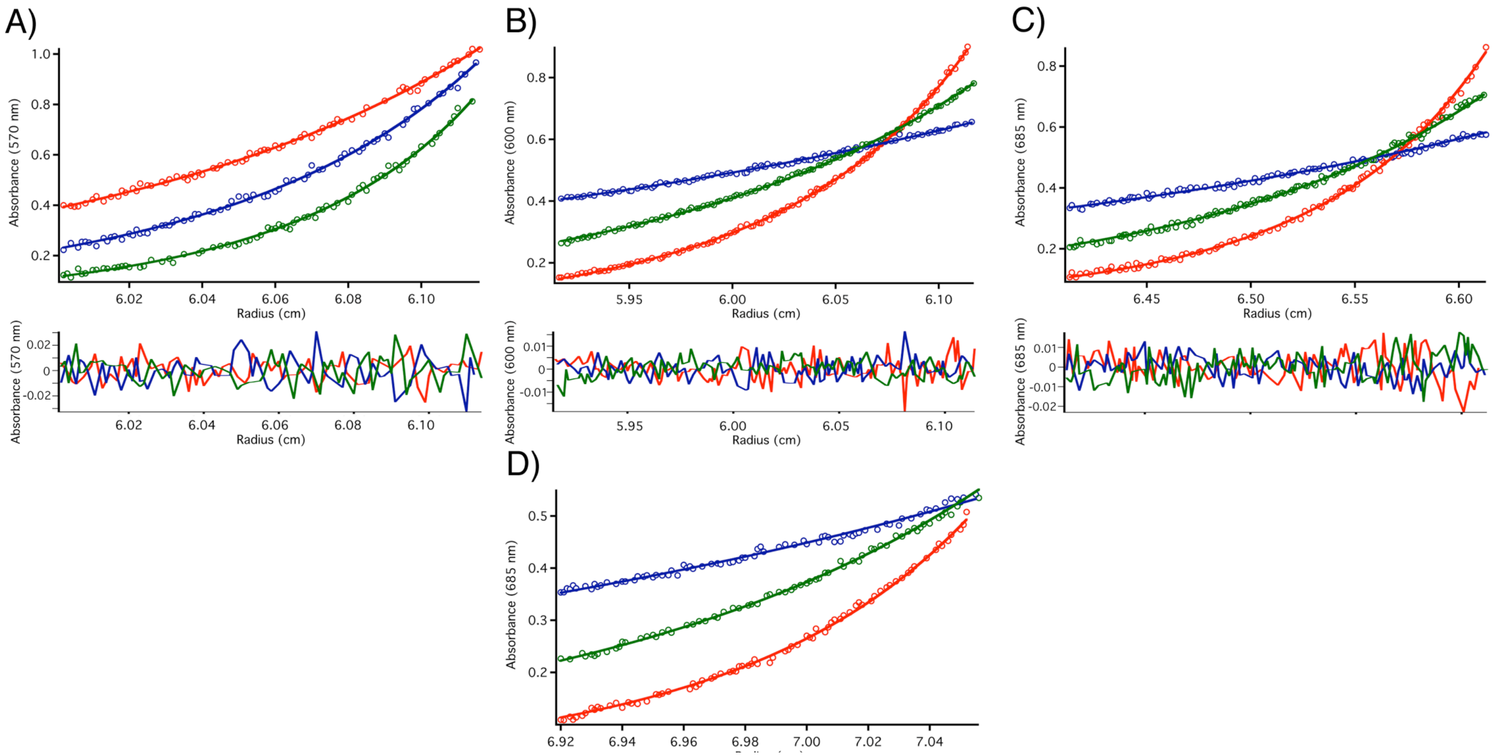


Figure 4.3. Sedimentation equilibrium data for MBPC-2 in the presence of Zn^{2+}. A) 50µM protein at 20,000, 25,000, and 30,000 rpm. B) 200µM protein at 10,000, 15,000, and 20,000 rpm. C) 400µM protein at 10,000, 15,000, and 20,000 rpm. D) 600µM protein at 10,000, 15,000, and 20,000 rpm. All samples were run in 20 mM TRIS (pH 7), at 25° C, with one molar equivalent of $ZnCl_2$. Scans were globally fit to a monomer-dimer-tetramer model yielding a minimized fit whose results and statistics are shown in Table 4.1.

obligate oligomer, (i.e. 100% of the protein would be in a higher-ordered assembly). As evidenced by the c(s) distributions, this is clearly not true, highlighting the limitations of this experimental procedure: varying the protein and Zn concentration while maintaining a constant 1:1 ratio of the two skews the overall results. This ratio, of course, must remain constant as too large of an excess of the metal can only result in an aggregation event, while an excess of protein would not fully populate the higher order species we wish to study. Therefore, these values must be taken as apparent values, not exact Kd's, and so only good for comparison of mutants or various metal-mediated assemblies of the same construct. Although this is true, it is obvious from these data that MBPC-2 does indeed form Zn-mediated oligomers more favorably than MBPC-1.

Comparison of the Zn_4:MBPC-2_4 and Zn_4:MBPC-2_4 tetramers. Although a Zn-mediated MBPC-2 tetramer is formed in solution as we had predicted, there remains the possibility that the resulting assembly is completely different from what we had initially hoped to engineer. Fortunately, Zn_4:MBPC-2_4 readily lent itself to crystallization, and we determined its structure at 1.9 Å resolution (PDB ID: 3C62).[20] The Zn_4:MBPC-2_4 architecture is similar to that of Zn_4:MBPC-1_4, with two interlaced V-shaped dimers that are slightly more open, an overall topology that is slightly flatter, and a nearly identical buried protein surface area (~5000 $Å^2$) (Figure 4.4). The Zn_4:MBPC-2_4 structure is indeed the "inverse" of Zn_4:MBPC-1_4 (Figure 4.4 A). Whereas the V-shapes are joined at the Helix3 C-termini in the latter, they are crosslinked at the N-terminus in the former; the transition between the two super-structures resembles a scissor motion. Surprisingly, Asp62 is not involved in Zn binding, as was the intention. Instead, each of the four Zn ions in the assembly are ligated by the His73/77 motif from one monomer, His59 from a

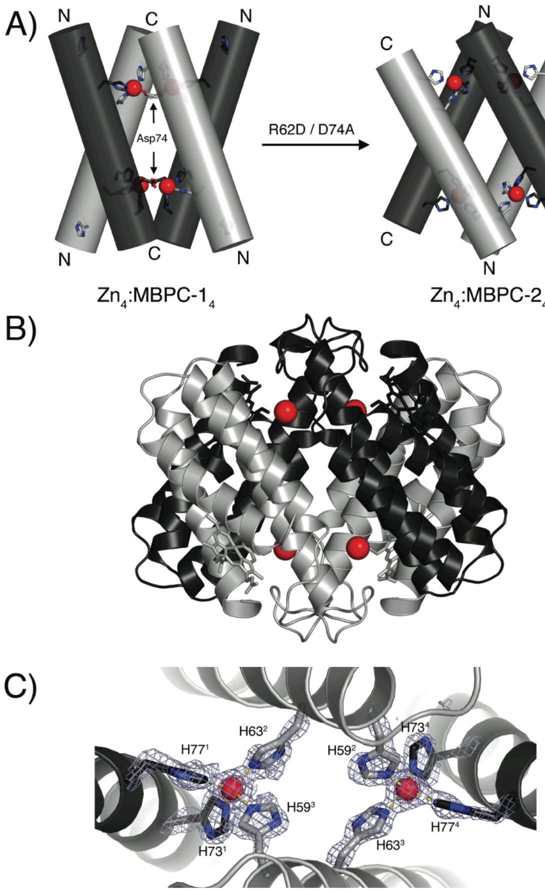


Figure 4.4. Crystal structure of Zn$_4$:MBPC-2$_4$. A) Cylindrical representations of Zn$_4$:MBPC-1$_4$ and Zn$_4$:MBPC-2$_4$ Helix3's and side chains involved in Zn coordination, viewed along a non-crystallographic twofold symmetry axis. Pairs of protein molecules that form the V-shaped dimers are colored alike. The N- and C-termini of the four Helix3's in each assembly are labeled accordingly. B) Ribbon representation of the Zn$_4$:MBPC-2$_4$ crystal structure. C) Close-up view of the Zn coordination environment down the non-crystallographic twofold symmetry axis and the corresponding simulated-annealing *Fo-Fc* omit electron density map (3.5σ).

second, and His63 from a third, yielding an ideal tetrahedral His^4-Zn coordination geometry with an average Zn-His bond distance of 2.05 Å (Figure 4.4 C). In this arrangement, each V-shape is stabilized by His59 and His63 coordination – instead of the expected Asp62 and His63 coordination – from two monomers, which splay apart to bind two Zn ions, thereby joining the Helix3 N-termini together. Interestingly, His^4-Zn coordination has rarely been observed in natural systems, one example being the HAP1 transcriptional factor that contains a structural $His^2(N\varepsilon)$,$His^2(N\delta)$-Zn site.[21]

With these two structures in hand, many questions are raised: Why does MBPC-2 not oligomerize through the His^3-Asp^1 Zn-coordination motif although it has access to it? Conversely, why does MBPC-1 not self-assemble through the same His^4 motif as MBPC-2? And finally, why is Zn_4:MBPC-2_4 more stable than Zn_4:MBPC-1_4 despite the fact that both assemblies are held together by four Zn ions in unstrained, tetrahedral coordination geometries? We thus decided to take a closer look at the protein interfaces in both assemblies to determine whether secondary interactions may influence metal induced self-assembly (Table 4.3).

As expected from a soluble protein whose surface is not optimized for self-association, the interfaces in both Zn_4:MBPC-1_4 and Zn_4:MBPC-2_4 structures feature overwhelmingly polar residues that are poorly packed. While the absolute stabilities of Zn_4:MBPC-1_4 and Zn_4:MBPC-2_4 and the energetic contributions of the interfacial residues are difficult to compute, a qualitative inspection of the interfaces in these assemblies reveal that there are fewer unfavorable and more favorable interactions in the latter compared to the former. Two such unfavorable interactions that stand out in Zn_4:MBPC-1_4 are: 1) the directly opposed positioning of two carboxylate residues from neighboring monomers (Asp60 and Glu86); and 2) the burial of Arg62 in the interface in an unfavorable geometry that places it in close proximity to Arg34. On the other hand, in Zn_4:MBPC-2_4, L38 side chains are more closely packed and buried in a much more favorable series of interactions.

Probing the salt bridges of the MBPC tetramers. In both the Zn_4:MBPC-1_4 and Zn_4:MBPC-2_4 tetramers there are surprisingly few interfacial H-bonding interactions given the extensive buried protein surfaces in both assemblies. Earlier surveys of oligomeric protein structures have indicated average interfacial H-bond densities ranging from 0.5 to 0.7 per 100 $Å^2$ of buried surface;[22] these values are considerably higher than 0.3 and 0.2 per 100 $Å^2$ observed in Zn_4:MBPC-1_4 and Zn_4:MBPC-2_4 complexes. Significantly, the majority of these H-bonding interactions involve two pairs of residues that form salt bridges: Arg34-Asp66 in Zn_4:MBPC-1_4 accounting for 9 out of a total of 15 interactions, and Arg34-Asp62 in Zn_4:MBPC-1_4 accounting for 7 out of a total of 10 (Figure 4.5).

In order to investigate the possible role of these salt bridges in the self-assembly of the two tetramers, we generated a series of Arg34 mutants of MBPC-1 and MBPC-2 aimed at weakening or abolishing these interactions. Figure 4.6 shows the oligomerization behavior of the Arg34 variants as determined by SV. As previously described, the higher relative stability of Zn_4:MBPC-2_4 can be seen in its nearly full population of the tetrameric form at 600 μM protein and Zn, whereas a significant population of dimeric species is still present with MBPC-1 under the same conditions. Upon the mutation of Arg34 to the shorter Lys, the peak for the tetrameric assembly disappears entirely for MBPC-1 (Figure 4.6 C, green line) and becomes broader for

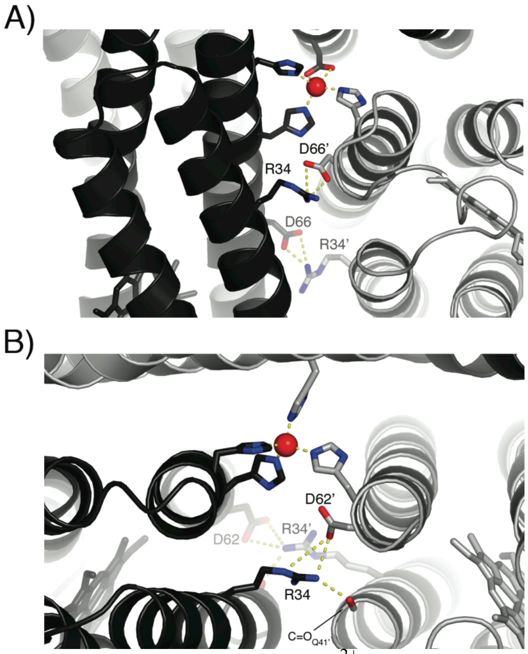


Figure 4.5. Hydrogen bonding interactions in MBPC Zn^{2+}-mediated tetramers. Interfacial H-bonding interactions in Zn_4:MBPC-1_4 (A) and Zn_4:MBPC-2_4 (B) assemblies that involve R34 and D66 in the former and R34 and D62 in the latter. The pairwise interactions shown above are repeated twice due to the internal twofold symmetry of each tetrameric assembly.

MBPC-2 (Figure 4.6 A, red line), consistent with the weakening of the salt-bridging interaction. The 1.75-Å crystal structure of the Zn_4:(R34K)MBPC-2_4 assembly (PDB ID: 3C63) reveals that Lys34 indeed maintains the salt bridge to Asp62 and stabilizes the parent tetrameric assembly, but its interactions are not as extensive compared to Arg34 (Figure 4.6 D, Table 4.3).

Dramatically, when Arg34 is mutated to Asp in MBPC-2 [(R34D)MBPC-2], a heterogeneous ensemble that contains higher order aggregates replaces the tetrameric form (Figure 4.6 B). Though the main peak in the c(s) distribution has a sedimentation coefficient of ~ 2 S, it also has an extremely large frictional ratio of 3.9. This value, which is a weight-averaged value calculated from fitting the SV data that describes the shape of the species in the sample, suggests that this peak corresponds to an extremely large protein aggregate. Similarly, when Arg34 is mutated to Ala in MBPC-1 [(R34A)MBPC-1], addition of equimolar Zn^{2+} causes the protein to fully aggregate out of solution. Shown in the red trace in Figure 4.6 C is a sample containing 600 μM protein and 300 μM Zn (a 2 protein:1 Zn ratio), which contains species at 5, 6, and 7 S, clearly describing the formation of this aggregate.

Conclusions

Although we were successful in the "inversion" of the Zn_4:MBPC-1_4 tetramer with the MBPC-2 mutant, we nonetheless discovered that the transformation was not identical to what we had envisioned. It is clear that, though minimal, the interfacial interactions between individual monomers in these metal-mediated assemblies are

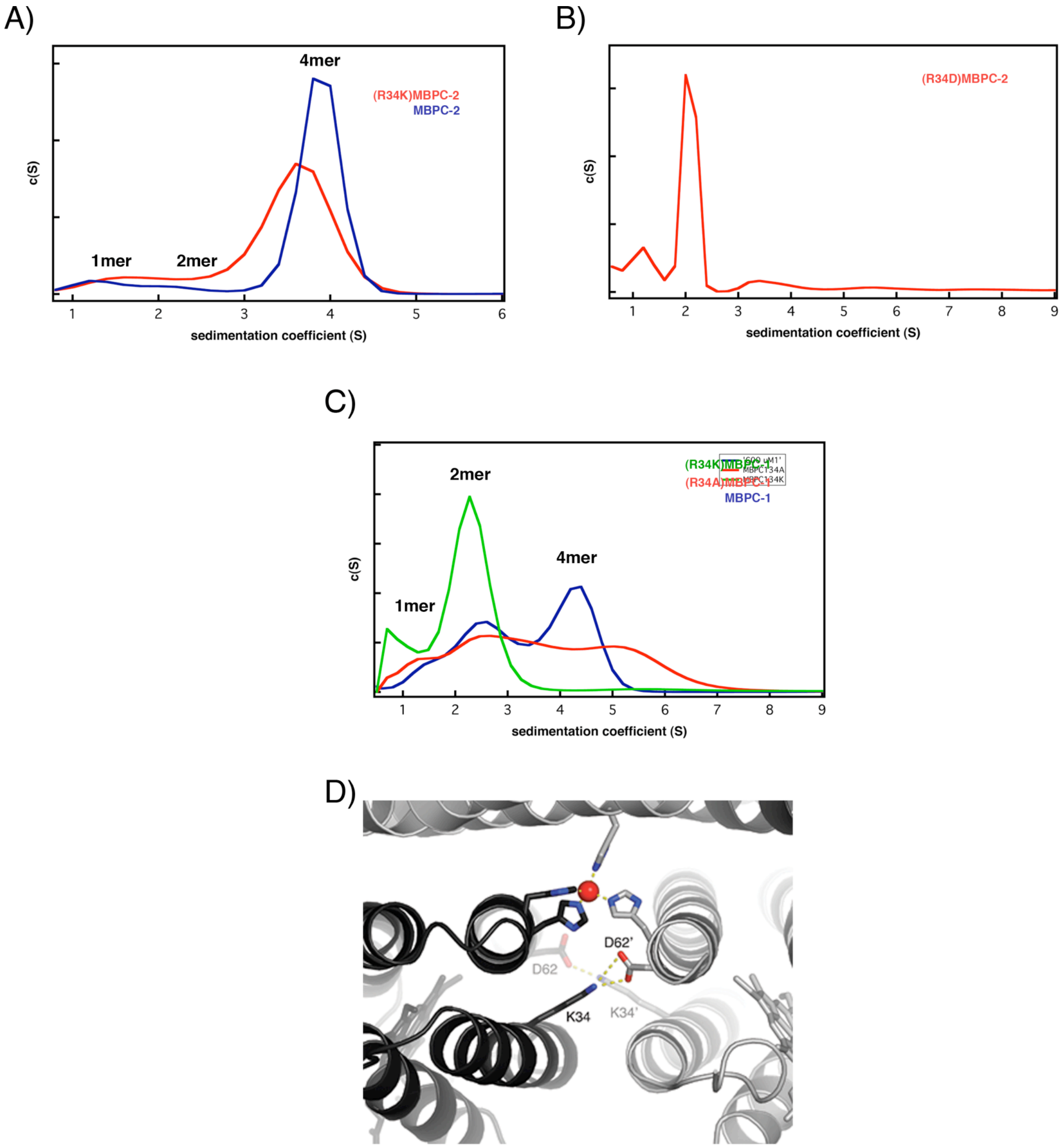


Figure 4.6. Arg34 mutants of MBPC-1 and MBPC-2. Sedimentation coefficient distributions of MBPC-2 (A and B) and MBPC-1 (C) Arg34 variants as determined by SV. All samples contain 600 µM protein and 600 µM Zn, with the exception of (R34A)MBPC-1 (red line in C), which contains 300 µM Zn. D) Interfacial salt-bridging interactions involving Lys34 and Asp62 residues in Zn_4:(R34K)MBPC-2_4. The pairwise interactions shown above are repeated twice due to the internal twofold symmetry of this tetrameric assembly.

important for determination of the final structure. The findings presented here demonstrate that the Arg34 interactions are clearly the guiding factor for the metal-induced oligomerization of MBPC-1 and 2. In the presence of limiting amounts of Zn, MBPC-1 or MBPC-2 molecules likely sample all of the conformations depicted in Figure 4.7, with the possible exception of **5** due to steric reasons. While monomeric and dimeric (**3**) forms dominate at low protein concentrations, the tetrameric, interlaced conformation **4**, with the highest Zn-binding cooperativity, is expected to be the prevalent species at intermediate concentrations. This is indeed borne out by the structural convergence of Zn_4:MBPC-1_4 and Zn_4:MBPC-2_4 assemblies despite differences in their metal coordination modes. Secondary interactions, particularly those involving Arg34, then, are crucial for fine-tuning the relative stabilities and geometric specificities of the different oligomeric states. MBPC-1 exclusively forms the type **4** conformation with the observed His^3Asp–coordination motif rather than the His^4-motif seen in Zn_4:MBPC-2_4, because a) the Arg34-Asp66 interaction (Figure 4.5 A) stabilizes the former geometry, and b) the latter geometry would align the original Arg62 in MBPC-1 across from Arg34 (Figure 4.5 B), leading to a destabilizing effect. Conversely, MBPC-2 specifically forms the type **4** conformation with His^4-coordination because it is energetically more favorable to utilize Asp62 in an interfacial salt bridge when a tetra-coordinate Zn-geometry is already accessible with four histidines. Lastly, a broader inspection of the interfaces in Zn_4:MBPC-1_4 and Zn_4:MBPC-2_4 reveals that the former assembly features a higher number of unfavorable interactions and side chain environments, which attests to its lower stability.

To what extent H-bonds and salt bridges contribute to the stability of proteins and

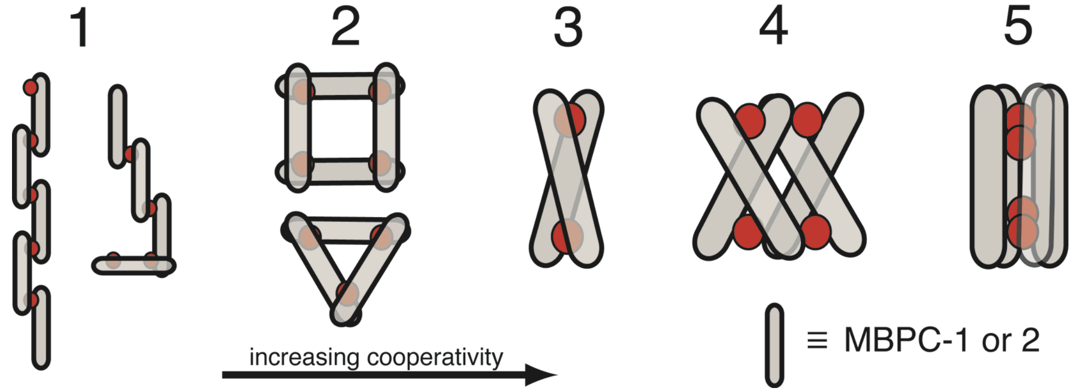


Figure 4.7. Possible Zn induced oligomerization states of MBPC-1 and 2 under limiting (≤1 equivalent) metal concentrations assuming a 4-coordinate Zn geometry. 1) extended structures/aggregates, 2) trimer or tetramer with His^4-coordination shared by two monomers, 3) dimer with His^4-coordination shared by two monomers, 4) tetramer with His^4 or His^3Asp-coordination shared by three monomers, 5) tetramer with His^4-coordination shared by four monomers.

PPIs is a matter of debate. Yet, it is generally thought that these alignment-dependent interactions are important in limiting the number of possible low energy docking conformations, thus playing a role in determining specificity. Our results provide clear evidence that salt-bridging and H-bonding interactions can dictate the geometric alignment of protein partners, leading to the population of discrete supra-molecular structures over other conformations of similar energy. The combined ability to direct PPIs through metal coordination and secondary interactions could provide the specificity required for the construction of complex, multi-component protein super-structures and the selective control of cellular processes that involve protein-protein association reactions.

Chapter 4 is reproduced in part with permission from: Salgado, E. N., Lewis, R. A., Faraone-Mennella, J., Tezcan, F. A. **2008**. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* 130, 6082–6084. Copyright 2008 American Chemical Society.

References

1. Jones, S., Thornton, J. M. **1996**. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA*. 93, 13-20.

2. Froloff, N., Windemuth, A., Honig, B. **1997**. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Protein Sci.* 6, 1293-1301.

3. Larsen, T. A., Olson, A. J., Goodsell, D. S. **1998**. Morphology of protein–protein interfaces. *Structure*. 6, 421-427.

4. Xu, D., Tsai, C. J., Nussinov, R. **1997**. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* 10, 999-1012.

5. Sheinerman, F. B., Norel, R., Honig, B. **2000**. Electrostatic aspects of protein–protein interactions *Curr. Opin. Struct. Biol.* 10, 153-159.

6. Grueninger, D., Treiber, N., Ziegler, M. O. P., Koetter, J. W. A., Schulze, M. S., Schulz, G. E. **2008**. Designed protein-protein association. *Science*. 319, 206-209.

7. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L., Gonzalez, C. **2002**. Computer-aided design of a PDZ domain to recognize new target sequences. *Nat. Struct. Biol.* 9, 621-627.

8. Shifman, J. M., Mayo, S. L. **2002**. Modulating calmodulin binding specificity through computational protein design. *J. Mol. Biol.* 323, 417-423.

9. Sun, Q.-F., Iwasa, J., Ogawa, D., Ishido, Y., Sato, S., Ozeki, T., Sei, Y., Yamaguchi, K., Fujita, M. **2010**. Self-assembled M24L48 polyhedra and their sharp structural switch upon subtle ligand variation. *Science*. 328, 1144-1147.

10. Kim, J., Chen, B., Reineke, T. M., Li, H., Eddaoudi, M., Moler, D. B., O'Keeffe, M., Yaghi, O. M. **2001**. Assembly of metal-organic frameworks from large organic and inorganic secondary building units: New examples and simplifying principles for complex structures. *J. Am. Chem. Soc.* 123, 8239-8247.

11. Salgado, E. N., Faraone-Mennella, J., Tezcan, F. A. **2007**. Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* 129, 13374-13375.

12. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A., Schuck, P. **2004**. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Analytical Biochemistry* 326, 234-256.

13. Vagin, A., Teplyakov, A. **1997**. MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.

14. Murshudov, G. N., Vagin, A. A., Dodson, E. J. **1994**. The CCP4 Suite: Programs for protein crystallography. *Acta Cryst. D*. 50, 760-763.

15. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S. K., J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. **1998**. Crystallography & NMR system: A new software system for macromolecular structure determination. *Acta Crystallogr. D.* 54, 905-921.

16. McRee, D. E. **1992**. A visual protein crystallographic software system for X11/XView. *J. Mol. Graphics*. 44-46.

17. Laskowski, R. A., Macarthur, M. W., Moss, D. S., Thornton, J. M. **1993**. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291.

18. DeLano, W. L. *The PYMOL molecular graphics system (http://www.pymol.org)*, 2003.

19. Krissinel, E., Henrick, K. **2007**. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 372, 774-797.

20. Salgado, E. N., Lewis, R. A., Faraone-Mennella, J., Tezcan, F. A. **2008**. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* 130, 6082–6084.

21. King, D. A., Zhang, L., Guarente, L., Marmorstein, R. **1999**. Structure of a HAP1-DNA complex reveals dramatically asymmetric DNA binding by a homodimeric protein. *Nat Struct Biol.* 6, 64-71.

22. Janin, J., Miller, S., Chothia, C. **1998**. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* . 204, 155-164.

Chapter 5

Metal-Templated Design of Protein Interface

Introduction

Early precursors to modern single- or multi-domain proteins likely emerged from intra- or inter-molecular association reactions between random peptide/protein surfaces.[1,2] Only those surfaces that provided sufficient interaction enthalpy to overcome the entropic cost of association were presumably subjected to natural selection to yield stable and functional protein folds or complexes.[3] Because such association surfaces comprise weak non-covalent bonds they must have been large enough to generate a sufficiently favorable binding enthalpy, which has a very low probability of occurrence by chance. In fact, it has been estimated that the likelihood of finding a foldable sequence in a random library of 80 amino-acid-long peptides that can fulfill the most basic function—binding a small molecule—is about 1 in 10^{11}.[4] In light of such low odds, metal coordination could have provided distinct evolutionary benefits in the emergence of folded and functional protein domains and complexes. From a structural perspective metal coordination is arguably the most efficient way to generate a folded domain while minimizing the sequence space required. Perhaps not coincidentally, Szostak and colleagues found that the best ATP-binding protein selected from a library of 6×10^{12} random peptide sequences contained a Zn coordination motif that was necessary for the structural organization of the ATP-binding site.[4] Similarly, Riechmann and Winter looked for chimeras of peptide segments that could bind anti-lysozyme antibodies.[5] They discovered that the most efficient binder incorporated an unforeseen heme group that dramatically stabilized the resulting dimer.

From a functional perspective metals possess properties such as Lewis acidity and redox reactivity that enable them to carry out catalytic transformations not accessible by organic building blocks. Given such immediate structural and functional benefits it is reasonable to ask whether metal coordination may have played a role in the early evolution of protein folds and complexes through an initial nucleation event. Motivated by this possibility we developed a rational design approach, metal template interface re-design (MeTIR), which mimics the time course of a hypothetical evolutionary pathway. We show that through MeTIR, non-interacting protein surfaces can first be rendered self-associating through an initial metal-templating event followed by the generation of stability, both through a small number of mutations on the protein surface. MeTIR yields unique protein architectures that stably self-assemble in the presence—or absence—of metals, thus providing a potential route for the de novo generation of protein interfaces and tunable metal coordination centers within biological scaffolds.

The underlying "template-and-stabilize" strategy for MeTIR is illustrated in Fig. 5.1. The key to MeTIR is the use of a folded, physiologically monomeric protein—instead of random polypeptide sequences or protein fragments—as a model system subject to optimization. The stability and rigidity of a folded protein ensure that its overall architecture is retained as its surface is modified and that it remains structurally tractable by crystallography. At the same time, using a monomeric protein ensures that its surface does not carry any bias towards self-association. With these factors in mind, our already implemented model protein, cyt cb_{562}, (illustrated as Species 1 in Fig. 5.1) serves as a perfect, stable scaffold for this purpose. As oligomeric assemblies of the mutant cyt cb_{562}, MBPC-1 (Figure 5.1 B), have been shown to form through the coordination of various divalent metal ions,[6-8] we next sought to demonstrate that the interactions along

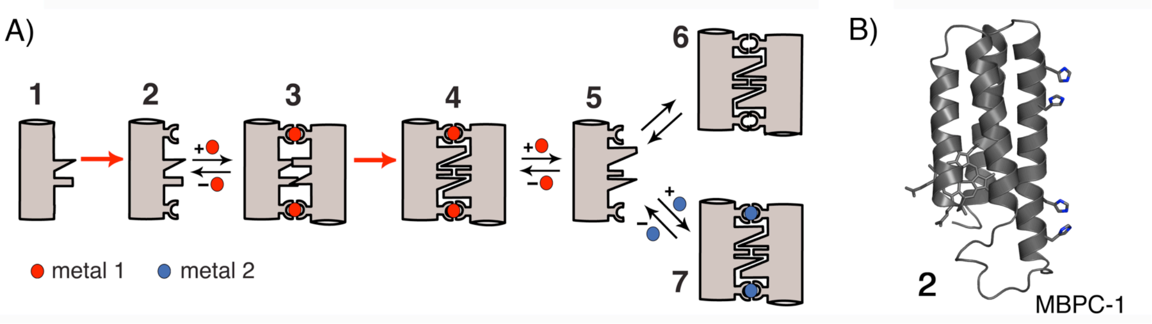


Figure 5.1. Cartoon outline for MeTIR and the species involved. A). Red arrows represent engineering or hypothetical evolutionary steps. Red and blue spheres represent metal ions 1 and 2 with different preferential coordination geometries. 1. Protein/peptide with a non-self-associating surface; 2. 1 modified with metal-coordinating groups; 3. initial metal1-templated protein/peptide complex with non-complementary interfaces; 4. metal1-templated protein/peptide complex with optimized, complementary interfaces; 5. protein/peptide with a self-associating surface; 6. metal-independent protein/peptide complex biased towards metal1 binding; 7. protein/peptide complex with distorted metal2 coordination. B) The structure of MBPC-1, corresponding to species 2, showing the two *i*, *i + 4* bis-His motifs on helix3.

these newly developed PPI interfaces could be re-designed to be more favorable. The resulting re-designed protein variant was, found to be capable of forming a thermodynamically more stable assembly than the parent MPBC-1, with identical organization and metal coordination environment to Zn_4:MBPC-1_4. Concurrently, these re-designed interfaces appear to enforce a tetrahedral geometry onto metals with non-tetrahedral preferences, and allow the protein to form a metal-independent assembly.

Materials and Methods

Protein Interface Redesign. For the cyt cb_{562} (PDB ID: 2BC5) analysis, the following criteria and programs were used for defining designable interface clusters: AREAIMOL[9] was used to calculate solvent accessible surface areas (SASA) values and residues with a SASA value below 10 Å^2 were flagged as undesignable; LIGPLOT[10] was used to identify and flag as undesignable residues contacting the heme and Zn ligands in cyt cb_{562} and Zn_4:MBPC-1_4 (PDB ID: 2QLA) respectively; and WHATIF[11] was used to calculate the optimal hydrogen bond network, with residues making side chain-main chain hydrogen bonds scoring above 0.45 being flagged as undesignable.

All sequence design and rotamer optimizations were performed with fixed backbone templates, using a variant of the RosettaDesign algorithms[12] used for optimizing multiple conformers for a single sequence,[13] in which each subunit of the oligomer represented a conformer. The variation from the previously published algorithm involves an improved rotamer optimization search algorithm in which: all rotamers are sampled in a random order; rotamers resulting in lower energies are duplicated in the rotamer set; and in each sampling iteration, the best of five independent simulated annealing trajectories are input into the next iteration where the rotamer set is reverted back to containing only unique rotamers. These improvements resulted in lower energies and increased sequence convergence with only a modest increase in sampling time.

For all rotamer optimizations, the Lennard-Jones van der Waals repulsive term was dampened and rotamers $\pm 1\sigma$ in torsion angles from canonical rotamers in $\chi 1$ for all amino acid types and in $\chi 2$ angles for aromatic amino acids were included in the rotamer set. For the native amino acid rotamer optimizations of Zn_4:MBPC-1_4, four independent trajectories were performed and the resulting residue energies and SASAprob values[14] were averaged and ranked as described above. For redesigns, twenty trajectories were performed and the sequence of the lowest energy trajectory was selected for experimental characterization. Interface residues were defined as residues in which atoms of a residue on one chain were within 5.5 Å of atoms on a residue in another chain. Neighboring residues were defined in a similar manner.

Docking Simulations. 3000 independent docking trajectories were carried out using RosettaDock.[15] The two monomers that form the **i1** interface in the Zn_4:RIDC-1_4 design (the backbones of which are identical in Zn_4:MBPC-1_4) were used as the starting structure for the docking simulations. One of the monomers was randomly spun along the axis connecting the centers of mass of both partners and the same monomer was also allowed to search a space of up to 3 Å normal to that axis, 8 Å in the plane perpendicular to the axis, and with up to an 8° tilt from the axis and an 8° additional spin around the axis. For docking simulations, these search parameters may be considered intermediary between a wide local search and a global search. Docking was performed with a full atom representation in which rotamers were optimized and rotamers deviating slightly from canonical rotamers were sampled. Docking simulations were performed prior to solving the crystal structures of RIDC-1_2 and Cu_2:RIDC-1_2.

Site Directed Mutagenesis and Protein Expression/Purification/Characterization. Site directed mutagenesis was performed on the pETc-b562 plasmid (denoted as wildtype), as above, with sequencing performed by UCSD Moores Cancer Center. Mutant plasmids were transformed into BL21(DE3) *E. coli* cells along with the *ccm* heme maturation gene cassette plasmid, pEC86. Expression and purification of the protein was performed as previously described.

Guanidine hydrochloride denaturation. In order to determine if the RIDC-1 and RIDC-2 constructs are still stable in comparison to MBPC-1, GuHCl denaturation was employed. Samples were performed at pH 7 in 20 mM TRIS, 5 mM EDTA with 5 μM protein in the presence of absence of 7 M GuHCl. The titrations were performed using an auto-titrator on an AVIV 62DS CD spectrometer. Each injection was allowed to equilibrate with stirring for 2 min., with the unfolding transition being monitored at 222 nm. The refractive indices of the solutions were determined with a Bausch and LombAbbe-3L refractometer.

Sedimentation Velocity. SV experiments were performed in order to determine the solution-state oligomerization behavior of RIDC-1 and RIDC-2. All metal containing SV experiments were performed in 20 mM TRIS (pH 7.0) using appropriate volumes of 5 mM $ZnCl_2$ or $CuSO_4$ in 20 mM TRIS (pH 7.0) to yield a 1:1 metal:protein ratio. All metal free samples were run in the same buffer and contained a final EDTA concentration of 5 mM.

SV measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor at 41,000 rpm for a total of 250 scans per sample. Wavelengths used for detection were as follows: 420 nm for 5 μM protein samples; 570 nm for 50 μM samples; and 582 nm for 200 μM samples

All SV data were processed as previously described. Vbar for each mutant was calculated assuming a partial specific volume of heme of 0.82 mg/ml. All data were processed in SEDFIT[16] with the following fixed parameters: buffer density (r) = 0.99764 g/ml; buffer viscosity = 0.0089485 poise; Vbar, which was calculated to be 0.7314 ml/g for RIDC-1, and 0.7373 ml/g for RIDC-2.

Sedimentation Equilibrium. SE experiments were performed on metal free samples of both RIDC-1 and RIDC-2 in order to determine the dissociation constant of the metal free dimers of each mutant. All samples contained 20 mM TRIS (pH 7.0), and a final EDTA concentration of 5 mM. Samples containing NaCl were similarly prepared but with the addition of 150 mM NaCl. Samples were made at loading concentrations of about 5 μM, 20 μM, and 40 μM.

SE measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor. Scans were taken at 14 and 16 hrs at wavelengths and rotor speeds between 20,000 and 30,000 rpm. The two scans were overlaid to ensure that equilibrium had been achieved, but only the 16 hr. scans were used for global analysis in SEDPHAT.[17]

For all data sets, the menisci were assigned based on 41,000 rpm meniscus depletion scans and were fixed as constants. Initially, species analysis was performed using single-species monomer and dimer models, as well as a monomer-dimer multi-species model. For each data set, the molecular masses of the species (monomeric RIDC-1, 12303 Da; monomeric RIDC-2, 12269 Da) were held constant, as was the position of the cell bottom. Buffer viscosity and buffer density were also held constant at the values given above for samples without additional NaCl. Samples containing 150 mM NaCl had a calculated buffer density and viscosity of 1.00382 g/ml and 0.00908 poise, respectively. In all instances, the best fit was derived from the monomer-dimer multi-species model, thus leading to the use of a monomer-dimer self association model to derive association constants.

When fitting the data towards the monomer-dimer self-association model, the molecular mass of the monomeric proteins and the extinction coefficient of the protein at each wavelength were all held constant. The association constant; initial loading concentrations of protein; and the bottom of the cell were treated as floating parameters. A minimized fit was attained once rigorous F-statistics analyses were performed on individual floating parameters. Standard deviations for the resulting log (K_a) values were determined through Monte-Carlo analysis within SEDPHAT. Final log (K_a) values, calculated K_d, and fit statistics for all fits are shown in Table 5.1.

Electron Paramagnetic Resonance (EPR) spectroscopy. X-band EPR spectra were obtained at 125 K on a Bruker ELEXSYS E500 spectrometer equipped with a Bruker ER4131VT variable-temperature unit. The EPR samples consisted of 150 μM RIDC-1 in 20 mM MOPS buffer (pH 7), with or without 100 μM copper (II) chloride dihydrate

Table 5.1. Results of global fitting of SE data for RIDC-1 and RIDC-2 with 5 mM EDTA. The chi-squared and rmsd values of the globally fit SE data for both mutants with and without 150 mM NaCl are given, along with resulting association and dissociation constants. These data correspond to the data shown in Figures 5.7 through 5.10.

Mutant		
RIDC-1	**χ^2 of Global Fit**	**Fit rmsd**
	1.084634	0.004513
	$Log_{10}(K_a)$, Monomer-Dimer	**Monomer-Dimer K_d**
	4.61099 ($\pm$4.097717 x 10^{-01})	2.45 x 10^{-5} M
RIDC-2	**χ^2 of Global Fit**	**Fit rmsd**
	1.67598	0.005853
	$Log_{10}(K_a)$, Monomer-Dimer	**Monomer-Dimer K_d**
	4.25507 ($\pm$3.306995x 10^{-01})	5.56 x 10^{-5} M
RIDC-1 150 mM NaCl	**χ^2 of Global Fit**	**Fit rmsd**
	1.095547	0.005131
	$Log_{10}(K_a)$, Monomer-Dimer	**Monomer-Dimer K_d**
	4.3783 ($\pm$ 3.489352x 10^{-01})	4.29 x 10^{-5} M
RIDC-2 150 mM NaCl	**χ^2 of Global Fit**	**Fit rmsd**
	1.577982	0.003407
	$Log_{10}(K_a)$, Monomer-Dimer	**Monomer-Dimer K_d**
	4.3973 ($\pm$ 3.576307x 10^{-02})	4.01 x 10^{-5} M

The spectrum of the copper-free sample was subtracted from that of the copper-containing sample to obtain the final spectrum. Spectra were recorded using the following spectrometer conditions: Microwave frequency, 9.389688 GHz; power, 0.63 mW; modulation amplitude, 0.10 mT.

Crystallography. All crystals were obtained by sitting drop vapor diffusion at room temperature. The crystallization conditions for the four different crystal forms described in this study are as follows: Zn_4:RIDC-1_4 – 1 μL of precipitant solution (100 mM HEPES, pH 7.0, 30% Jeffamine ED-2001 (pH 7.0), 1.5 mM $ZnCl_2$) and 2 μL protein (1.5 mM in 20 mM TRIS, pH 7.0) in the sitting drop; Zn_4:RIDC-2_4 – 1 μL of precipitant solution (100 mM BIS-TRIS, pH 6.5 , 25% PEG 3350, 3.4 mM $ZnCl_2$) and 2 μL protein (1.7 mM in 20 mM TRIS, pH 7.0) in the sitting drop; RIDC-1_2 – 1 μL of precipitant solution (26% PEG 2000) and 2 μL protein (3.8 mM in 20 mM TRIS, pH 7.0) in the sitting drop; Cu_2:RIDC-1_2 – 1 μL of precipitant solution (100 mM HEPES, pH 7.5, 200 mM NaCl, 25% PEG 3350, 10.7 mM $CuSO_4$) and 2 μL protein (5.4 mM in 20 mM TRIS, pH 7.0) in the drop. Appropriate crystals were transferred to a solution of mother liquor containing 20% glycerol as a cryoprotectant and frozen in liquid nitrogen or directly in the cryostream prior to measurement. X-ray diffraction data for Zn_4:RIDC-1_4 and Zn_4:RIDC-2_4 were collected at 100 K at the Stanford Synchrotron Radiation Laboratory (BL 9-2 and 7-1, respectively) using 1-Å radiation. The data were integrated and scaled using Denzo/SCALA. Data for metal-free RIDC-1_2 and Cu_2:RIDC-1_2 were collected at 100 K using a Bruker Apex II CCD detector and monochromatized Cu-Kα radiation (1.54 Å) produced by a Siemens sealed tube source. The data were processed using SAINT and Bruker SADABS. All structures were determined through molecular

Table 5.2. X-ray data collection and refinement statistics for Zn-mediated RIDC-1 and RIDC-2 assemblies.
[‡] $R_{sym} = \Sigma\Sigma_j|I_j - \langle I\rangle|/\Sigma\Sigma_j|I_j|$.
[§] $R = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$.
[II] Free R calculated against 7.2% of the reflections removed at random for both structures.
[¶] Root mean square deviations from bond and angle restraints.
[*] Numbers in parentheses correspond to the highest resolution shell.

	Zn_4:RIDC-1_4	**Zn_4:RIDC-2_4**
Residues in complex	4 x (106 + 1 Heme + 1 Zn)	4 x (106 + 1 Heme + 1 Zn)
No. of tetramers / asymmetric unit	2	1
Waters in asymmetric unit	136	213
Unit cell dimensions (Å)	57.2 x 87.9 x 152.2	48.4 x 65.2 x 70.8
	$\alpha = \beta = \gamma = 90^\circ$	$\alpha = \gamma = 90^\circ$, $\beta = 104.70^\circ$
Symmetry group	$P2_12_12_1$	$P2_1$
Resolution (Å)	76.0 - 2.35	30.7 - 2.0
X-ray wavelength (Å)	0.979	0.976
Number of Unique Reflections	32072	26855
Redundancy	3.4	2.6
Completeness (%)*	98.3 (98.3)	92.5 (92.4)
⟨I / sI⟩*	5.3 (1.4)	11.0 (2.7)
Rsym[‡] (%)*	11.8 (54.8)	4.4 (17.6)
R[§] (%)*	22.6 (29.5)	22.8 (27.1)
Free R[II] (%)*	27.8 (35.8)	27.2 (32.1)
Rms Bnd[¶] (Å)	0.011	0.009
Rms Ang[¶] (°)	1.28	1.19
Ramachandran plot (%)		
Residues in most favored regions	97.2	96.1
Residues in add.l allowed regions	2.8	3.9
Residues in generously allowed regions	0.0	0.0
Residues in disallowed regions	0.0	0.0

Table 5.3. X-ray data collection and refinement statistics for RIDC-1_2 and Cu_2:RIDC-1_2 assemblies.
$^{‡}R_{sym} = \Sigma\Sigma_j|I_j - <I>|/\Sigma\Sigma_j|I_j|$.
$^{§}R = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$.
^{II}Free R calculated against 6.9% of the reflections removed at random for the structures.
¶Root mean square deviations from bond and angle restraints.
*Numbers in parentheses correspond to the highest resolution shell.

	RIDC-1_2	**Cu_2:RIDC-1_2**
Residues in complex	2 x (106 + 1 Heme)	4 x (106 + 1 Heme + 1 Zn)
No. of tetramers / asymmetric unit	1	1
Waters in asymmetric unit	186	206
Unit cell dimensions (Å)	49.3 x 55.5 x 72.4	66.2 x 87.0 x 80.8
	$\alpha = \beta = \gamma = 90^{o}$	$\alpha = \beta = \gamma = 90^{o}$
Symmetry group	$P2_12_12_1$	$C222_1$
Resolution (Å)	27.8 - 2.1	22.2 - 2.2
X-ray wavelength (Å)	1.542	1.542
Number of Unique Reflections	12023	11999
Redundancy	3.5	2.9
Completeness (%)*	98.7 (96.5)	99.5 (98.4)
⟨I / sI⟩*	13.4 (4.1)	6.5 (2.0)
Rsym‡ (%)*	7.8 (18.7)	16.8 (48.2)
R§ (%)*	20.0 (21.8)	20.4 (23.3)
Free R^{II} (%)*	27.1 (20.6)	27.0 (27.9)
Rms Bnd¶ (Å)	0.007	0.008
Rms Ang¶ (^{o})	0.99	1.1
Ramachandran plot (%)		
Residues in most favored regions	98.5	97.9
Residues in add.l allowed regions	1.5	2.1
Residues in generously allowed regions	0.0	0.0
Residues in disallowed regions	0.0	0.0

replacement with MOLREP[18] using the cyt cb_{562} monomer structure (PDB ID:2BC5) as the search model. Rigid-body, positional, thermal and TLS refinement with REFMAC[19, 20] using appropriate non-crystallographic symmetry restraints, along with manual rebuilding, and water placement with XFIT[21] produced the final models. All figures were produced with PYMOL.[22]

Results and Discussion

Analysis of protein interfaces. We have previously shown that MBPC-1 self-assembles into discrete oligomeric structures upon metal coordination by the bis-His motifs. Given the non-self-associating surface of cyt cb_{562}, the oligomerization of MBPC-1 is entirely driven by metal coordination. Consequently, the supra-molecular arrangement of MBPC-1 is dictated by the inner-sphere metal coordination geometry: the octahedral and the square-planar coordination preferences of Ni^{2+} and Cu^{2+} respectively lead to a C_3-symmetrical trimer and a C_2-symmetrical dimer, whereas tetrahedral Zn^{2+} coordination yields a D_2-symmetrical tetramer (Figure 5.2). These three structures are collectively illustrated as Species 3 (Fig. 5.1 A), whose interfaces lack complementary non-covalent interactions that would typically be expected to drive protein–protein associations. According to our hypothetical evolutionary pathway, Species 3 therefore represents the initial peptide or protein ensemble that has assembled around a metal ion, a process that has required a minimal number of mutations in the amino acid sequence.

The Ni- and Cu-driven MBPC-1 complexes feature small (<1000 $Å^2$) interfaces. The Zn-driven tetramer (Zn_4:MBPC-1_4), in contrast, provides an extensive buried surface

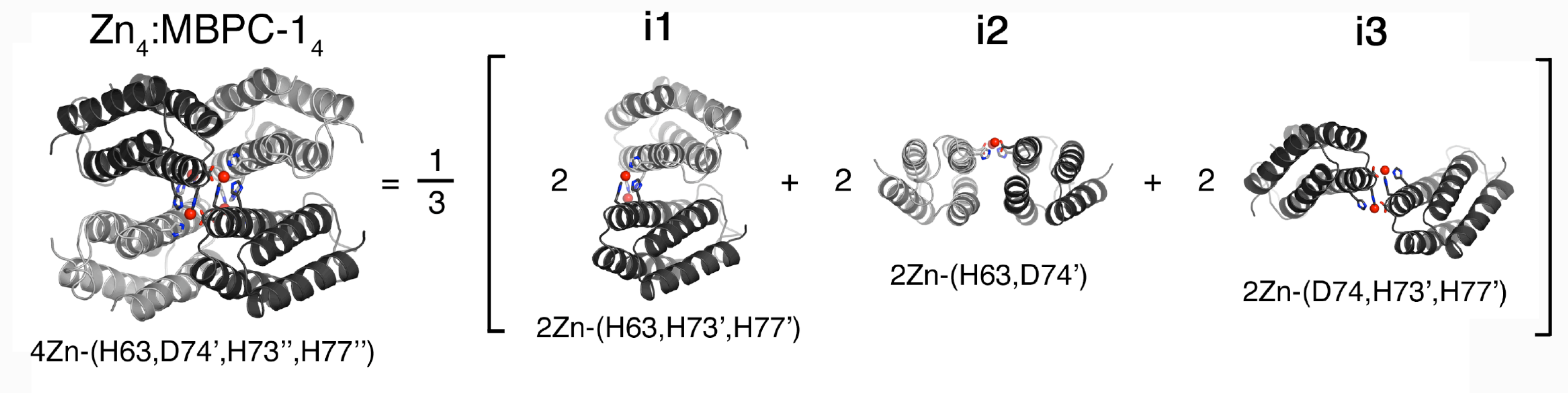


Figure 5.2. Interfaces of Zn_4:MBPC-1_4. The three pairs of interfaces (i1, i2, i3) formed within the D_2-symmetrical Zn_4:MBPC-1_4 tetramer are shown with the Zn coordination environment in each interface listed below.

(~5000 Å^2) with intimate contacts among its four subunits. Hence, Zn_4:MBPC-1_4 was chosen as the focus of surface redesign efforts to generate a stable non-covalent bonding network around the metal centers. Owing to its D_2 (222) symmetry, Zn_4:MBPC-1_4 can be dissected into three pairs of C_2- symmetrical interaction surfaces (Figure 5.2). As seen in most natural D_2-symmetrical protein complexes,[23] two of the three interfaces (i1 and i2) in Zn_4:MBPC-1_4 are significantly more extensive than the third (i3), which is almost entirely made up by metal coordinating residues. Each interface is held together by one or two Zn ions using alternative combinations of four coordinating residues (the engineered His73 and His77, and the native His 63 and Asp74). Importantly, the third engineered residue, His59, was found not to be involved in Zn coordination, meaning that the total number of required mutations to obtain a tetrameric complex is only two.

Together the three pairs of interfaces make an interwoven assembly linked together by four Zn ions. From a redesign perspective i1, i2, and i3 can be considered as independent targets since the amino acid side chains that make up their cores do not overlap. However, because the three interfaces constitute a cooperative assembly, their contributions to the stability of the entire Zn-mediated tetrameric architecture should be synergistic. The buried surface areas in i1, i2, and i3 in the Zn_4:MBPC-1_4 complex are 1080, 870, and 490 Å^2, respectively and comprised almost entirely of polar side chains that are forced together by Zn coordination. Given its small size we did not pursue the redesign of i3 and first sought to optimize the largest interface i1, followed by i2.

General re-design strategy. To achieve a successful redesign of interfaces we developed a strategy that explicitly addressed a) preserving the fold of the cyt cb_{562} monomer and b) mutating the minimal number of residues that might have the maximal impact on binding affinity in the Zn_4:MBPC-1_4 complex. For the first goal, we analyzed the structures and flagged as "undesignable'' all sequence positions in which the residues contacted the heme groups or Zn atoms (in Zn_4:MBPC-1_4), residues that had low solvent accessible surface areas (SASA), and residues that were involved in side chain-main chain hydrogen bonds.

For the second goal, the remaining set (designable residues) were subjected to rotamer optimizations employing a variant of the RosettaDesign algorithms[12] used for optimizing multiple conformers for a single sequence.[13] In this case, each monomeric subunit of the tetramer represented a conformer. The designable residues were then ranked from high to low according to the Rosetta energy values and quality of packing scores.[14] It was then determined which of those residues at the top of the list (i.e., poorly packed residues and residues with high energy) were at the interface. The neighboring residues of each non-optimal designable interface residue were then enumerated, which yielded clusters of designable residues. These clusters were subsequently used for redesign (i.e., concerted optimization of amino acid types at all cluster sequence positions). Throughout all procedures the backbones of the monomeric subunits were held fixed as intended by templating.

Re-design of i1. The interface i1 is formed between two monomers found in a crisscrossed antiparallel arrangement that allows close contacts between four α-helices (helices 2 and 3 from each monomer), in turn yielding the most extensive and intimate inter-protein surface (Figure 5.2 and Figure 5.3 A). The average gain in solvation energy upon the formation of i1 in the native structure was calculated to be only −1.5 kcal⁄mol,

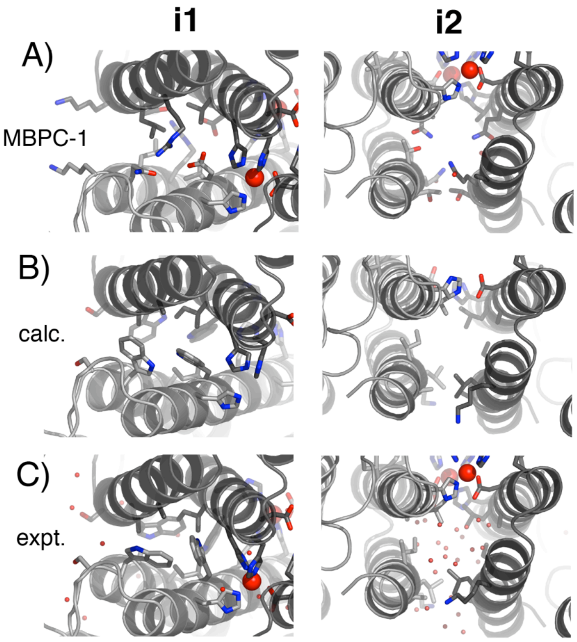


Figure 5.3. Side chain conformations in interfaces i1 and i2. Side chains in the core regions of the two interfaces in A) Zn_4:MBPC-1_4, B) the Rosetta-calculated model, and C) Zn_4:RIDC-2_4. Highlighted are six positions in each interface that were subjected to redesign, as well as those involved in Zn coordination. Water molecules observed in the Zn_4:RIDC-2_4 structure are shown as small red spheres.

suggesting that the non-covalent interactions in i1 are not expected to yield a stable dimer.[24] Our design strategy yielded a set of six mutations for il: Arg34Ala, Leu38Ala, Gln41Trp, Lys42Ser, Asp66Trp, structure indicates that they make up the bulk of i1, collectively burying ~600 Å^2. In the redesigned i1 interface the mutated residues were predicted to form a well-packed hydrophobic core in which Ala34/Trp66' (where the apostrophe indicates the second monomer) and Ala38/Trp41/Ile69' form knob-in-hole interactions (Figure 5.3 B). Although the potential contribution of position 42 to i1 stability is less obvious, calculations suggested a strong preference for Ser to the native Lys in this position. Consequently, all of the six prescribed mutations were incorporated into MBPC-1 to generate the construct RIDC-1 (Rosetta Interface Design Cytochrome-1).

Redesign of i2. We then addressed the redesign of the interface i2. The monomers that comprise i2 are positioned side by side, whereby the majority of inter-protein contacts are now made between two helices 4 aligned anti-parallel to each other. As expected again from a monomeric protein, the interfacial contacts in i2 are non-complementary and calculated to yield an average gain of solvation energy of only −3 kcal⁄mol. In contrast to i1, the side-by-side alignment of monomers in i2 precludes the formation of a well-packed and solvent-protected core (Figure 5.3 A). For this interface our design strategy converged on smaller hydrophobic residues (Ile67Leu, Gln71Ala, Ala89Lys, Gln93Leu, Thr96Ala, and Thr97Ile) (Figure 5.3 B) at six surface positions that contribute a large fraction (~400 Å^2) of the buried surface in i2. In the calculated model of redesigned i2, Leu93 and Ala96 side chains from each monomer inter-digitate across the center of i2 to form a modest-sized hydrophobic patch. Ala71 and Ile97 side chains are predicted to form intra-protein knob-in-hole interactions, whereas Ile67 appears mostly to add hydrophobic bulk in the interface. Given the small and less-than-optimally packed core of redesigned i2, we predicted that it would not contribute significantly to the stability of the Zn induced tetramer on its own. Hence, we constructed the second-generation variant, RIDC-2, which includes the six calculated mutations in i2 in addition to the six incorporated into RIDC-1.

Stability and Zn-Mediated Oligomerization of RIDC-1 and RIDC-2. Both RIDC-1 and RIDC-2 involve a significant increase in surface hydrophobicity compared to the parent construct MBPC-1. We therefore asked if they would be stable and maintain the native tertiary fold. Both variants are expressed in large yields in bacterial cultures, remain soluble, and feature heme groups with wild-type absorption features, suggesting that they are correctly folded. Chemical denaturation titrations indicate that both RIDC-1 and RIDC-2 are at least as stable as MBPC-1 (Figure 5.4) although RIDC-2 displays non-two-state unfolding behavior due likely to the presence of additional hydrophobic mutations on its surface.

The effect of interface redesign on the overall stability of the Zn induced tetrameric assembly was assessed by SV experiments (Figure 5.5). Previous SV measurements on MBPC-1 indicated that the predominant species in solution at low protein and equimolar Zn concentrations was monomeric (Figure 5.5 A). Dimeric and tetrameric species become significantly populated only at MBPC-1 and Zn concentrations over 100 μM (1 : 1 protein : Zn) with increasing concentrations favoring the population of the tetrameric form. As a benchmark, the relative populations of monomeric, dimeric, and tetrameric species at 600 μM MBPC-1 and Zn are 12%, 35%, and 50%, respectively

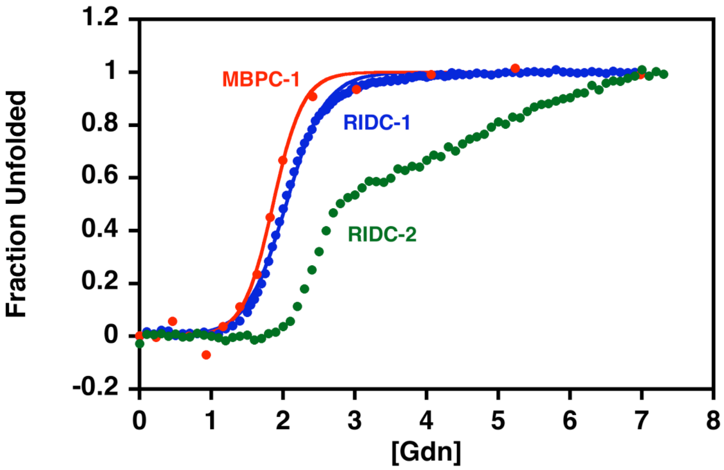


Figure 5.4. Chemical denaturation profiles for MBPC-1, RIDC-1 and RIDC-2 monitored by CD spectroscopy (at 222 nm). The unfolding of MBPC-1 and RIDC-1 is well described by a two-state model, yielding ΔG_f of -22.1 kJ x mol^{-1} and -17.8 kJ x mol^{-1}, respectively. The titrations were carried out using 5 μM protein in 20 mM TRIS buffer (pH 7) and 5 mM EDTA. Interestingly, RIDC-2 appears even more stable than both MBPC-1 and RIDC-1, judging from the onset of denaturation at higher guanidine hydrochloride (GuHCl) concentrations, although its unfolding does not appear to be two-state. We suggest that the additional hydrophobic residues present on the RIDC-2 surface lead to the formation of a folding intermediate not observed in the other two variants.

(Figure 5.5 B). In the cases of RIDC-1 and RIDC-2, on the other hand, there is only a very small detectable amount (approximately 6% of total species) of the monomeric form even at 5 μM protein and equimolar Zn. By 200 μM RIDC-1 or RIDC-2 and Zn, the tetrameric form is fully (>95% of total) populated (Figure 5.5 C-F). Because protein oligomerization is intimately coupled to metal binding in the tetrameric structures, it is challenging to obtain individual stability constants for the numerous equilibria leading to their formation. Nevertheless, the SV population distributions at different protein and Zn concentrations indicate that the interfacial mutations in RIDC-1 and RIDC-2 stabilize the Zn induced tetramer by nearly two and three orders of magnitude, respectively, over the parent MBPC-1 species.

Next, we set out to establish whether the supra-molecular architecture of Zn_4:MBPC-1_4 is maintained in the RIDC-1 and RIDC-2 tetramers as intended by the template-and-stabilize strategy. The crystal structures of Zn_4:RIDC-1_4 and Zn_4:RIDC-2_4 were determined at 2.4 Å and 2.0 Å resolution, respectively. A backbone superposition of the MBPC-1, RIDC-1, and RIDC-2 tetramers (Species 4 in Figure 1) indicates that they are topologically identical (rmsd over 424 αC's = 0.4 Å and 0.6 Å) despite a total of 24 mutations on the former and 48 on the latter complex relative to Zn_4:MBPC-1_4 (Figure 5.6 A). The tetrahedral Zn-coordination geometry in Zn_4:RIDC-1_4 and Zn_4:RIDC-2_4 is largely unchanged from Zn_4:MBPC-1_4, although the higher resolution structures of the former two reveal that Asp74 actually coordinates Zn in a monodentate fashion (Figure 5.6 B). Likely due to the same reason, the plane of the His73 imidazole ring in Zn_4:RIDC-1_4 and Zn_4:RIDC-2_4 is found to be rotated by 90° compared to that originally modeled in the Zn_4:MBPC-1_4 structure.

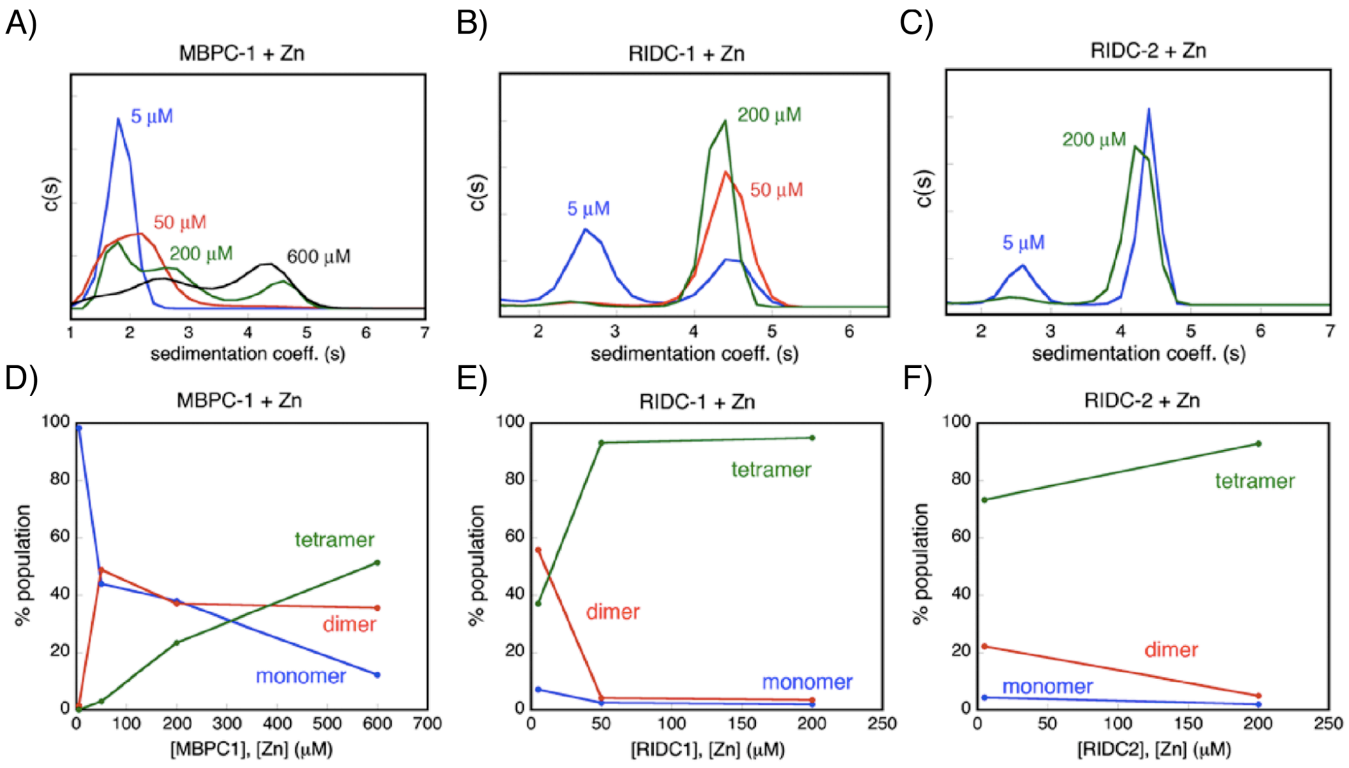


Figure 5.5. Sedimentation coefficient distributions of RIDC-1 and RIDC-2. Distributions for various concentrations of MBPC-1 (A), RIDC-1 (C) and RIDC-2 (E) in the presence of equimolar Zn^{2+} are shown. B, D, and F) Percent population of monomeric, dimeric and tetrameric species based on sedimentation coefficient distributions.

The redesigned i1 interface, which features the same substitutions in RIDC-1 and RIDC-2, is structurally superimposable in both Zn_4:RIDC-1_4 and Zn_4:RIDC-2_4 structures with small variations in side chain conformations. As predicted, the engineered Trp41 and Trp66 side chains provide the bulk of the hydrophobic core in i1, which now buries ~1500 $Å^2$ of surface area and is largely devoid of solvent molecules (Figure 5.3 C and Figure 5.6 C). The Zn-coordination sphere in Zn_4:RIDC-1_4 and Zn_4:RIDC-2_4 is intimately linked to the i1 core through interactions between the coplanar His73' and Trp66 aromatic rings (Figure 5.6 B and C). The Trp66 side chain is further anchored by H-bonding between the indole imine and the Ile69' backbone carbonyl group across the interface. Interestingly, there appears to be some fluidity within i1 because the electron density corresponding to the Trp41 side chain is best accommodated with two overlapping conformations of the indole ring (Figure 5.6 D).

In contrast to i1, the redesigned i2 is replete with solvent molecules (Figure 5.3 C). As expected from the small sizes of the substituted residues, the increase in buried surface area in i2 upon redesign is minimal (20 $Å^2$) with a modest calculated gain in solvation free energy (−1.5 kcal⁄mol). Taken together the structural details of i1 and i2 are consistent with the results from SV measurements that redesign of i1 leads to a significant stabilization of the Zn induced tetramer and that of i2 has a small incremental effect.

Metal independent dimerization of RIDC-1. Since the redesign of i1 generates an extensive set of hydrophobic interactions in Zn_4:RIDC-1_4 and Zn_4:RIDC-2_4, we examined whether these interactions could also sustain stable monomer–monomer interactions in the absence of metal coordination. SV and SE experiments show that

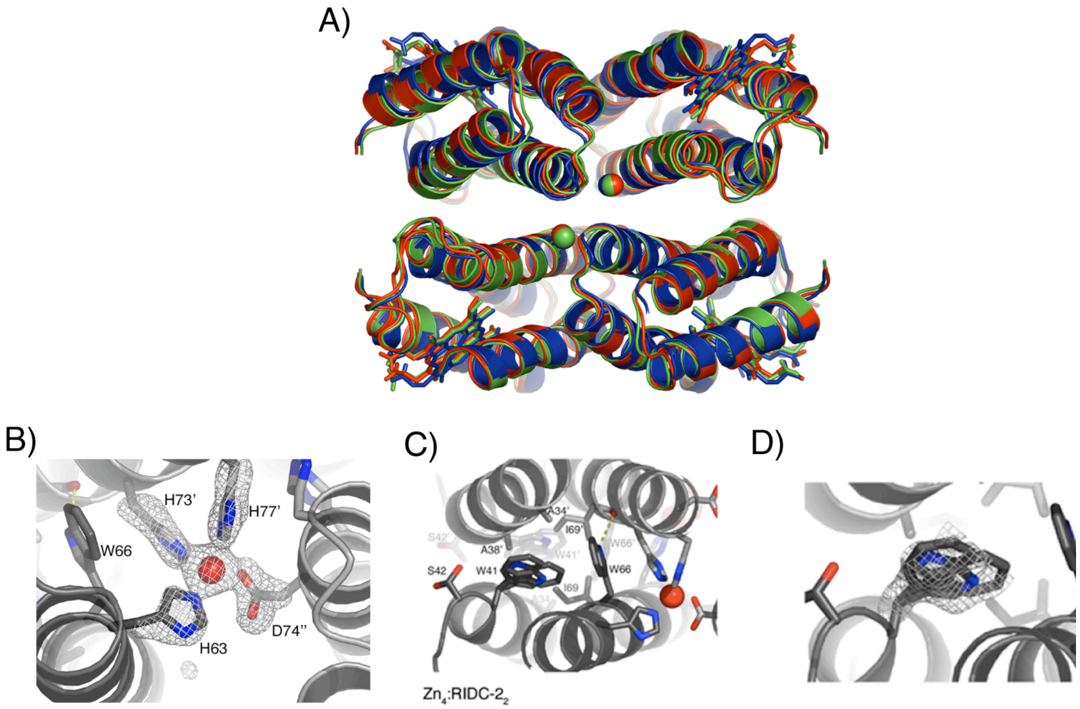


Figure 5.6. Crystal structure of Zn$_4$:RIDC-1$_4$ tetramer. A) Backbone superposition of Zn$_4$:MBPC-1$_4$ (green), Zn$_4$:RIDC-1$_4$ (blue), and Zn$_4$:RIDC-2$_4$ (red). B) Tetrahedral Zn-coordination environment in Zn$_4$:RIDC-2$_4$, with the corresponding Fo-Fc omit electron density map (3.2σ). C) i1 packing interactions in Zn$_4$:RIDC-2$_2$. D) Electron density showing the fluxionality of the Trp41 side chain in Zn$_4$:RIDC–2$_2$.

RIDC-1 and RIDC-2 indeed form a metal-independent dimeric species (corresponding to Species 6 in Figure 5.1) (Figure 5.7 and Figure 5.8). The monomer-dimer K_d have been determined by SE measurements to be 25 μM for RIDC-1_2 and 55 μM for RIDC-2_2 under low ionic-strength conditions (20 mM TRIS buffer, pH 7, 5 mM EDTA, Table 5.1), and 43 μM and 40 μM, respectively, under high ionic-strength conditions (150 mM NaCl in addition, Figure 5.9 and Figure 5.10, Table 5.1). Similar Kd's for RIDC-1_2 and RIDC-2_2 suggest that dimerization in both cases involves the same protein interface (i.e., i1), while the lack of an apparent ionic-strength dependence in these values implies that dimerization is mainly driven by hydrophobic interactions.

Crystals of RIDC-1_2 were obtained at protein concentrations (>3 mM) that should favor the formation of the metal-independent dimer. The 2.0 Å resolution structure reveals a pair of RIDC-1 molecules in the asymmetric unit that arrange in an anti-parallel fashion to form a 1300-$Å^2$ interface (Figure 5.11). This C_2-symmetrical interface is formed largely along the helix 3 from each monomer, primarily utilizing the engineered hydrophobic residues and those that are originally involved in Zn coordination. As detailed in Figure 5.11 B, His73 and Leu76 from one monomer form a patch of hydrophobic contacts across the interface with Trp41', His63', Ile69', and Trp66', whose indole group extends out toward the solvent without being involved in obvious CPIs. The orientations of Trp41 and Trp66 side chains are stabilized by H-bonds to the Arg62 backbone carbonyl and the Asp74' carboxylate groups, respectively. The Trp41– Arg62 interaction is part of a larger H-bonding circuit that involves a conserved water molecule in addition to His77' and Glu81'. The latter closes the circuit through extensive salt bridges to the Arg62 side chain.

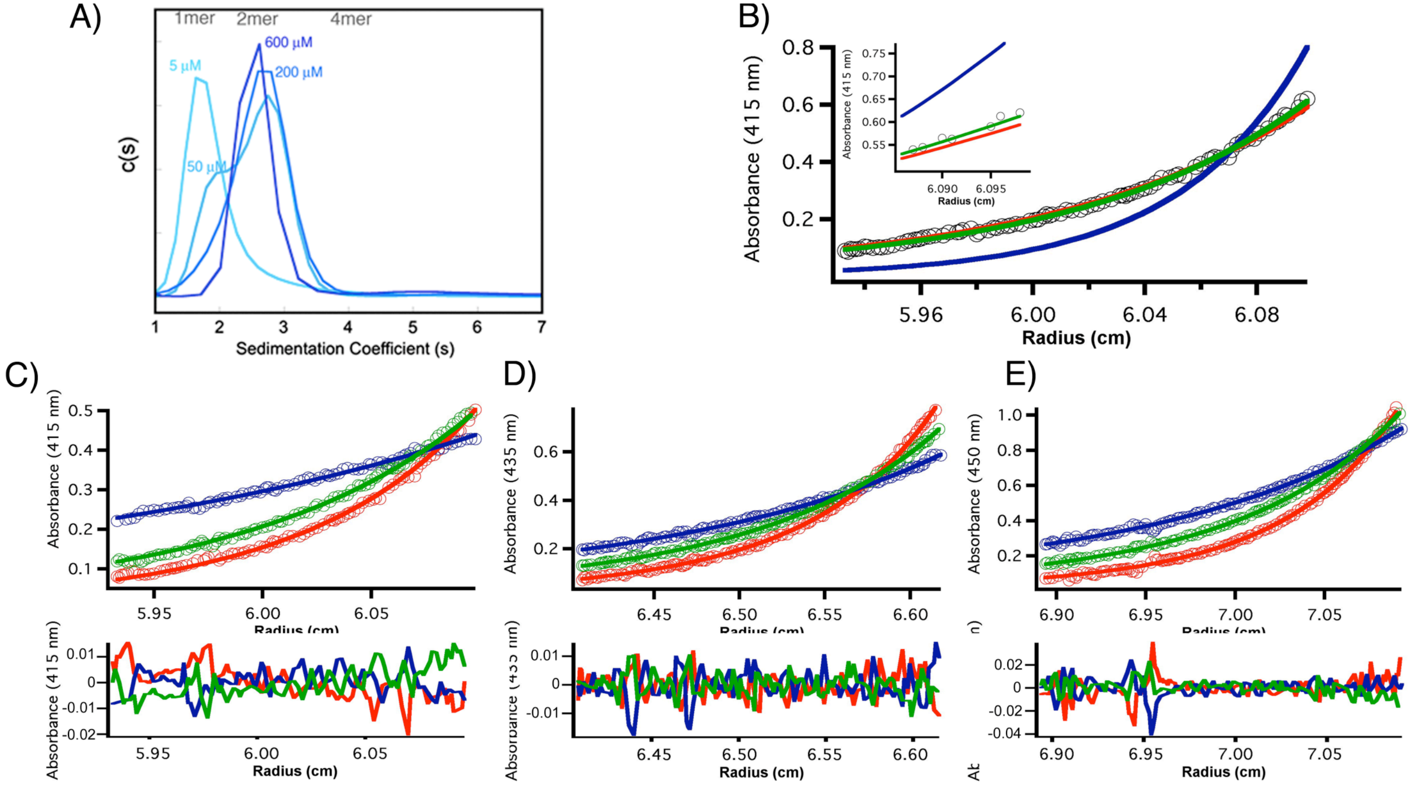


Figure 5.7. SE and SV data for RIDC-1 in the presence of 5 mM EDTA. A) Sedimentation coefficient distributions for various concentrations of RIDC-1 in the presence of 5 mM EDTA. B) Comparison of species analysis models used to globally fit SE data. Only data and fits for the 20,000-rpm scan of the 5 μM protein sample are shown (Black Circles), along with fits for the single species monomer (Red) and dimer (Blue), and monomer-dimer equilibrium (Green) models. (Inset) Close-up of the fits of the three different models. C) 5 μM protein at 20,000, 30,000, and 35,000 rpm monitored at 415 nm. D) 20 μM protein at 20,000, 25,000, and 30,000 rpm monitored at 435 nm. E) 40 μM protein at 20,000, 25,000, and 30,000 rpm monitored at 450 nm. All samples were run in 20 mM TRIS buffer (pH 7), at 25° C, with 5 mM EDTA. Scans shown in B), C) and D) were globally fit to a monomer-dimer model yielding a minimized fit whose results and statistics are shown in Table 5.1.

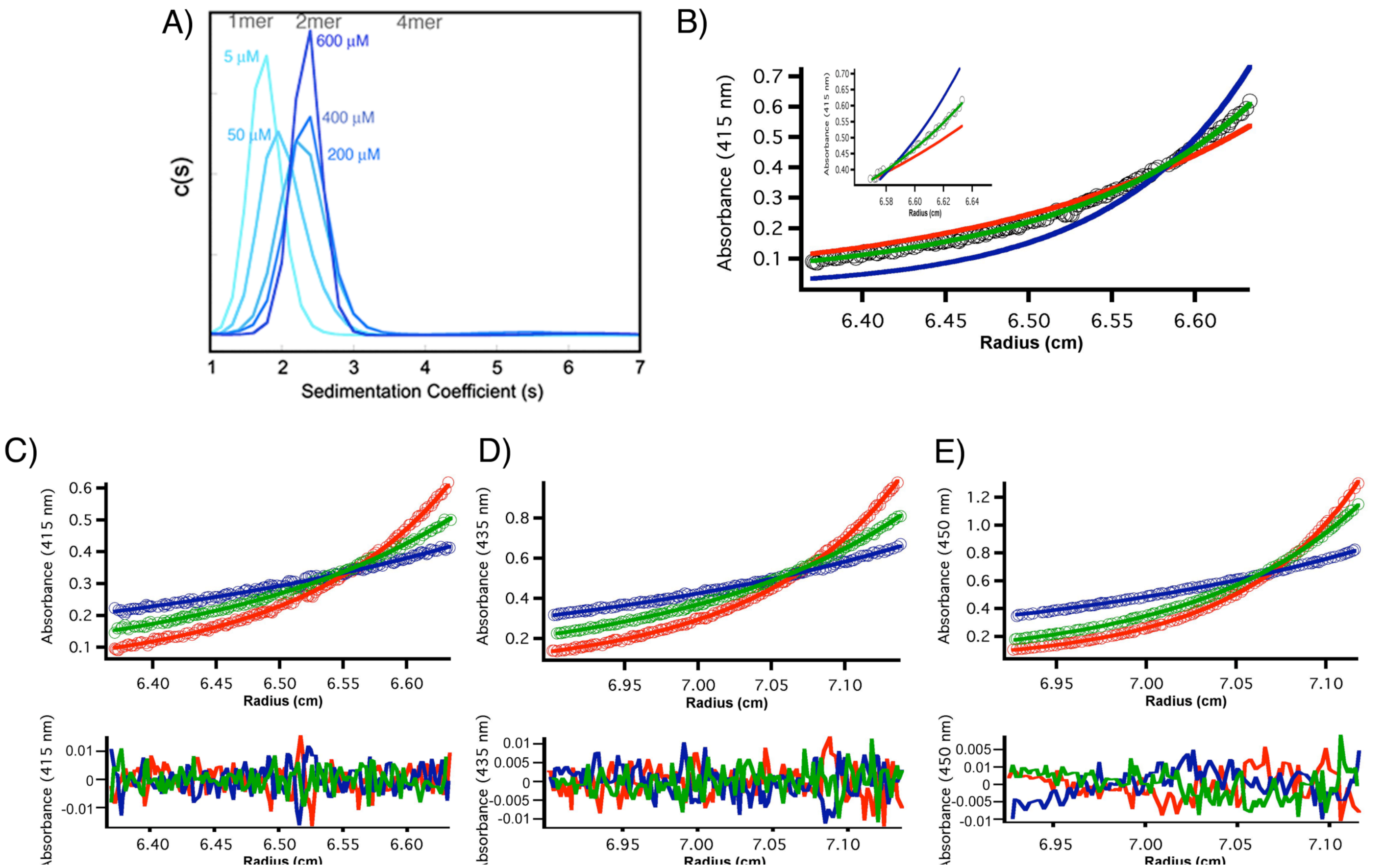


Figure 5.8. SE and SV data for RIDC-2 in the presence of 5 mM EDTA. A) Sedimentation coefficient distributions for various concentrations of RIDC-2 in the presence of 5 mM EDTA. B) Comparison of species analysis models used to globally fit SE data. Only data and fits from the 30,000 rpm scan of the 5 µM protein sample are shown (black circles) with fits from the single species monomer (red) and dimer (blue), and multi-species monomer-dimer (green) models superimposed. (inset) Close-up of the fits of the three different models. C) 5 µM protein at 20,000, 30,000, and 35,000 rpm detected at 415 nm. D) 20 µM protein at 15,000, 20,000, and 25,000 rpm detected at 435 nm. E) 40 µM protein at 15,000, 20,000, and 25,000 rpm detected at 450 nm. All samples were run in 20 mM TRIS buffer (pH 7), at 25° C, with 5 mM EDTA. Scans shown in B), C) and D) were globally fit to a monomer-dimer model yielding a minimized fit whose results and statistics are shown in Table 5.1.

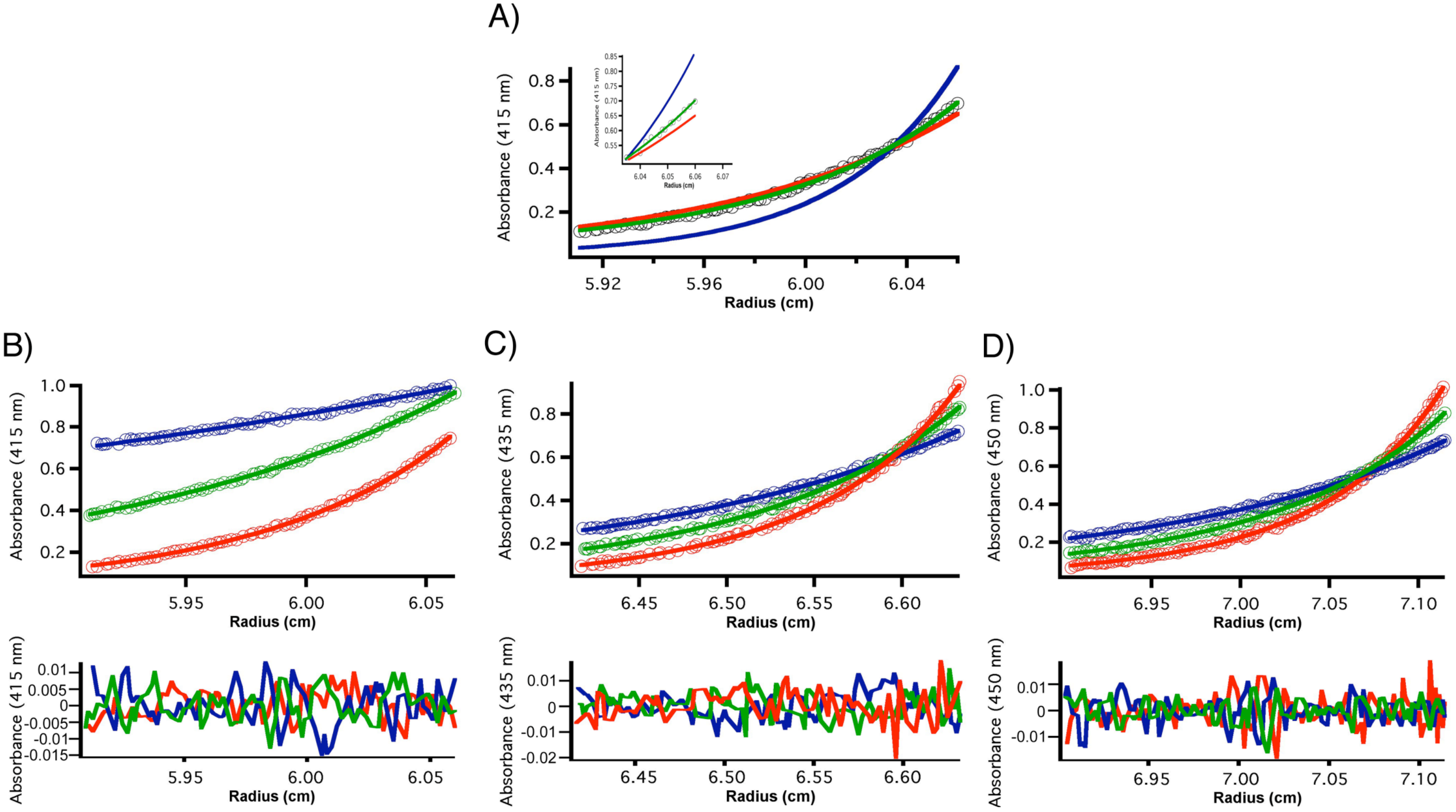


Figure 5.9. Sedimentation equilibrium data for RIDC-1 with 5 mM EDTA and 150 mM NaCl. A) Comparison of species analysis models used to globally fit SE data. Only data and fits from the 35,000 rpm scan of the 5 μM protein sample are shown (black circles) with fits from the single species monomer (red) and dimer (blue), and multi-species monomer-dimer (green) models superimposed. (inset) Close-up of the fits of the three different models. B) 5 μM protein at 15,000, 25,000, and 35,000 rpm detected at 415 nm. C) 20 μM protein at 20,000, 25,000, and 30,000 rpm detected at 435 nm. D) 40 μM protein at 20,000, 25,000, and 30,000 rpm detected at 450 nm. All samples were run in 20 mM TRIS buffer (pH 7), at 25° C, with 5 mM EDTA and 150 mM NaCl. Scans shown in B), C) and D) were globally fit to a monomer-dimer model yielding a minimized fit whose results and statistics are shown in Table 5.1.

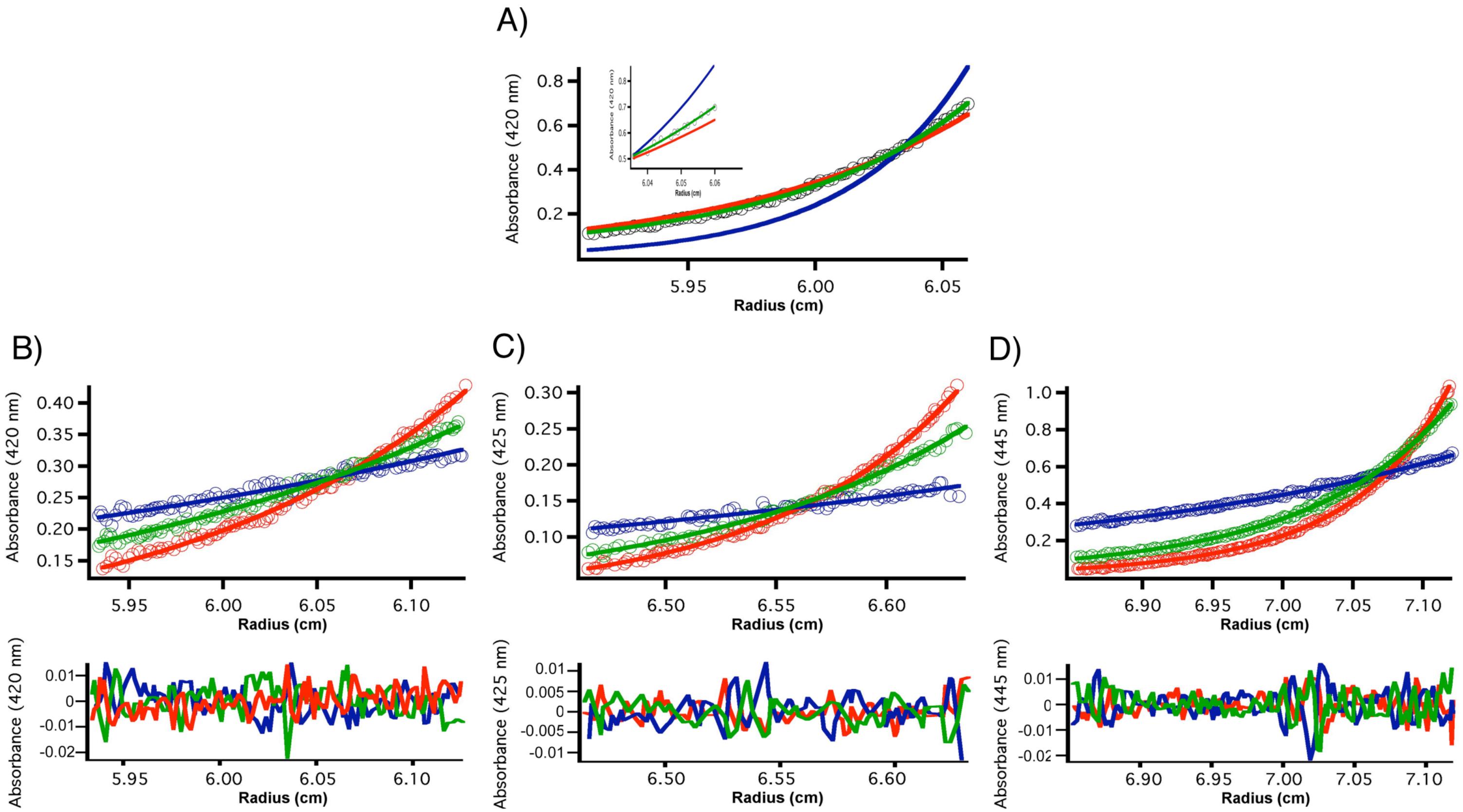


Figure 5.10. Sedimentation equilibrium data for RIDC-2 with 5 mM EDTA and 150 mM NaCl. A) Comparison of species analysis models used to globally fit SE data. Only data and fits from the 25,000 rpm scan of the 5 μM protein sample are shown (black circles) with fits from the single species monomer (red) and dimer (blue), and multi-species monomer-dimer (green) models superimposed. (inset) Close-up of the fits of the three different models. B) 5 μM protein at 15,000, 20,000, and 25,000 rpm detected at 420 nm. C) 20 μM protein at 15,000, 25,000, and 30,000 rpm detected at 425 nm. D) 40 μM protein at 15,000, 25,000, and 30,000 rpm detected at 445 nm. All samples were run in 20 mM TRIS buffer (pH 7), at 25° C, with 5 mM EDTA and 150 mM NaCl. Scans shown in B), C) and D) were globally fit to a monomer-dimer model yielding a minimized fit whose results and statistics are shown in Table 5.1.

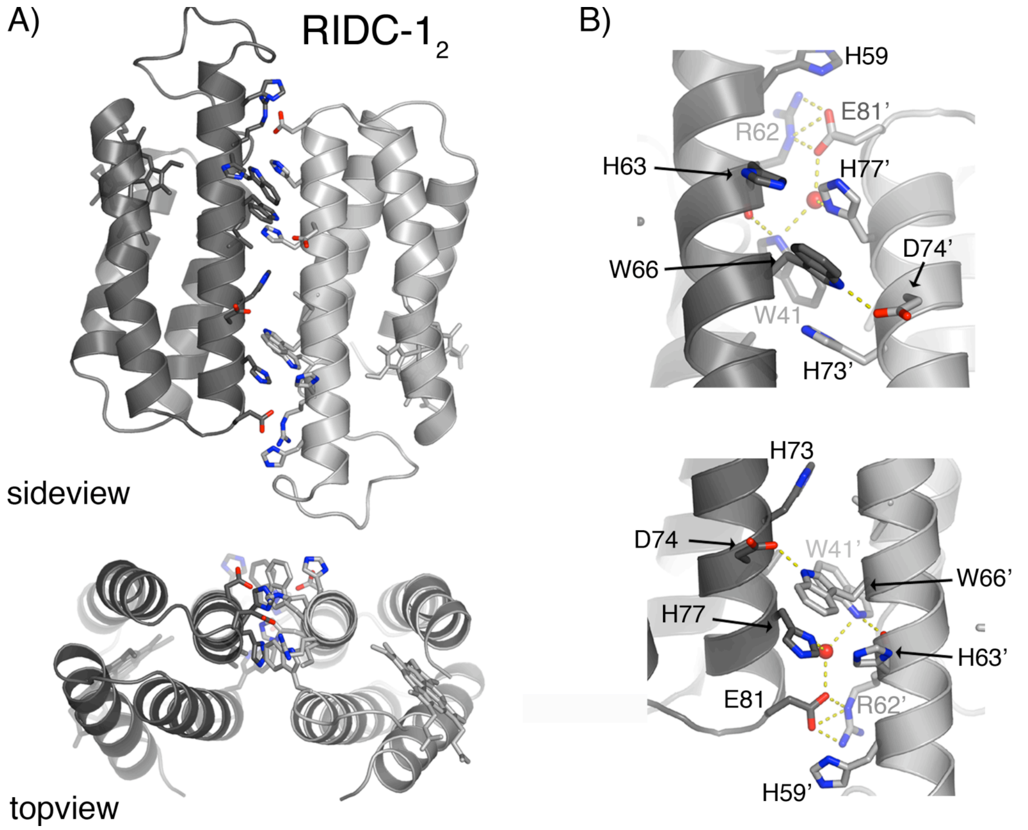
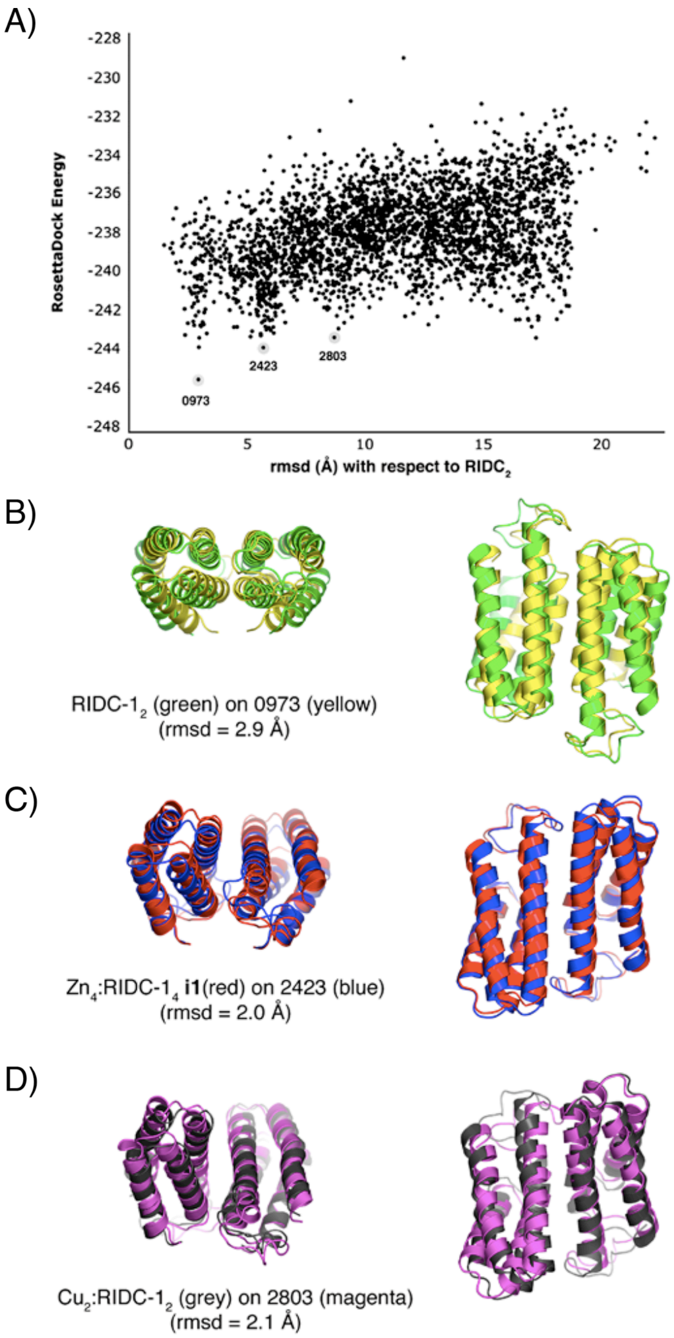


Figure 5.11. Crystal structure of the RIDC-1$_2$ dimer. A) Side and top views of the RIDC-1$_2$ crystal structure, whereby the interfacial residues are shown as sticks. B) Close-up views of the two (nearly) symmetrical interaction zones in the dimer interface detailing the hydrophobic and H-bonding contacts.

Overall, the modest size of this interface is consistent with the measured dissociation constants in the micromolar range, which are typical of transient protein complexes.[25,26] Based on the information available, such as the ionic-strength independence of dimerization despite the observed salt-bridging interactions, we propose that RIDC-1_2 in solution is a collection of several structurally similar conformers afforded by the fluidity of engineered hydrophobic interactions in i1. The crystal structure of RIDC-1_2 likely represents one of the most stable conformations favored over others through lattice-packing interactions. Docking simulations that were run prior to the determination of the RIDC-1_2 crystal structure corroborate these conclusions. The simulations run in the absence of metal coordination (Figure 5.12 and Table 5.4) reveal a shallow funnel toward the docking geometry observed in the RIDC-1_2 crystal structure, suggesting that the engineered hydrophobic interactions do not impose strong specificity, but that they do orient the monomers to form complexes that closely approximate the conformation induced by Zn coordination. Ultimately the observation that RIDC-1 can form a metal-independent dimer suggests that the contribution of i1 mutations to Zn_4:RIDC-1_4 stability is not only enthalpic, but also entropic. Dimerization of RIDC-1 halves the number of protein components toward tetramerization, while pre-organizing the Zn-coordinating residues (H63, H73’, and H77’) into close proximity.

Cu-mediated dimerization of RIDC-1. An implied outcome of the increased stability of the Zn induced tetramer due to interface redesign is an increased preference for the formation of tetrahedral Zn-coordination environment, which in turn should translate into greater Zn binding affinity/specificity. As mentioned previously, it is not possible to experimentally dissect metal-protein binding equilibria from protein–protein



RIDC-1$_2$ (green) on 0973 (yellow)
(rmsd = 2.9 Å)

Zn$_4$:RIDC-1$_4$ **i1**(red) on 2423 (blue)
(rmsd = 2.0 Å)

Cu$_2$:RIDC-1$_2$ (grey) on 2803 (magenta)
(rmsd = 2.1 Å)

Figure 5.12. Results from RosettaDock docking simulations (see also the corresponding Table 5.4). A) Correlation between the docking energy and deviation from the RIDC–12 dimer crystal structure for 3,000 decoys. B - D) Superpositions of RIDC–1$_2$, Zn$_4$:RIDC–1$_4$ and Cu$_2$:RIDC–1$_2$ crystal structures with calculated decoy structures. The rmsd values shown in (A) are with respect to the RIDC $-$ 1$_2$ structures, whereas those in (B), (C) and (D) are for the shown superpositions.

Table 5.4. Results from RosettaDock docking simulations. A corresponding correlation between RosettaDock energies and rms deviations from the RIDC–12 structure is shown in Figure 5.10.

Superpositions*

Structure 1	Structure 2	RMSD**	RosettaDock Energy
RIDC-1$_2$	**Zn$_4$:RIDC1$_4$**	6.164	N/A
RIDC-1$_2$	*0973****	2.894	-245.62
RIDC-1$_2$	*2423*	5.644	-243.98
RIDC-1$_2$	*2803*	8.651	-243.45
Zn$_4$:RIDC1$_4$	*0973****	5.384	-245.62
Zn$_4$:RIDC1$_4$	*2423*	1.998	-243.98
Cu$_2$:RIDC1$_2$	*2803*	2.090	-243.45

*Crystal structures are indicated by bold font and docking simulation decoys are indicated by italic font. The decoys 0973, 2423, 2803 presented in this table are highlighted in Figure 5.12, and respectively have energies 3.66, 2.87, and 2.61 standard deviations form the simulation mean.

**RMSD in Å over all α-Cs. In the case of tetrameric crystal structures, only the monomers forming the i1 interface are used in the superposition.

***0973.pdb is the lowest energy decoy of the simulation.

association equilibria that collectively lead to the formation of the tetramer. As an indirect means to assess the increase in Zn binding affinity/specificity due to interface redesign, we investigated the oligomerization behavior of RIDC-1 in the presence of Cu^{2+}, which prefers non-tetrahedral coordination geometries. Cu^{2+} was previously observed to induce the dimerization of MBPC-1 through the square-planar coordination of His59, His63, His73’, and His77’ (Figure 5.13 B). If the surface mutations to convert MBPC-1 to RIDC-1 indeed lead to increased preference towards Zn, this should be reflected in the distortion of the Cu^{2+} coordination environment in RIDC-1.

SV measurements indicate that Cu^{2+} binding leads to the exclusive formation of a dimer (Cu_2:RIDC-1_2) at all RIDC-1 concentrations used (5–600 μM) (Figure 5.13 A). The crystal structure of Cu_2:RIDC-1_2 (Species 7 in Fig. 1) was determined at 2.2 Å resolution. The crystallographic asymmetric unit of Cu_2:RIDC-1_2 crystals contains one dimer that displays a crisscrossed monomer– monomer alignment (Figure 5.13 B). This alignment is significantly distorted from the anti-parallel arrangement of monomers seen in Cu_2:MBPC-1_2 toward that observed across i1 in Zn induced tetramers. The Cu_2:RIDC-1_2 backbone can be superimposed with each dimeric half of Zn_4:RIDC-1_4 that contains the i1 interface with an overall rmsd of 1.4 Å (Figure 5.13 C) and features the same set of hydrophobic contacts in the interface (Figure 5.13 D). As a consequence of the crisscrossed arrangement of monomers in Cu_2:RIDC-1_2, His59 is now pushed out of the Cu coordination sphere, leaving only His63, His73’, and His77’ as the protein based ligands (Figure 5.14 A).

The two Cu coordination sites in the dimer interface display small differences. Cu1 exhibits a slightly distorted square-pyramidal geometry with one equatorial and one

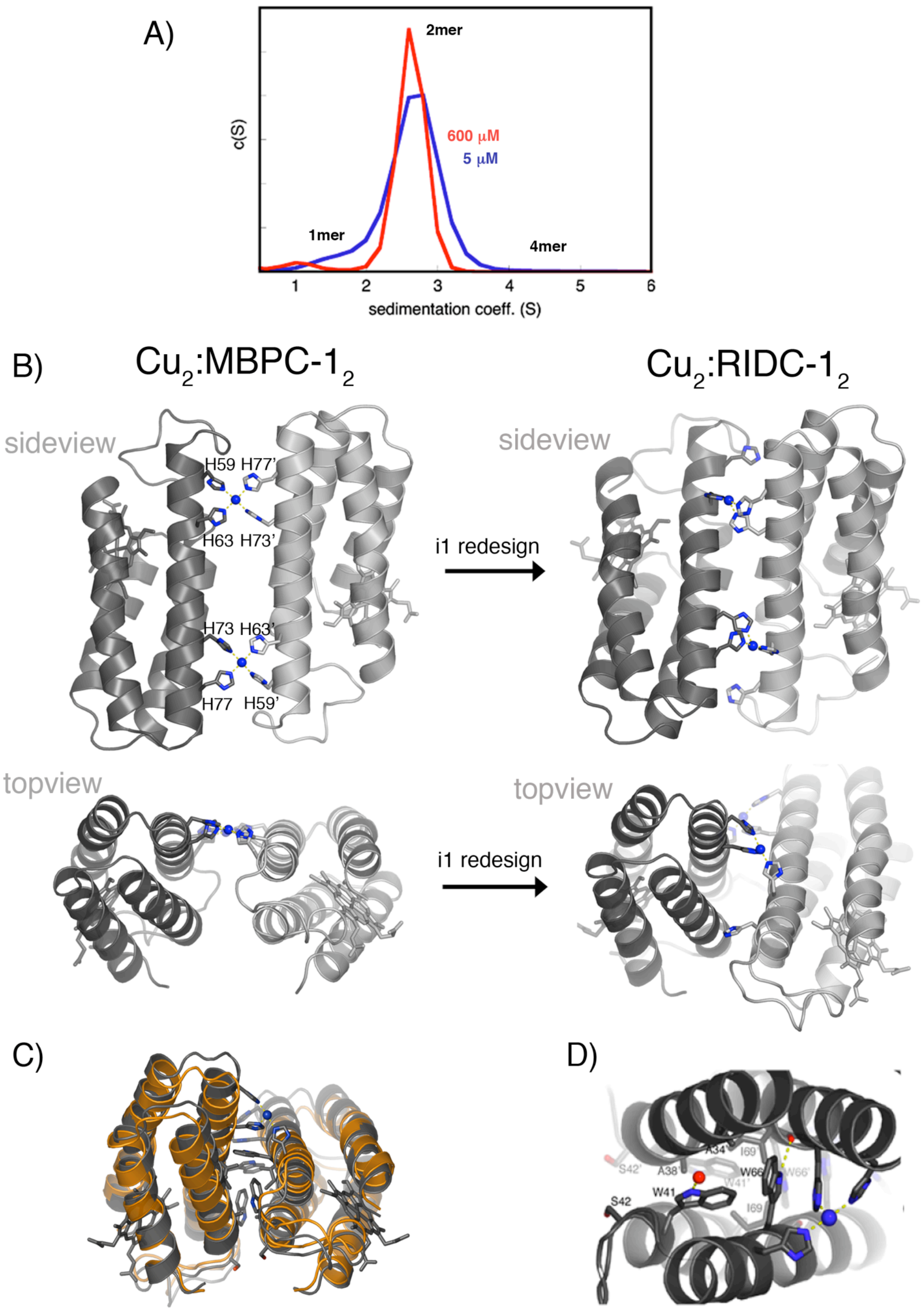


Figure 5.13. Cu^{2+}-mediated RIDC-1 dimer. A) Sedimentation coefficient distributions for 5 and 600 μM RIDC-1 in the presence of equimolar Cu^{2+}, showing the exclusive formation of a dimeric species in solution. B) The influence of Zn-templated interfacial mutations in i1 on the conformations of Cu-mediated dimeric assemblies. C) Backbone superposition of Cu_2:RIDC-1_2 (gray) and a dimeric half of Zn_4:RIDC-1_4 (orange) that contains i1. Interfacial residues of Cu_2:RIDC-1_2 are shown as sticks. D) i1 packing interactions in Cu_2:RIDC–1_2.

axial water molecule, both of which are H-bonded to Glu81 carboxylate from a symmetry-related dimer in the crystal lattice. Cu2, in contrast, is found in an approximately trigonal bipyramidal geometry, whereby the crystallographically related Glu81 side chain is now directly coordinated (Figure 5.14 B). Because SV studies indicate the exclusive formation of a dimer, we suggest that in solution the Cu centers in Cu_2:RIDC-1_2 adopt a square-planar or pyramidal geometry with three equatorial His ligands and one or two solvent molecules. This coordination geometry is supported by the axial EPR spectrum of Cu_2:RIDC-1_2 (Figure 5.14 C).

As previously discussed, the engineered hydrophobic interactions in i1 are sufficiently flexible that they conform to the preferred square-planar coordination geometry of Cu^{2+} without paying a large energetic penalty (Figure 5.12 D). Clearly the enforcement of tetrahedral or near-tetrahedral coordination geometries on Cu^{2+} as intended by Zn templating would require more extensive and specific non-covalent interactions. A fitting example in this case is provided by blue copper proteins that use a rigid network of non-covalent interactions (i.e., the entire protein fold) to produce the "rack-effect" or the "entatic state" for Cu.[27] Nevertheless, the redesigned surface in i1 provides sufficient driving force for the formation of the crisscrossed supra-molecular arrangement over the anti-parallel alignment seen in Cu_2:MBPC-1_2 forcing Cu to adopt three—instead of four—histidine ligands. Thus, the memory of Zn coordination engraved into non-covalent interactions through templating is still imposed on Cu, ultimately enforcing an unsaturated coordination environment.

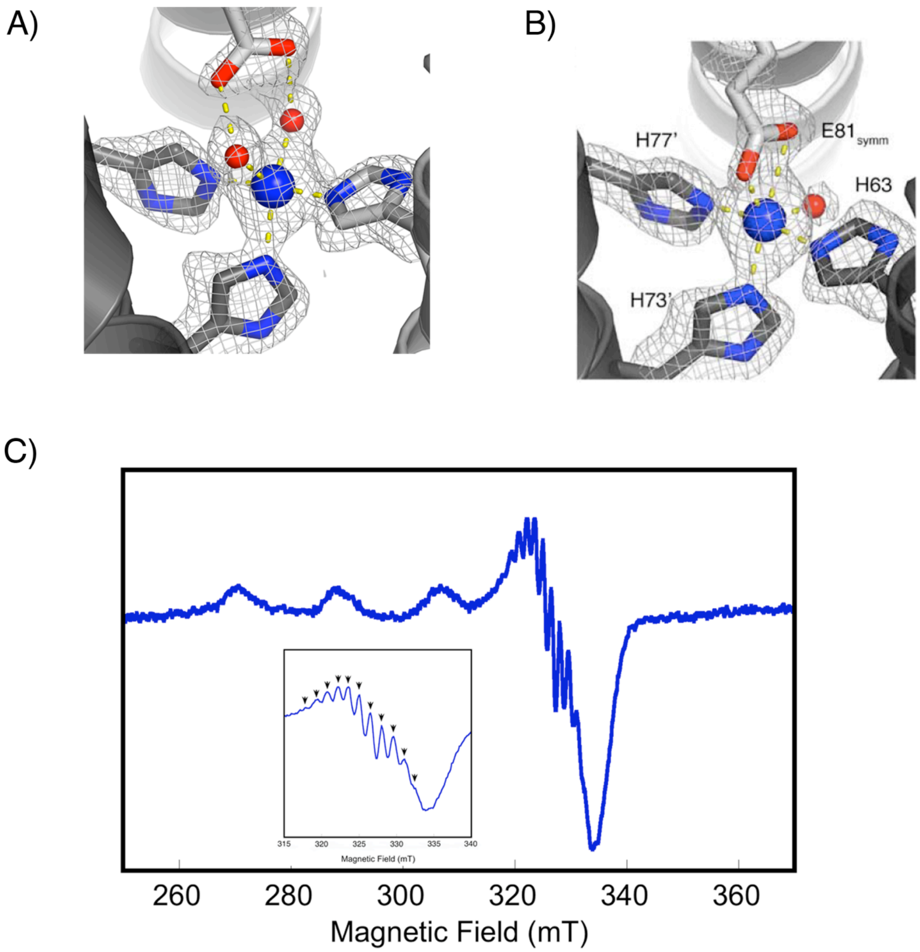


Figure 5.14. Binding mode of Cu in Cu_2:RIDC1_2. A) Cu coordination environment in Cu_2:RIDC-1_2 (Site 1), highlighting the open coordination sites occupied by two water molecules. The Glu81 side chain from a crystallographic symmetry-related dimer that forms H-bonds to the coordinated water molecules is shown in light gray. B) Cu coordination environment in Cu_2:RIDC–1_2 (Site 2), where the Glu81 side chain from a crystallographic symmetry related dimer is directly coordinated to Cu. The Fo-Fc omit electron density map in both (A) and (B) is contoured at 3σ. C) The X-band EPR spectrum of Cu_2:RIDC–1_2 (150 μM RIDC–1 + 100 μM Cu^{2+}) collected at 125 K. (inset) Close-up view of the g⊥ band, highlighting the superhyperfine splitting pattern. There are at least ten discernible superhyperfine peaks, whereas seven are expected for three equivalent nitrogen donors. We attribute this to a small degree of fluxionality and/or deviations from ideal square planar geometry in the Cu-coordination environment.

Conclusions

All biologically relevant transition metal ions form stable complexes with peptides, sometimes even when they lack metal coordinating side chain functionalities.[28] This is evident from the fact that unstructured polypeptides and folded proteins alike often form aggregates in the presence of high concentrations of transition metals. In modern-day organisms cellular metal concentrations are tightly regulated and the availability of free transition metal ions is generally low.[29] It is safe to assume, on the other hand, that during early evolution of folded proteins such advanced regulatory mechanisms did not exist. Therefore, the interactions between peptides/proteins and metals were likely under thermodynamic control and governed by environmental concentrations of soluble metal ions.[30] Under such conditions it would have been nearly impossible for any polypeptide chain to avoid being associated with metal ions. One possible outcome of such metal-polypeptide association reactions is precipitation. At the same time it is probable that some resulted in the formation of discrete soluble structures formed around metal-coordinating nuclei. If any such metal-nucleated structures inferred some benefit to an organism (e.g., generation of stable structures that did not deleteriously associate with existing cellular components, sequestration of essential or toxic metals, reactivity), they could have been subject to natural selection and evolution into more stable architectures or advanced functions. Our findings suggest that it is feasible for a non-self-associating protein surface to assemble into higher order architectures through a small number of mutations that enable metal coordination. Once the entropic cost of association is overcome, the resulting non-covalent interfaces can be optimized through additional mutations, in turn leading to more stable architectures that can form even in the absence of metal coordination. Given the lack of detailed information regarding the cellular environment during the early emergence of proteins, any such evolutionary model is speculative. Yet, the fact that a large fraction of known proteins contain metal ions as integral components and that many metal active sites are located in interfaces between secondary substructures or domains,[31] suggest that evolutionary pathways involving initial metal-mediated nucleation events are possible.

From a practical protein interface design viewpoint, MeTIR offers important advantages. In principle MeTIR can be implemented on any protein surface large enough for the incorporation of stable metal binding motifs. As demonstrated here, metal–protein interactions mediated by such motifs can be strong enough to hold together extensive—and originally repulsive— protein surfaces that are amenable to structural characterization and subsequent (and iterative) re-design to generate associative interactions. In this regard the utility of crystal structures as a starting point for designing protein structures and interfaces has been well documented.[32,33] The most successful *de novo* interface design efforts to date have focused on coiled-coil motifs with pre-determined docking orientations and knowledge-based energy functions.[34,35] In contrast MeTIR is not restricted to any particular type of protein interface or a particular fold for the partners in the complex; the fact that the protein building blocks used in this study are helical bundles is coincidental and purely due to practical reasons (stability, solubility, crystallizability). While the dependence on crystallographic information can be regarded as a limitation, MeTIR represents a promising step toward the ultimate goal of designing arbitrary protein interfaces from scratch.

From a practical inorganic chemical viewpoint our results show that individual proteins can be utilized as large polydentate ligands that bring along the advantage of having extensive functionalizable surfaces. While the complex nature of protein surfaces renders the control of metal coordination challenging, it allows the tuning of the metal coordination environment through distant non-covalent interactions. Such interactions, after all, account for the exquisite control of metal selectivity and reactivity in natural metalloproteins and enzymes through the formation of an extensive three-dimensional bonding network.

Chapter 5 is reproduced in part from: Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B., Tezcan, F. A. **2010**. Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. USA*. 107, 1827-1832.

References

1. Soding, J., Lupas, A. N. **2003**. More than the sum of their parts: On the evolution of proteins from peptides. *BioEssays*. 25, 837–846.

2. Bogarad, L. D., Deem, M. W. **1999**. A hierarchical approach to protein molecular evolution. *Proc Natl Acad Sci USA*. 96, 2591–2595.

3. Andre, I., Strauss, C. E. M., Kaplan, D. B., Bradley, P., Baker, D. **2008**. Emergence of symmetry in homooligomeric biological assemblies. *Proc Natl Acad Sci USA*. 105, 16148–16152.

4. Keefe, A. D., Szostak, J. W. **2001**. Functional proteins from a random-sequence library. *Nature*. 410, 715–718.

5. Riechmann, L., Winter, G. **2006**. Early protein evolution: Building domains from ligand binding polypeptide segments. *J Mol Biol*. 363, 460–468.

6. Salgado, E. N., Faraone-Mennella, J., Tezcan, F. A. **2007**. Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* 129, 13374-13375.

7. Salgado, E. N., Lewis, R. A., Faraone-Mennella, J., Tezcan, F. A. **2008**. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* 130, 6082–6084.

8. Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., Tezcan, F. A. **2009**. Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical assemblies through Cu^{II} and Ni^{II} coordination. *Inorganic Chemistry*. 48, 2726-2728.

9. Lee, B., Richards, F. M. **1971**. Interpretation of protein structures—Estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.

10. Wallace, A. C., Laskowski, R. A., Thornton, J. M. **1995**. Ligplot—A program to generate schematic diagrams of protein ligand interactions. *Prot. Eng.* 8, 127–134.

11. Vriend, G. **1990**. What if—A molecular modeling and drug design program. *J. Mol. Graph*. 8, 52.

12. Liu, Y., Kuhlman, B. **2006**. RosettaDesign server for protein design. *Nucl. Acids. Res*. 34, W235–238.

13. Ambroggio, X. I., Kuhlman, B. **2006**. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* 128, 1154–1161.

14. Leaver-Fay, A., Butterfoss, G. L., Snoeyink, J., Kuhlman, B. **2007**. Maintaining solvent accessible surface area under rotamer substitution for protein design. *J. Comp. Chem.* 28, 1336–1341.

15. Gray, J. J., al., e. **2003**. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* 331, 281–299.

16. Schuck, P. **2000**. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* 78, 1606-1619.

17. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A., Schuck, P. **2004**. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Analytical Biochemistry* 326, 234-256.

18. Vagin, A., Teplyakov, A. **1997**. MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.

19. Murshudov, G. N., Vagin, A. A., Dodson, E. J. **1997**. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.D.* 53, 240-255.

20. Murshudov, G. N., Vagin, A. A., Dodson, E. J. **1994**. The CCP4 Suite: Programs for protein crystallography. *Acta Cryst. D*. 50, 760-763.

21. McRee, D. E. **1992**. A visual protein crystallographic software system for X11/XView. *J. Mol. Graphics*. 44-46.

22. DeLano, W. L. *The PYMOL molecular graphics system (http://www.pymol.org)*, 2003.

23. Kyte, J., *Structure in Protein Chemistry*. Garland Science: New York, 2007.

24. Krissinel, E., Henrick, K. **2007**. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 774–797.

25. Janin, J., Miller, S., Chothia, C. **1998**. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* . 204, 155-164.

26. Nooren, I. M. A., Thornton, J. M. **2003**. Structural characterisation and functional significance of transient protein–protein interactions. *J. Mol. Biol.* 325, 991–1018.

27. Gray, H. B., Malmstrom, B. G., Williams, R. J. P. **2000**. Copper coordination in blue proteins. *J. Biol. Inorg. Chem.* 5, 551–559.

28. Martell, A. E., Smith, R. M., *Critical Stability Constants*. Plenum Press: New York, 1974–1989.

29. Finney, L. A., O'Halloran, T. V. **2003**. Transition metal speciation in the cell: Insights from the chemistry of metal ion receptors. *Science*. 300, 931–936.

30. Frausto da Silva, J. J. R., Williams, R. J. P., *The Biological Chemistry of the Elements*. Oxford Univ. Press: Oxford, 2001.

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

32. Joachimiak, L. A., Kortemme, T., Stoddard, B. L., Baker, D. **2006**. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein–protein interface. *J. Mol. Biol.* 361, 195–208.

33. Bryson, J. W., Desjarlais, J. R., Handel, T. M., DeGrado, W. F. **1998**. From coiled coils to small globular proteins: Design of a native-like three-helix bundle. *Protein Sci.* 7, 1404–1414.

34. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., Kim, P. S. **1998**. High-resolution protein design with backbone freedom. *Science*. 282, 1462–1467.

35. Grigoryan, G., Reinke, A. W., Keating, A. E. **2009**. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature*. 458, 859–864.

Chapter 6

Formation of a Synthetic Duplicate-Domain Protein: Covalent Stabilization of a Metal-Templated Protein Tetramer

Introduction

Thorough examination of proteins at both the genetic and structural levels have allowed for a number of theories describing possible evolutionary mechanisms through which functional single proteins, oligomeric assemblies, and protein families may have arisen.[1-3] While the lack of a fossil record of primordial proteins makes the direct observation of evolutionary trajectories described by these theories impossible, observed patterns in both modern day proteins and designed proteins/peptides can serve as a qualitative means by which we can begin to understand how twenty amino acids have been adapted to create architectures of such varying structures and functions.

Our work involving Metal-Templated Interface Redesign (MeTIR)[4] adds to these ideas by suggesting a possible mechanism through which modern proteins could have evolved their current PPIs concomitantly with structural or functional metal centers. Briefly, an initial minimal set of random mutations along the surface of a pre-existing protein results in the coordination of a metal ion present in the surrounding environment. This coordination event precipitates an interaction between individual molecules, which, through later random mutations made along the newly formed interaction interfaces, yields a complex that is more thermodynamically stable than the parent assembly. Further mutations could then serve to rigidify, functionalize, or completely eliminate the metal center. In the first two cases, a resulting reinforcement of geometric restraints at the metal center would likely lead to increased metal selectivity. While speculative, this concept provides a reasonable evolutionary pathway and exists as a possible method through which researchers could create new and functional protein assemblies.

We sought to further “evolve” our MeTIR derived proteins by employing other previously proposed concepts of protein evolution. To this end, we set out to utilize the ideas of gene duplication and fusion,[5-7] suggested to be the evolutionary mechanism behind such naturally occurring enzymes as yeast extracellular glucoamylase,[8] human glyoxalase I,[9] and aspartate/ glutamate racemaces.[10] As an evolutionary tool, the concept of gene duplication is useful as it can effectively relieve selective pressure on the gene, allowing evolution of new functionalities, sub-functionalities, or often times inactivation to a pseudogene, while minimizing the effects of deleterious mutations that may occur. Further mutations to the intron DNA separating duplicate genes on the same chromosome could then result in a translated sequence serving as a protein linker (gene fusion) between what are now two domains of a single protein. From here, further mutations can occur at each domain independently, vastly increasing the chances of evolving multiple functionalities from similar genetic sequences and protein architectures.[11,12]

Towards the end goal of creating a similar duplicate-domain protein, we constructed a mutant of the previously described RIDC-1 protein[4] containing a cysteine (Cys) residue at position 82, (^{C82}RIDC-1), that we have crosslinked using a Cys specific bis(maleimido)ethane (BMOE) moiety. Initial experiments involving the resulting crosslinked construct, ^{C82}RIDC-$1_{2,BMOE}$, suggest that our artificial duplicated and fused gene product is capable of fully forming a Zn^{2+}-templated assembly structurally identical to the non-crosslinked RIDC-1 tetramer (Zn_4:RIDC-1_4), at lower protein concentrations than the parent mutant. We further demonstrate that the use of longer crosslinkers does not appear to affect the formation of this assembly.

Similar early evidence also suggests that ^{C82}RIDC-$1_{2,BMOE}$ has an increased affinity for binding Zn^{2+} over other divalent metal ions, and demonstrates a possible selectivity for Zn over Cu^{2+}. Elucidation of a unique, D_2 symmetric tetramer apparently mediated by the binding of a sulfate anion, implies that we may have not only imparted this preference to bind Zn, but a general specificity for molecules with tetrahedral geometries. These data highlight how we can apply theories derived from naturally evolved proteins on to our engineered systems in order to expand their functionalities even beyond metal binding.

Materials and Methods

Site Directed Mutagenesis and Protein Expression/Purification/Characterization. Site directed mutagenesis was performed on the pETc-b562 plasmid (denoted as wildtype), as above, and with sequencing performed by Retrogen, Inc. Mutant plasmids were transformed into BL21(DE3) *E. coli* cells along with the *ccm* heme maturation gene cassette plasmid, pEC86. Expression was performed as previously described.

Mutant-expressing cells were then sonicated for 5 min. (30 sec pulse at 50% amplitude with 1 min rest between pulses), brought to pH 9 with the addition of NaOH and centrifuged at 16,000 g, 4°C, for 10 min. Multiple cycles of addition of small volumes of HCl followed by centrifugation at 16,000 g, 4°C, for 10 min. were performed until the pH was brought to 5. The protein was then purified by ion-exchange chromatography on a CM-Sepharose matrix (Amersham Biosciences) using a NaCl gradient in 5 mM sodium acetate, pH 5. After exchange into 10 mM sodium phosphate, pH 8, the protein was further purified using an Uno-Q (BioRad) anion exchange column on a DuoFlow chromatography workstation (BioRad) using a NaCl gradient. Purity was determined by SDS-PAGE gel electrophoresis. Verification of mutations was made through MALDI mass spectrometry.

Maleimide crosslinking of ^{C82}*RIDC-1.* Purified protein was buffer exchanged in to 20 mM TRIS (pH 7) and 50 mM dithiothreitol (DTT) and concentrated. Concentrated protein was run through a 10 DG de-salting column (BioRad) and collected in an appropriate volume of 20 mM TRIS (pH 7) to yield a protein concentration of 100 μM. A final concentration of 100 μM BMOE, bis(maleimide) butane (BMB), or bis(maleimide) hexane (BMH) (Pierce), was added to the sample. Crosslinker addition began as protein eluted from the column at 1/10th of the final volume of crosslinker to be added. Nine more additions of the crosslinker, at the same 1/10th volume, was added every 30 sec. for 2.5 min. The reaction was then allowed to proceed at RT, with stirring, for 30 min. Crosslinking reactions were quenched by the addition of 50 mM DTT and allowed to incubate at RT for 15 min. Protein used for crystallizing Zn_4:^{C82}RIDC-$1_{2,BMOE}$ with $CuSO_4$ was crosslinked in the presence of 100 μM $ZnCl_2$. Roughly 50% of the initially purified ^{C82}RIDC-1 protein (yields in the 50–100 mg range), was successfully crosslinked with yields varying depending on crosslinker length and speed of elution/ crosslinker addition.

Crosslinked protein (^{C82}RIDC-$1_{2,BMOE}$; ^{C82}RIDC-$1_{2,BMB}$; ^{C82}RIDC-$1_{2,BMH}$), was then concentrated and purified through multiple SEC runs in 20 mM TRIS (pH 7) with 150 mM NaCl until all monomeric protein was separated out. Dimer purity was assessed by SDS-PAGE, while appropriate constructs were verified by MALDI mass spectrometry. Final protein was incubated with 50 mM EDTA overnight at 4°C followed by buffer exchange with a 10 DG de-salting column in to 20 mM TRIS (pH 7). Protein crosslinked in the presence of $ZnCl_2$ was not treated with EDTA at any time, but was dialyzed in 3 × 1 L of 20 mM TRIS (pH 7).

Sedimentation Velocity. SV experiments were performed in order to determine the solution-state oligomerization behavior of ^{C82}RIDC-$1_{2,BMOE}$, ^{C82}RIDC-$1_{2,BMB}$, and ^{C82}RIDC-$1_{2,BMH}$. All metal containing SV experiments were performed in 20 mM TRIS (pH 7.0) with 2.5 μM crosslinked dimer with a final metal concentration of 5 μM (2:1 metal:crosslinked protein ratio). All metal free samples were run in the same buffer and protein concentration with a final EDTA concentration of 5 mM.

SV measurements were made on a Beckman XL-I Analytical Ultracentrifuge (Beckman-Coulter Instruments) using an An-60 Ti rotor at 41,000 rpm for a total of 250 scans per sample. Data was collected at 415 nm. Vbar for each mutant was calculated assuming a Vbar of heme of 0.82 mg/ml and a Vbar for the crosslinker calculated as in Durchschlag and Zipper.[13] All data were processed in SEDFIT[14] with the following fixed parameters: buffer density (r) = 0.99764 g/ml; buffer viscosity = 0.0089485 poise; Vbar, which was calculated to be 0.7310 ml/g for ^{C82}RIDC-$1_{2,BMOE}$ and ^{C82}RIDC-$1_{2,BMB}$; 0.7306 ml/g for ^{C82}RIDC-$1_{2,BMH}$.

Crystallography. All crystals were obtained by sitting drop vapor diffusion at room temperature. The crystallization conditions for the five different crystal forms described in this study are outlined in Table 6.1 to Table 6.3, along with sites and wavelengths of collection, collection statistics, and refinement statistics. All protein was stored in 20 mM

Table 6.1. Crystallization conditions, X-ray data collection, and refinement statistics for SO_4:^{C82}RIDC-$1_{2,BMOE}$.

[‡] $Rsym = \Sigma\Sigma j|Ij - \langle I \rangle| / \Sigma\Sigma j|Ij|$.

[§] $R = \Sigma||Fobs| - |Fcalc|| / \Sigma|Fobs|$ (2 σ cutoff).

[II] Free R calculated against 7% of the reflections removed at random.

[¶] Root mean square deviations from bond and angle restraints.

[*] Numbers in parentheses correspond to the highest resolution shell, 2.80 Å.

	SO_4:^{82C}RIDC-$1_{2,BMOE}$
Mother liquor	2 M ammonium sulfate with 5 mM EDTA
Concentration of protein	1.66 mM
μl protein : μl mother liquor	1:1
X-ray source	SSRL BL 9-2
Residues in complex	4 x (106 + 1 Heme)
No. of complexes / asymmetric unit	1
Metal ions in asymmetric unit	N.A.
Waters in asymmetric unit	107
Unit cell dimensions (Å)	a = b = c =172.544
	α = β = γ = 90
Symmetry group	*P*$4_3$32
Resolution (Å)	70.44 - 2.80
X-ray wavelength (Å)	0.9795
Number of Unique Reflections	21053
Redundancy	13
Completeness (%)*	99.83 (99.94)
⟨I / σI⟩*	6.1 (2.0)
Rsym[‡] (%)*	10.6 (38.9)
R[§] (%)*	19.1 (26.7)
Free R[II] (%)*	24.0 (31.0)
Rms Bnd[¶] (Å)	0.005
Rms Ang[¶] (°)	0.728
Ramachandran plot (%)	
Residues in most favored regions	95.9
Residues in add.l allowed regions	4.1
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0

Table 6.2. Crystallization conditions, X-ray data collection, and refinement statistics for Zn$_4$:^{C82}RIDC-1$_{2,BMOE}$ and Zn$_4$:^{C82}RIDC-1$_{2,BMOE}$ - Cu.

‡Rsym= $\Sigma\Sigma j|\, I_j - \langle I \rangle | / \Sigma\Sigma j|\, I_j|$.

§R= $\Sigma|\,|F_{obs}| - |F_{calc}|\,| / \Sigma|F_{obs}|$ (2 σ cutoff).

IIFree R calculated against 7% of the reflections removed at random.

¶Root mean square deviations from bond and angle restraints.

*Numbers in parentheses correspond to the highest resolution shell (2.30 Å for Zn$_4$:^{C82}RIDC-1$_{2,BMOE}$; 2.64 Å for Zn$_4$:^{C82}RIDC-1$_{2,BMOE}$ – Cu).

	Zn$_4$:^{82C}RIDC-1$_{2,BMOE}$	**Zn$_4$:^{82C}RIDC-1$_{2,BMOE}$ - Cu**
Mother liquor	100 mM HEPES pH 7.5, 10% PEG 3350, 4.6 mM ZnCl$_2$	100 mM HEPES pH 7.5, 12% PEG 3350, 2.46 mM CuSO$_4$
Concentration of protein	1.15 mM	2.46 mM
μl protein : μl mother liquor	2:1	1:1
X-ray source	Stanford Synchrotron Radiation Laboratory (SSRL) BL 9-2	SSRL BL 7-1
Residues in complex	4 x (106 + 1 Heme) + 4 Zn	4 x (106 + 1 Heme) + 4 Zn
No. of complexes / asymmetric unit	1	1
Metal ions in asymmetric unit	11 Zn	5 Zn, 7 Cu
Waters in asymmetric unit	39	13
Unit cell dimensions (Å)	a = b =52.088, c = 253.943	a = b = 52.53, c = 255.72
	$\alpha = \beta = 90$, $\gamma = 120$	$\alpha = \beta = 90$, $\gamma = 120$
Symmetry group	$P6_1$	$P6_1$
Resolution (Å)	45.11 - 2.30	45.59 - 2.64
X-ray wavelength (Å)	0.9795	1.265 (near Zn K edge) 1.377 (Cu K edge)
Number of Unique Reflections	15859	10772
Redundancy	11.1	15.8
Completeness (%)*	99.2 (97.3)	99.4 (97.3)
$\langle I / \sigma I \rangle$*	10.6 (3.1)	3.8 (1.8)
Rsym‡ (%)*	5.2 (24.8)	12.0 (36.2)
R§ (%)*	20.0 (21.2)	17.3 (15.7)
Free RII (%)*	24.9 (34.3)	21.0 (21.2)
Rms Bnd¶ (Å)	0.018	0.011
Rms Ang¶ (°)	0.776	0.663
Ramachandran plot (%)		
Residues in most favored regions	94.6	92.9
Residues in add.l allowed regions	5.1	7.1
Residues in generously allowed regions	0.3	0.0
Residues in disallowed regions	0.0	0.0

Table 6.3. Crystallization conditions, X-ray data collection, and refinement statistics for Zn$_4$:^{C82}RIDC-1$_{2,BMB}$ and Zn$_4$:^{C82}RIDC-1$_{2,BMH}$.

[‡] $Rsym = \Sigma\Sigma j |\, Ij - \langle I \rangle | / \Sigma\Sigma j |\, Ij |$.

[§] $R = \Sigma | |Fobs| - |Fcalc| | / \Sigma |Fobs|$ (2 σ cutoff).

[II] Free R calculated against 7% of the reflections removed at random.

[¶] Root mean square deviations from bond and angle restraints.

[*] Numbers in parentheses correspond to the highest resolution shell (1.84Å for Zn$_4$:^{C82}RIDC-1$_{2,BMB}$; 2.70 Å for Zn$_4$:^{C82}RIDC-1$_{2,BMH}$.).

	Zn$_4$:^{82C}RIDC-1$_{2,BMB}$	**Zn$_4$:^{82C}RIDC-1$_{2,BMH}$**
Mother liquor	100 mM Bis-TRIS pH 6.5, 14% PEG 3350, 2.84 mM ZnCl$_2$	50 mM Bis-TRIS pH 6.5, 20% pentaerythritol ethoxylate [15/4 EO/OH], 200 mM ammonium sulfate, 2.12 mM ZnCl$_2$
Concentration of protein	710 μM	1.07 mM
μl protein : μl mother liquor	2:1	1:1
X-ray source	Bruker Apex II CCD detector and monochromatized Cu-Kα radiation (Siemens sealed tube source)	SSRL BL 9-2
Residues in complex	4 x (106 + 1 Heme) + 4 Zn	4 x (106 + 1 Heme) + 4 Zn
No. of complexes / asymmetric unit	1	1
Metal ions in asymmetric unit	4	4
Waters in asymmetric unit	470	51
Unit cell dimensions (Å)	63.626 x 76.399 x 93.265	48.298 x 61.843 x 70.087
	$\alpha = \beta = \gamma = 90^o$	$\alpha = \gamma = 90^o$, $\beta = 102.22$
Symmetry group	$P2_12_12_1$	$P2_1$
Resolution (Å)	22.65 - 1.84	47.20 - 2.70
X-ray wavelength (Å)	1.542	0.9795
Number of Unique Reflections	37037	10569
Redundancy	4.6	3.7
Completeness (%)*	99.4 (97.9)	98.8 (99.8)
$\langle I / \sigma I \rangle$*	14.55 (3.19)	2.5 (6.5)
Rsym[‡] (%)*	5.7 (32.4)	9.3 (10.4)
R[§] (%)*	19.0 (26.2)	25.5 (31.3)
Free R[II] (%)*	24.8 (35.3)	30.4 (39.9)
Rms Bnd[¶] (Å)	0.024	0.016
Rms Ang[¶] (°)	1.87	1.50
Ramachandran plot (%)		
Residues in most favored regions	96.9	92.3
Residues in add.l allowed regions	3.1	7.4
Residues in generously allowed regions	0.0	0.3
Residues in disallowed regions	0.0	0.0

Table 6.4. PISA server output (Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html))[15] for the analysis of the protein-protein interfaces formed in SO_4:^{C82}RIDC-$1_{2,BMOE}$. The interactions of residues involved in interfacial salt-bridges in each assembly are highlighted in red.

SO_4:RIDC-$1_{2,BMOE}$

Interface between chains	Residue/Atom	Dist (Å)	Residue/Atom
A-C	A:HIS 77[ND1]	2.55	C:GLU 81[OE1]
A-D	A:GLU 86[OE1]	2.92	D:HIS 63[NE2]
	A:GLU 86[OE2]	3.23	D:HIS 63[NE2]
	A:GLU 81[OE1]	2.88	D:TRP 66[NE1]
	A:HIS 63[NE2]	2.51	D:GLU 86[OE1]
	A:HIS 63[NE2]	3.43	D:GLU 86[OE2]
	A:TRP 66[NE1]	3.21	D:GLU 81[OE2]
B-C	B:GLU 86[OE1]	3.00	C:HIS 63[NE2]
	B:GLU 86[OE2]	3.01	C:HIS 63[NE2]
	B:GLU 81[OE2]	2.96	C:TRP 66[NE1]
	B:HIS 63[NE2]	3.12	C:GLU 86[OE1]
	B:HIS 63[NE2]	3.02	C:GLU 86[OE2]
	B:TRP 66[NE1]	3.26	C:GLU 81[OE2]
B-D	B:HIS 77[ND1]	2.71	D:GLU 81[OE1]
A-SO_4	A:HIS 77[NE2]	3.65	G:SO_4 13[O4]
B-SO_4	B:HIS 77[NE2]	3.54	G:SO_4 13[O4]
C-SO_4	C:HIS 77[NE2]	3.06	G:SO_4 13[O4]
D-SO_4	D:HIS 77[NE2]	3.16	G:SO_4 13[O4]

TRIS (pH 7). Appropriate crystals were transferred to a solution of mother liquor containing 20% glycerol as a cryoprotectant and frozen in liquid nitrogen prior to data collection at 100 K.

Data were integrated using MOSFLM[16, 17] and scaled in SCALA,[17] except for Zn_4:^{C82}RIDC-$1_{2,BMB}$, which was processed using SAINT and Bruker SADABS. All structures were determined through molecular replacement with MOLREP[17,18], using the RIDC-1 monomer structure (PDB ID:3HNI) as the search model. Rigid-body, positional, thermal and TLS refinement with REFMAC[17,19] using appropriate non-crystallographic symmetry restraints. The Zn_4:^{C82}RIDC-$1_{2,BMOE}$ and Zn_4: ^{C82}RIDC-$1_{2,BMOE}$ co-crystallized with $CuSO_4$ were found to have 48% and 44% twin fractions, respectively. Both data sets were refined using intensity based twin refinement as implemented in REFMAC. All models were manually rebuilt, along with placement with waters, in COOT, producing the final models. All figures were produced with PYMOL.[20]

Mag-Fura-2 Competition Assays. 1 mg Mag-Fura-2 (Invitrogen) was resuspended in 1 ml Millipore purified H_2O. Concentration was determined by absorbance at 369 nm.[21] Ni^{2+} and Co^{2+} control titrations were performed in a 1-cm cuvette with a 2 ml sample volume containing 0.5 μM Mag-Fura-2 and 1 mM $CaCl_2$ in 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS, pH 7) with 150 mM NaCl. The sample was allowed to equilibrate with stirring for 3 min. after addition of the competing metal before recording excitation scans with emission at 505 nm. Zn^{2+} control experiments were performed with 1 mM $NiSO_4$ in place of $CaCl_2$. Data collected at an excitation wavelength of 330 nm were then fit to a 1 site binding model using Dynafit[22] assuming a Mag-Fura-2 Ca^{2+} K_d of 25 μM, which the manufacturer has found to be constant over a wide variety of pH and ionic strength conditions.[21] The resulting K_d values are listed in Table 6.5.

^{C82}RIDC-$1_{2,BMOE}$ competition assays were performed in a 1-cm cuvette with a 1.5 ml sample volume containing 25 μM dimer and 10 μM Mag-Fura-2 in MOPS (pH 7) with 150 mM NaCl. As with the controls, samples were allowed to equilibrate, with stirring, for 3 min, before recording excitation scans with emission at 505 nm. Ni^{2+} and Co^{2+} bound to Mag-Fura-2 result in quenching of the florescence signal, which was subsequently followed at an excitation wavelength of 372 nm and fit within Dynafit. Zn^{2+} binding results in an increased fluorescence at 323 nm, which was used for the final fit. All data fit to a (4 × 1) binding model consisting of four binding sites with equivalent K_d values listed in Table 6.5.

Results and Discussion

Covalent crosslinking of Zn_4:RIDC-1_4 i3 interface. Application of MeTIR to the Zn_4:MBPC-1_4 tetramer attempted to stabilize this Zn-templated oligomer. The resulting re-designed constructs, RIDC-1 and RIDC-2, were found to form the desired tetrameric assembly more readily, with RIDC-2 demonstrating full conversion to tetramer at 5 μM Zn/protein, and RIDC-1 forming ~50% tetramer at these concentrations. This phenomenon can be attributed to both a decreased entropic factor derived from the ability of these variants to form a metal independent dimer, as well as an increased enthalpic contribution from more favorable interactions along i1 and i2, respectively.

We believed that application of the ideas of gene duplication and fusion to RIDC-

1 would lead to a new construct similarly capable of full conversion to the Zn-mediated tetramer at low concentrations due to the much larger entropic push from a fully populated apo-dimer, even without the inclusion of the re-designed i2 interface. Another possible consequence of the duplicate domain variant could be the formation of an apo-tetramer driven by inter-dimer interactions along i1, allowing for pre-formation of a tetrahedral Zn binding site which could lend this assembly a degree of Zn binding preference, as well as serve as a stable platform for functionalization of the metal center.

Ideally, accomplishing this goal would mean the design of a short loop region connecting N- and C- terminal ends of neighboring protomers in the tetrameric structure. Unfortunately, as demonstrated in Figure 6.1 A, these termini are far too separated to make this design feasible. Instead, we sought to employ tools derived from organic chemistry to cross-link two monomers across one of the interfaces of the tetramers, accomplishing not only the dimerization of protomers, but also inherently stabilizing the interface.

Introduction of an organic crosslinker across the i2 interface runs the risk of distorting its spatial arrangement and thus disturbing the tetrameric complex. The i3 interface, however, is small (~490 $Å^2$ of buried surface area) and lies at the vertex of the V-shaped sub-structure formed by the head-to-head arrangement of monomers (Figure 6.1). Introduction of a crosslinker at the vertex of the V's is the least likely place to disrupt interactions along the PPI interfaces, while also allowing for close spatial positioning of crosslinker reactive residues. The best candidate for placement of these reactive residues was determined, by inspection of Zn_4:RIDC-1_4, to be position 82 as it is located at the top of the vertex, but not directly in the i3 interface (Figure 6.1 A, magenta

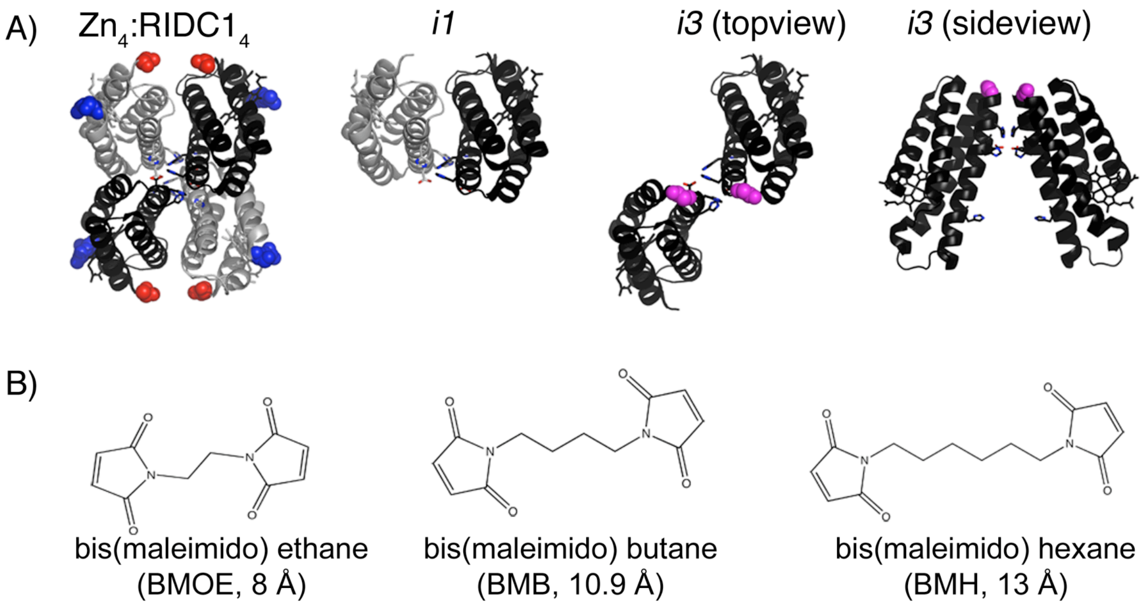


Figure 6.1. Optimal i3 crosslinking site of Zn$_4$:RIDC-1$_4$. A) Arrangement of the Zn$_4$:RIDC-1$_4$ tetramer, with the N- and C- termini highlighted as spheres in red and blue, respectively. Both the i1 and i3 interfaces are shown, with Gly82 highlighted in the latter as magenta spheres. B) Cys reactive maleimide crosslinkers available form Pierce.

spheres).

Mutation of the native Gly82 to Cys would result in a site that can be specifically crosslinked, as there are no surface exposed Cys residues in the rest of the protein. Simple modeling of this mutation along with the 8 Å long, Cys specific BMOE moiety (Figure 6.1 B) suggests a perfect fit of the crosslinker at this site. We therefore pursued the generation of this construct, ^{C82}RIDC-$1_{2,BMOE}$. As similar compounds with longer linker arms exist, we also chose to examine crosslinking of RIDC-1 with BMB (10.9 Å) and BMH (13 Å), yielding the constructs ^{C82}RIDC-$1_{2,BMB}$ and ^{C82}RIDC-$1_{2,BMH}$.

Formation of a stable ^{C82}RIDC-$1_{2,BMOE}$ dimer. The first question that arises after chemically crosslinking ^{C82}RIDC-1 monomers is whether or not this process allows for the pre-formation of a metal independent tetramer. As described in Figure 6.2 A, there is in only a small amount of protein forming a species larger than a dimer at low concentrations with all three constructs. Increasing the concentration of protein in the sample from 2.5 to 25 μM dimer (Figure 6.2 B) reveals that the bulk of the protein remains in the dimeric state while increasingly larger species begin to appear. These larger species most likely represent the formation of aggregate assemblies, as opposed to discrete structures, possibly driven by interactions along the re-designed i1 interface.

Although structures of these dimeric species have not been obtained, models of possible BMOE crosslinked dimers can be used to determine theoretical sedimentation coefficients of different dimer conformations using the program, Hydropro.[23] The first of these models, shown in Figure 6.3 A, assumes a linear shape with no contacts between the two protomers. The final model, Figure 6.3 D, shows a hand modeled “closed” form

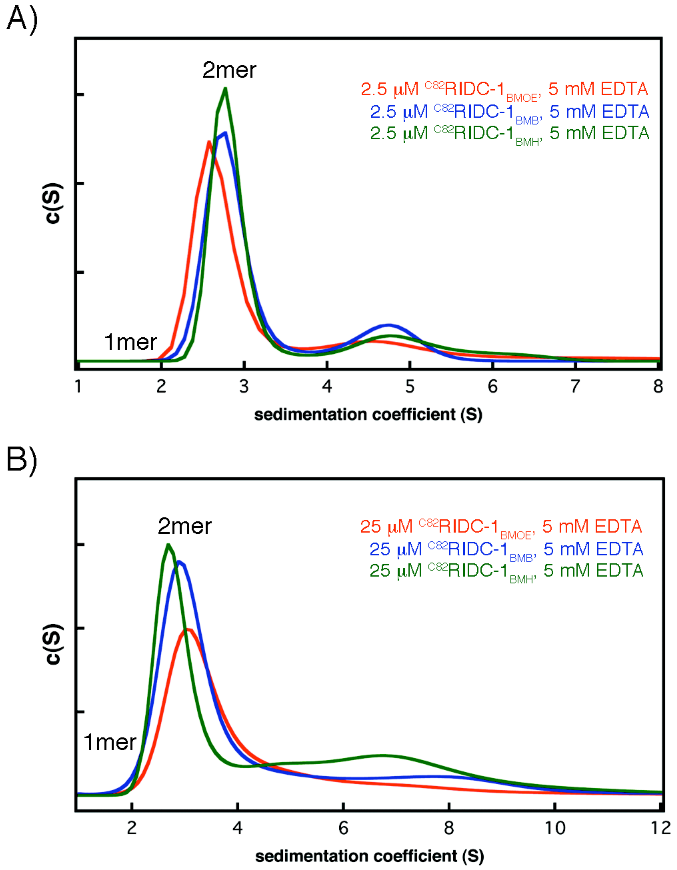


Figure 6.2. Metal independent crosslinked dimers. A) Sedimentation coefficient distribution of 2.5 µM ^{C82}RIDC-1$_{2,BMOE}$ (red), ^{C82}RIDC-1$_{2,BMB}$ (blue), and ^{C82}RIDC-1$_{2,BMH}$ (green) with 5 mM EDTA. B) Sedimentation coefficient distribution of 25 µM ^{C82}RIDC-1$_{2,BMOE}$ (red), ^{C82}RIDC-1$_{2,BMB}$ (blue), and ^{C82}RIDC-1$_{2,BMH}$ (green) with 5 mM EDTA. Monomer and dimer peaks are indicated, while larger, aggregate peaks are unlabeled due to the ambiguity of their oligomeric composition.

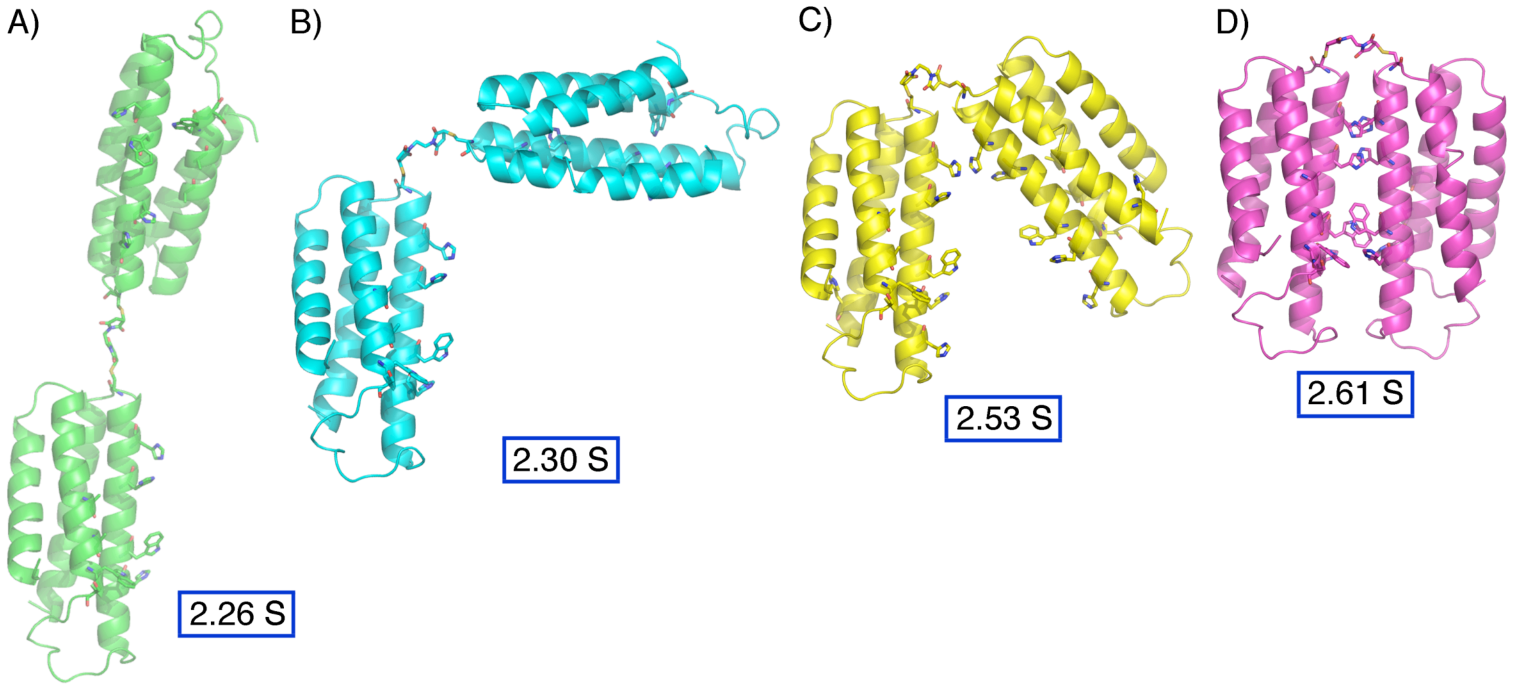


Figure 6.3. Possible ^{C82}RIDC-1$_{2,BMOE}$ dimers and their theoretical sedimentation coefficients. A) A linear dimer of ^{C82}RIDC-1 protomers crosslinked with a BMOE moiety has a theoretical sedimentation coefficient of 2.26 S, determined using hydropro and displayed in a blue box below the model. B-C) Increasingly more “closed” dimers correspond to increasing sedimentation coefficients of 2.30 and 2.53 S. D) A fully “closed” dimer, which lends itself to a number of stacking interactions between engineered His and Trp residues, has the largest theoretical sedimentation coefficient of 2.61 S, most closely resembling the peaks at about 2.8 S seen in Figure 6.2 A.

wherein the protomers form various stacking interactions between the numerous His and Trp residues. Figures 6.3 B and C show intermediates between these two states, thus yielding a series of different shaped structures whose calculated sedimentation coefficients range from about 2.2 S to 2.6 S going from the linear to “closed” dimer. As the experimentally determined sedimentation coefficients for the dimers of all three crosslinked constructs are between 2.6 and 2.8 S, it can be suggested that the dimers formed in solution more closely resemble the “closed” dimer in Figure 6.3 D, with its intra-protomer contacts, rather than the more “open” dimers described by the other models.

Crosslinker length does not affect Zn^{2+}-mediated tetramer formation or structure. While the pre-formation of a tetrameric assembly has obviously not been achieved, the chemical crosslinking of the ^{C82}RIDC-1 protein should nonetheless result in a stabilization of a Zn-mediated assembly identical to those previously described, with the only difference lying in the presence of the covalent crosslinker across i3. In order to assess this, we once again determined the solution state oligomeric assemblies of the three maleimide crosslinked constructs, this time in the presence of the template Zn^{2+}. The resulting sedimentation coefficient distributions demonstrate that, regardless of the length of the crosslinker we have chosen to employ, all of the constructs are capable of forming discrete, Zn-mediated tetrameric assemblies in solution (Figure 6.4 A).

Much as the re-design of i2 added enough of a thermodynamic push to allow almost exclusive Zn-mediated tetramer formation of RIDC-2 at 5 μM,[4] we see here that our covalent crosslinkers have been able to serve a similar purpose along i3, with almost full conversion to Zn-mediated tetramers even at a relatively low concentration of 2.5 μM

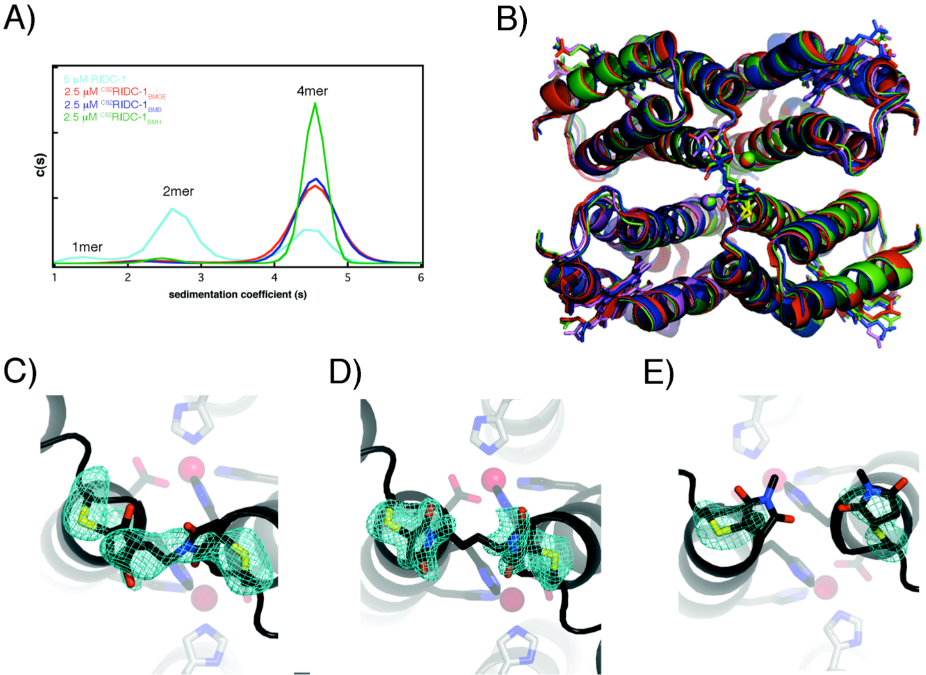


Figure 6.4. Zn^{2+}-mediated crosslinked tetramers. A) Sedimentation coefficient distributions of 5 µM RIDC-1 with 1:1 Zn^{2+} (cyan), compared to 2.5 µM ^{C82}RIDC-$1_{2,BMOE}$ (red), ^{C82}RIDC-$1_{2,BMB}$ (blue), and ^{C82}RIDC-$1_{2,BMH}$ (green) with 5 µM Zn^{2+}. B) Overlay of Zn_4:RIDC-1_4 (red), Zn_4:^{C82}RIDC-$_{2,BMOE}$ (blue), Zn_4:^{C82}RIDC-$1_{2,BMB}$ (green), and Zn_4:^{C82}RIDC-$_{2,BMH}$ (magenta) tetramers. C) BMOE linker in Zn_4:^{C82}RIDC-$_{2,BMOE}$ highlighted with the corresponding Fo-Fc omit electron density map (2.5 σ). D) BMB linker in Zn_4:^{C82}RIDC-$1_{2,BMB}$ highlighted with the corresponding Fo-Fc omit electron density map (3.2 σ) .E) BMH linker in Zn_4:^{C82}RIDC-$1_{2,BMB}$ highlighted with the corresponding Fo-Fc omit electron density map (2.8 σ).

dimer. As the building block of these assemblies are the RIDC-1 protein, this observation, as postulated previously, is likely derived from the decreased entropic factor that arises from the formation of a permanent dimer, as opposed to the elimination of unfavorable interactions provided by the i2 redesign.

Given that previous experiments have shown that, despite the fluidity of the i1 interface, the overall architectures of the Zn-mediated tetramers of MBPC-1 and RIDC-1 remain identical, we felt confident that the same would remain true for all three of our maleimide constructs. To prove this assertion we attempted to crystallize ^{C82}RIDC-$1_{2,BMOE}$, ^{C82}RIDC-$1_{2,BMB}$, and ^{C82}RIDC-$1_{2,BMH}$ in the presence of Zn^{2+}. After successfully obtaining crystals of all three metal-mediated assemblies, shown in Figure 6.4 B, we find that our hypothesis has borne itself true: all three of the crosslinked tetramers do indeed align well with the structure of Zn_4:RIDC-1_4, (rmsd over 424 αC's of 0.71, 0.50, and 0.44 Å from shortest to longest linker), even down to the tetrahedral Zn binding sites. In fact, the only differences between these structures–Zn_4:^{C82}RIDC-$1_{2,BMOE}$; Zn_4:^{C82}RIDC-$1_{2,BMB}$; and Zn_4:^{C82}RIDC-$1_{2,BMOE}$–is the crosslinker connecting the head-to-head aligned monomers of the V-shaped sub-structures.

As observed in Figure 6.4 C, electron density for the BMOE moiety is clear, with this compound being in a fully extended conformation as opposed to the highly distorted form seen in the metal free structure. The density for the corresponding BMB crosslinker, on the other hand, is far less well defined (Figure 6.4 D), while only the maleimide rings of BMH can actually be modeled (Figure 6.4 E). Considering the increasing flexibility of the compounds from the ethane to hexane linker arm, this observation is not surprising, and can go to further support the hypothesis that both ^{C82}RIDC-$1_{2,BMB}$ and ^{C82}RIDC-

$1_{2,BMH}$, at least, form "closed" dimers. The formation of these "closed" structures would then necessitate a classical 3D domain swap with a second dimer, a phenomenon commonly associated with duplicate domain proteins that oligomerize,[24, 25] to form the final Zn-mediated tetramer. These results are also consistent with the ability of metal coordination to serve as the main driving force behind all of the coordination induced PPIs we have thus far described.

Evidence for increased Zn^{2+} affinity of ^{C82}RIDC-$1_{2,BMOE}$. As shown in Figure 6.6 A, ^{C82}RIDC-$1_{2,BMOE}$ is still capable of binding other divalent metal ions such as Ni^{2+}, Cu^{2+}, and Co^{2+}, resulting in complete formation of a metal-mediated tetrameric species. Assuming we have accomplished our goal of imparting metal selectivity on to this construct, we should see that ^{C82}RIDC-$1_{2,BMOE}$ affinities for these various metals should be lower than for the template Zn^{2+} ion. To probe this idea, metal binding competition assays were performed with the metal chelating fluorophore, Mag-Fura-2. Control experiments were performed to determine the affinity of Mag-Fura-2 for Zn^{2+}, Ni^{2+}, and Co^{2+} (Figure 6.5), with the resulting Mag-Fura-2 K_d values, listed in Table 6.5, being similar to previously reported values.[26, 27]

Titrations of Cu^{2+} in to a sample containing ^{C82}RIDC-$1_{2,BMOE}$ and Mag-Fura-2 are complicated by a large number of relatively weak binding events that cause the binding isotherm to continue on in excess of 12 equivalents of the metal, making these titrations inconclusive. Similar titrations with a stronger chelator, Fura-2 revealed that the crosslinked construct was not able to compete with the fluorophore, hence we were unable to perform competition assays with Cu. Nonetheless, ^{C82}RIDC-$1_{2,BMOE}$ – Mag-Fura-2 competition experiments with Zn, Ni, and Co, shown in Figure 6.6 B-D,

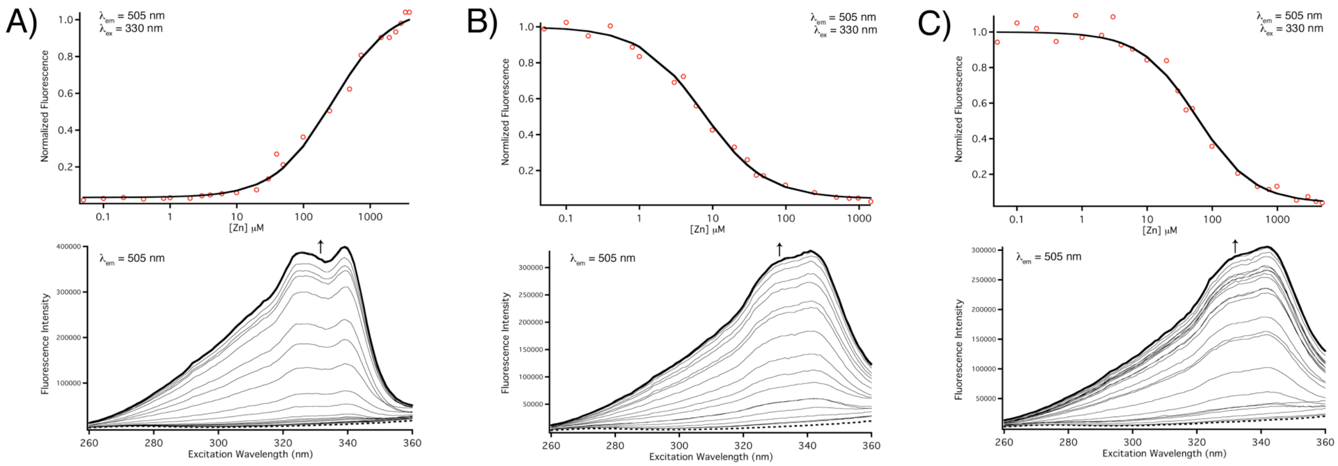


Figure 6.5. Mag-Fura-2 control metal titrations. A, top) Titration of $ZnCl_2$ in to 0.5 μM Mag-Fura-2 incubated with 1 mM $NiSO_4$. A, bottom) Corresponding excitation scans of the Zn^{2+} titration. B, top) Titration of $NiSO_4$ in to 0.5 μM Mag-Fura-2 incubated with 1 mM $CaCl_2$. B, bottom) Corresponding excitation scans of the Ni^{2+} titration. C, top) Titration of $CoCl_2$ in to 0.5 μM Mag-Fura-2 incubated with 1 mM $CaCl_2$. A, bottom) Corresponding excitation scans of the Co^{2+}. Divalent metal binding affinities of Mag-Fura-2. Titrations were fit using data collected at an excitation wavelength of 330 nm and emission wavelelngth of 505 nm employing a 1:1 Metal:Mag-Fura-2 model in Dynafit. All titrations were performed at 25 °C in 20 mM MOPS (pH7) and 150 mM NaCl. Corresponding K_d values are listed in Table 6.5.

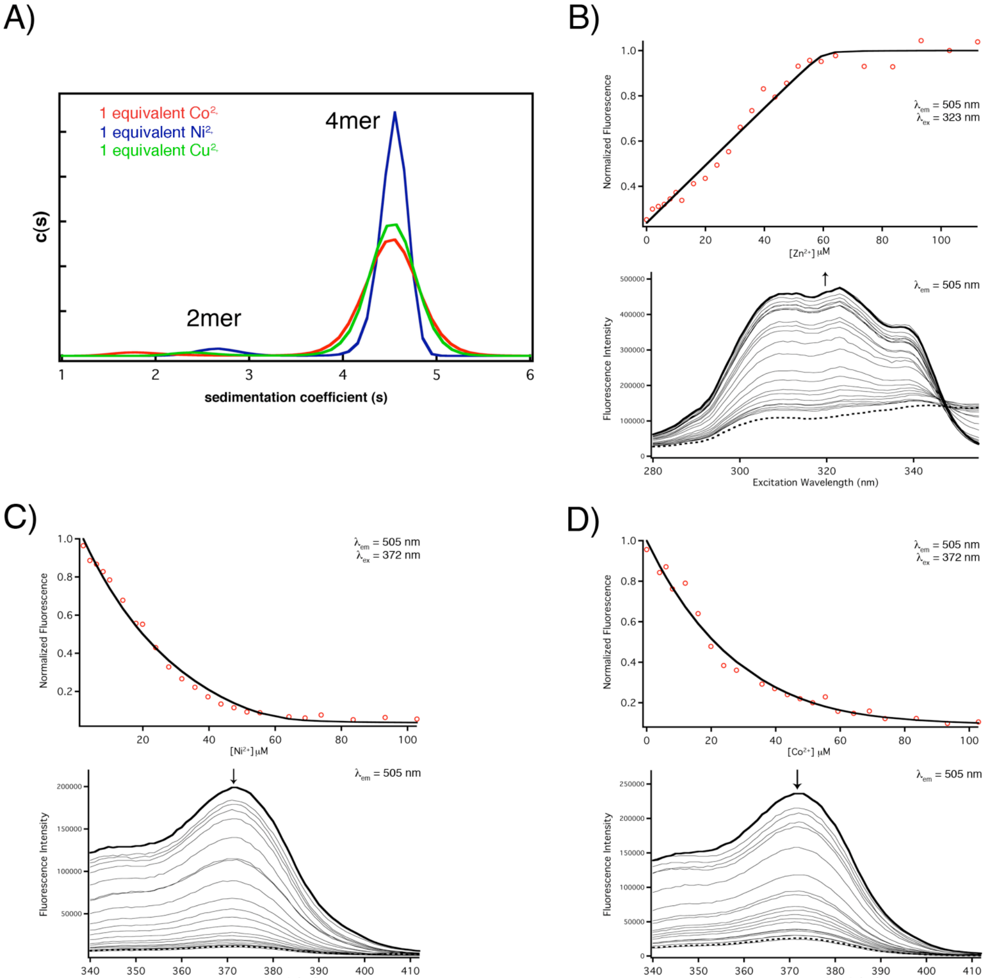


Figure 6.6. Divalent metal binding affinities of ^{C82}RIDC-1$_{2,BMOE}$. A) Sedimentation coefficient distributions of ^{C82}RIDC-1$_{2,BMOE}$ with 1:1 Ni^{2+} (blue), Co^{2+} (red), and Cu^{2+} (green). B, top) Titration of ZnCl$_2$ in to 10 μM Mag-Fura-2 incubated with 25 μM ^{C82}RIDC-1$_{2,BMOE}$. B, bottom) Corresponding excitation scans of the Zn^{2+} titration. C, top) Titration of NiSO$_4$ in to 10 μM Mag-Fura-2 incubated with 25 μM ^{C82}RIDC-1$_{2,BMOE}$. C, bottom) Corresponding excitation scans of the Ni^{2+} titration. D, top) Titration of CoCl$_2$ in to 10 μM Mag-Fura-2 incubated with 25 μM ^{C82}RIDC-1$_{2,BMOE}$. D, bottom) Corresponding excitation scans of the Co^{2+}. Titrations were fit using data collected at an excitation wavelength of 323 nm for Zn titrations and 372 nm for both Ni and Co titrations, with emission at 505 nm. All samples were fit using a single binding event model (4 x 1) in Dynafit. All titrations were performed at 25 °C in 20 mM MOPS (pH7) and 150 mM NaCl. Corresponding K$_d$ values, derived from triplicate experiments and fits, are listed in Table 6.5.

Table 6.5. Dissociation constants Mag-Fura-2 and ^{C82}RIDC-1$_{2,BMOE}$ titrations. Values shown are the average of three replicate titrations with corresponding errors shown in parenthesis.

Metal	**K_d (μM)**	
	Mag-Fura-2	**^{C82}RIDC-1$_{2,BMOE}$**
Zn	0.047 (0.005)	0.041 (0.010)
Ni	0.175 (0.014)	0.54 (0.055)
Co	1.41 (0.22)	9.29 (1.02)

demonstrate that a total of four binding events occur, consistent with the formation of a tetramer similar to the Zn-mediated assemblies described above.

A single binding event (4 × 1) model is found to fit best to these data. This model assumes that each protomer in the final tetramer has one metal binding site, and each of these sites has an equivalent K_d, effectively making metal binding independent of oligomerization, to a first approximation. The resulting Kd's of Zn, Ni, and Co binding were determined to be 41 nM, 540 nM, and 9.29 μM, respectively. Despite the fact that these values may not be exact, due to the absence of protein oligomerization in the model, they do nonetheless demonstrate a relative increase in affinity for the template Zn^{2+} ion over Ni^{2+} and Co^{2+}.

To determine if Zn could out-compete Cu for binding to ^{C82}RIDC-$1_{2,BMOE}$, we set out to crystallize the protein in the presence of both of the metal ions. Here, which ever metal preferentially binds to the complex can be elucidated by collecting x-ray diffraction data at or near the K edges of the respective metals. For this, samples of ^{C82}RIDC-1 were crosslinked with 1:1 Zn, purified, and dialyzed into 20 mM TRIS (pH 7), never being treated with EDTA. A crystallization drop was then set up containing the Zn incubated protein, as well as $CuSO_4$ from the mother liquor. Data from the resulting structure was then collected at the K edge energies of the respective metals (Figure 6.7 A), yielding the Cu and Zn anomalous difference maps shown in Figure 6.7 B and C, respectively. There is clear anomalous signal indicating Cu ions at seven different positions in the crystal structure. Interestingly, none of these atoms are located at the Zn binding sites, but are rather found exclusively at the periphery of the tetramer. Zn ions, however, can be seen,

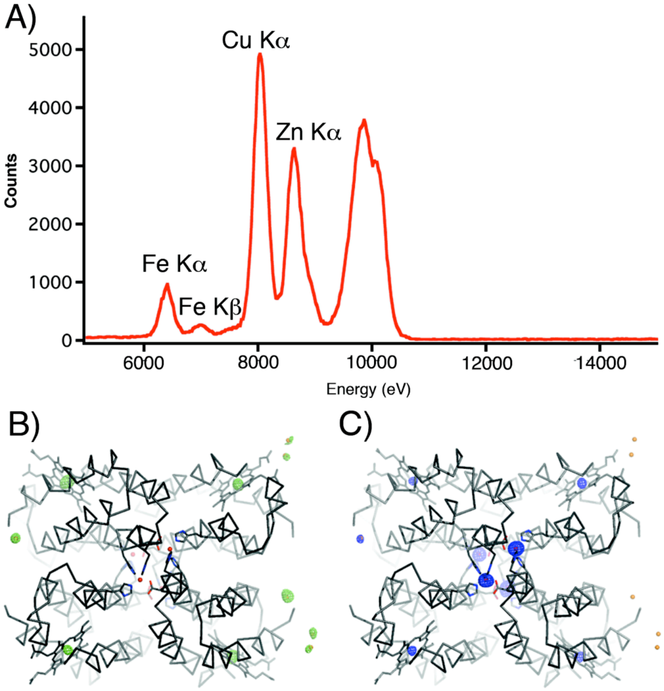


Figure 6.7. Evidence for ^{C82}RIDC-1$_{2,BMOE}$ preference for binding Zn^{2+} over Cu^{2+}. A) X-ray fluorescence excitation scan, performed at the Zn^{2+} K edge, demonstrates the presence of both Cu^{2+} and Zn^{2+} in a crystal of ^{C82}RIDC-1$_{2,BMOE}$. Anomalous difference maps calculated at the A) Cu K edge (9000 eV, green) and B) near the Zn K edge (9800 eV, blue) for ^{C82}RIDC-1$_{2,BMOE}$ crystallized in the presence of both Zn^{2+} and Cu^{2+} contoured to 5 s. An anomalous peak can be seen for the four heme Fe atoms at both energies. Cu can be seen at multiple different sites at crystal contacts in B), while the core metal sites of the tetramer only demonstrate the presence of Zn^{2+} in C).

from the Zn K edge anomalous data, at the expected His^3Asp binding site described in our previous oligomers. This, in conjunction with the Mag-Fura-2 titrations, leads to the suggestion that the BMOE crosslinking of RIDC-1 has, indeed, allowed for the formation of a metal selective protein assembly.

Evidence for binding of tetrahedral anions by ^{C82}*RIDC-1*$_{2,BMOE}$. While attempting to crystallize the ^{C82}RIDC-1$_{2,BMOE}$ dimer, a crystallization condition was employed consisting of only 2 M ammonium sulfate and EDTA. The resulting crystal, which diffracted to 2.8 Å and contained about 70% solvent, was found to contain not a crosslinked dimer, but rather a dimer of dimers, or a tetramer (Figure 6.8). Initial analysis of this structure allowed us to imagine this tetramer as the higher ordered species described in SV experiments in the absence of added metal. However, closer inspection revealed the presence of a sulfate anion at the heart of the tetramer, bound through hydrogen bonds formed between the sulfate and His77 from all four protomers (Figure 6.8 D and E, Table 6.4), which results in an overall architecture (SO_4:^{C82}RIDC-1$_{2,BMOE}$) dissimilar from the metal-mediated oligomers previously described.

Akin to the Zn_4:MBPC-1_4 structure, few hydrogen bonding or electrostatic interactions are found to mediate the tetramer beyond those mediating the protein-sulfate interaction. There are, however, repeating His63-Glu86' and Trp66-Glu81' hydrogen bonding pairs formed between the two sets dimers at the vertices of the crosslinked V's (Figure 6.8 B, where the apostrophe signifies a residue on an opposing crosslinked dimer; Table 6.4). Sealing the sides of the tetramer is a largely hydrophobic "zipper" formed by an anti-parallel arrangement of residues Thr31/Ala34/Ala35/Ala38/Trp41 between opposing crosslinked pairs (Figure 6.8 C).

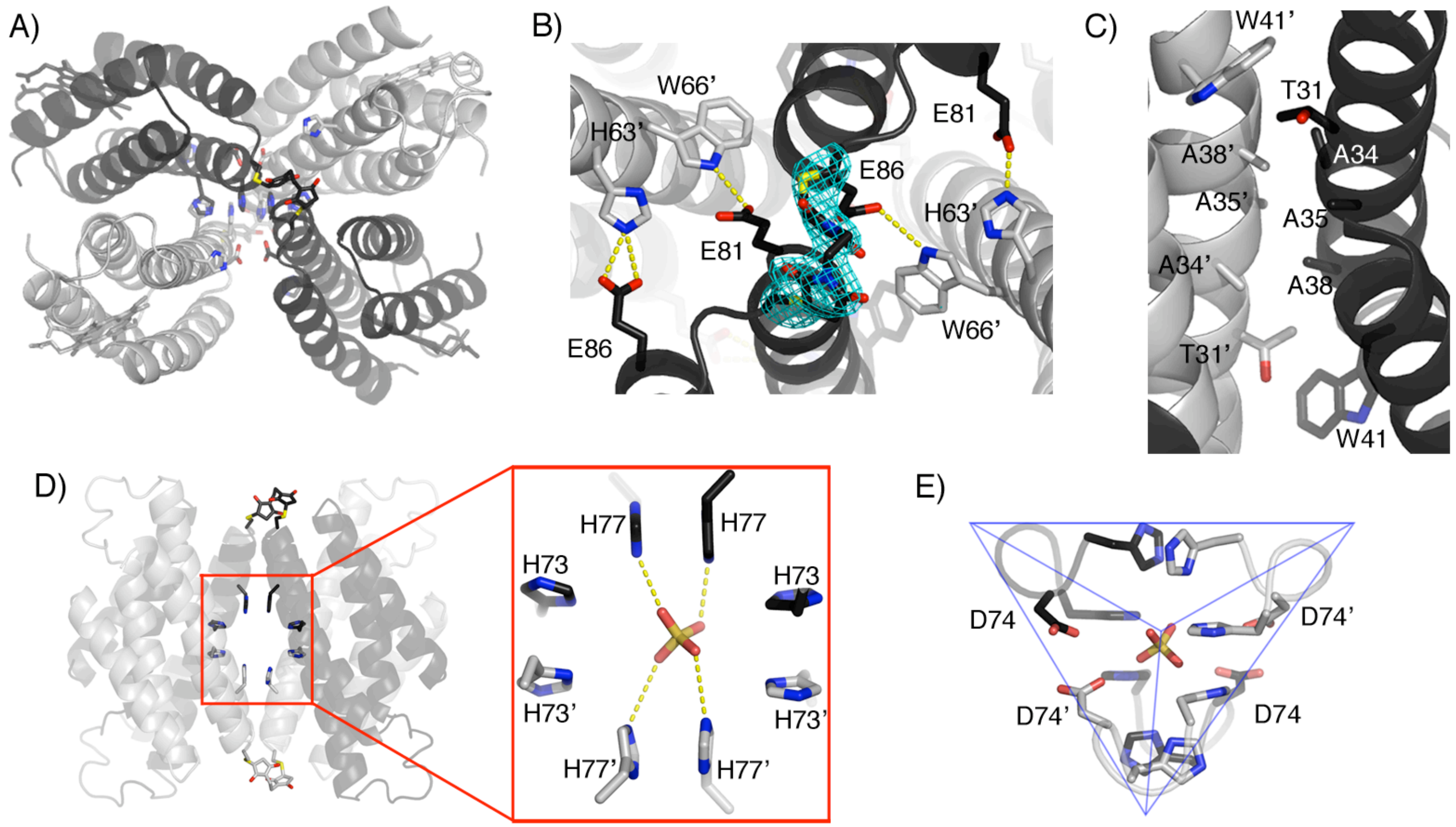


Figure 6.8. Crystal structure of SO_4:^{C82}RIDC-$1_{2,BMOE}$. A) Top view of the ^{C82}RIDC-$1_{2,BMOE}$ crystal structure wherein the BMOE moiety and the putative Zn binding residues are shown as sticks. B) Close-up view of the two (nearly) symmetrical interaction zones in the tetramer interface detailing the h-bonding contacts. The BMOE linker is also highlighted with the corresponding Fo-Fc omit electron density map (5 σ). C) Close-up of the i1 hydrophobic "zipper" that seals the sides of the assembly. D) His stacking "pinwheel" interactions within the core of the complex. A zoomed in view (red box) shows these interactions, and corresponding h-bonding interactions between the His77 residues and the sulfate anion, in better detail. E) The tetrahedral His73-Asp74-His77 "cage" which surrounds the sulfate anion.

The striking feature of this structure is a series of His73 and His77 stacking interactions that form a “pinwheel” around the central sulfate ion (Figure 6.8 D). These interactions mediate both intra- and inter-crosslinked dimer contacts, the former of which are formed by the stacking of His77 imidazole rings from crosslinked pairs, while the latter are formed by His73-His73’ stacking between opposite dimers. This “pinwheel” is then closed in to a tetrahedral “cage” by the neighboring Asp74 residues, resulting in an encapsulation of the sulfate ion. (Figure 6.7 E).

Simplification of the Zn_4:RIDC-1_4 and SO_4:^{C82}RIDC-$1_{2,BMOE}$ tetramers to their respective helix3’s, followed by alignment of one of the protomeric subunits quickly makes it obvious that these two assemblies do not overlay (Figure 6.9 A). From this simplified view, one finds that there is a 14.6 Å contraction of the distance between the vertices of the V’s between Zn_4:RIDC-1_4 and SO_4:^{C82}RIDC-$1_{2,BMOE}$ (Figure 6.9 B). This contraction is concurrent with a 7^o “scissor”-like closure of the SO_4:^{C82}RIDC-$1_{2,BMOE}$ V, causing a severe folding in of the maleimide moiety which forces the two rings of the BMOE to stack with one another (Figure 6.9 C). All of this outlines a thermodynamically favored arrangement made possible by the fluidity in i1. Despite this fact, a rough overall organization described by the interlacing of two V-shaped sub-structures is maintained, suggesting the possibility that ^{C82}RIDC-$1_{2,BMOE}$ not only has a preference for binding tetrahedral Zn^{2+}, but tetrahedral small molecules as well.

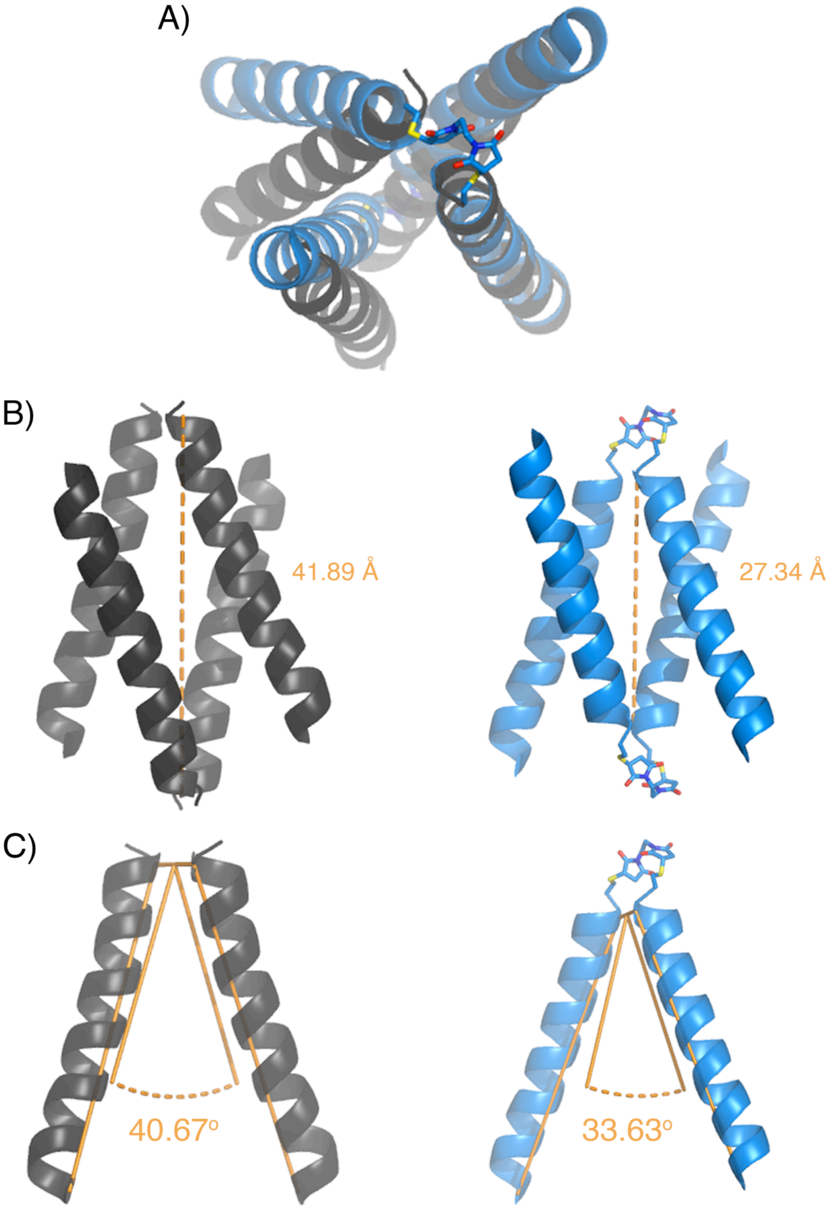


Figure 6.9. Comparison of SO_4:^{C82}RIDC-$1_{2,BMOE}$ and Zn_4:RIDC-1_4 tetramers. A) Top view of an overlay of helix3 V's of Zn_4:RIDC-1_4 (black) and SO_4:^{C82}RIDC-$1_{2,BMOE}$ (blue). B) Side view of the tetramers demonstrating the 14 Å shift in distances between V vertices from Zn_4:RIDC-1_4 to SO_4:^{C82}RIDC-$1_{2,BMOE}$. C) Side view of a single V of both assemblies showing the 7° "scissoring" closure between the two structures. Distances were measured from the Cys82 Cα of one vertex to the Cys82 Cα of the opposing V. The V dihedral was measured as follows: His59 Cα – Cys82 Cα – Cys82' Cα –His59' Cα, where ' denotes the opposite V pair.

Conclusions

By crosslinking two of our MeTIR derived RIDC-1 variants, we have successfully created a duplicate-domain construct capable of increased formation of the previously described Zn-mediated tetramer. The mechanism of this oligomerization appears to involve a “closed” dimer that undergoes a 3D domain swap with a second dimer on binding the template metal ion, an action that appears to be independent of the length of the linker arm used. The implication of this 3D domain swapping lies in the hypothesis that this phenomenon may impart allostery between active sites in naturally occurring protein assemblies, suggesting that allostery could be engineered in to functionalized versions of these Zn centers in the future, thereby expanding the possibilities for future protein systems based on RIDC-1.

While we originally assumed Zn binding to MBPC-1, and in turn RIDC-1, was cooperative, competition assays with Mag-Fura-2 suggest that this is not true for the BMOE-crosslinked dimer as there is no increased affinity for the metal ions as each consecutive one is bound. Even though the binding model we have chosen does not include a factor for the association of the dimeric subunits, it does allow us to calculate relative affinities of ^{C82}RIDC-1$_{2,BMOE}$ for divalent Zn, Ni, and Co, which suggest a greater affinity for Zn over the other two metals. This, in conjunction with the crystal structure containing both Zn and Cu, lead us to propose that crosslinking RIDC-1 may have imparted Zn binding selectivity on to the protein.

A crystal structure of this same construct forming a distinctly new tetramer via binding sulfate has led us to propose that the selectivity describe above extends to tetrahedral molecules in general, and not only Zn. To test this idea, future SV experiments containing ^{C82}RIDC-$1_{2,BMOE}$ and various ratios of sulfate, molybdate, phosphate, and carbonate will be performed. All four of these molecules have a minus two charge assuming that, at the pH of these experiments, pH 7, the phosphate exists as hydrogen phosphate; while all the carbonate have a tetrahedral geometry, with carbonate being trigonal planar, thus serving as a control for tetrahedral binding preference.

As the average bond length of the molybdenum-oxygen bond in molybdate is longer than the sulfur-oxygen bond in sulfate, stronger H-bonding between the protein and the anion in the tetrahedral "cage" of the assembly would be expected. Direct coordination to the molybdenum by both Asp and His residues from the "cage" is also possible, leading to an even stronger binding of molybdate than sulfate to ^{C82}RIDC-$1_{2,BMOE}$. While the bond lengths in hydrogen phosphate are similar to sulfate, the presence of the hydrogen on the molecule could allow for H-bonding to the Asp74 residues of the "cage", again possibly making this interaction stronger than sulfate.

Though initial experiments have suggested that the ideas outlined above are true, further studies are necessary before drawing final conclusions. Performing these solution studies, as well as further binding and crystallization experiments, can shed light on the sulfate-mediated structure we have elucidated, while at the same time adding yet another "evolution" to our cyt *cb*$_{562}$ system; namely the ability to strongly and selectively bind biologically relevant small molecules.

References

1. Wilson, A. C., Carlson, S. S., White, T. J. **1977**. Biochemical evolution. *Ann. Rev. Biochem.* . 46, 573-639.

2. Golding, G. B., Dean, A. M. **1998**. The structural basis of molecular adaptation. *Mol. Biol. Evol.* 15, 355-369.

3. Pal, C., Papp, B., Lercher, M. J. **2006**. An integrated view of protein evolution. *Nat. Rev. Genet.* 7, 337-348.

4. Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B., Tezcan, F. A. **2010**. Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. USA.* 107, 1827-1832.

5. Zhang, J. **2003**. Evolution by gene duplication: an update. *Trends in Ecology & Evolution*. 18, 292-298.

6. Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., Teichmann, S. A. **2004**. Structure, function and evolution of multidomain proteins. *Current Opinion in Cell Biology*. 14, 208-216.

7. Lynch, M., Conery, J. S. **2000**. The evolutionary fate and consequences of duplicate genes. *Science*. 290, 1151-1155.

8. Yamashita, I., Nakamura, M., Fukui, S. **1987**. Gene fusion is a possible mechanism underlying the evolution of *STA1*. *J. Bacteriol.* 169, 2142-2149.

9. Cameron, A. D., Olin, B., Ridderstrom, M., Mannervik, B., Jones, T. A. **1997**. Crystal structure of human glyoxalase I—Evidence for gene duplication and 3D domain swapping. *EMBO J.* 16, 3386-3395.

10. Liua, L., Iwata, K., Yohdab, M., Mikia, K. **2002**. Structural insight into gene duplication, gene fusion and domain swapping in the evolution of PLP-independent amino acid racemases. *FEBS Letters*. 528, 114-118.

11. Teichmann, S. A., Park, J., Chothia, C. **1998**. Structural assignments to the Mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements. *Proc. Natl. Acad. Sci. USA*. 95, 14658-14663.

12. Wang, Y., Gu, X. **2001**. Functional divergence in the caspase gene family and altered functional constraints: statistical analysis and prediction. *Genetics*. 158, 1311-1320.

13. Durchschlag, H., Zipper, P. **1994**. Calculation of the partial volume of organic compounds and polymers. *Progr. Colloid Polym. Sci.* 94, 20-39.

14. Schuck, P. **2000**. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* 78, 1606-1619.

15. Krissinel, E., Henrick, K. **2007**. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 372, 774-797.

16. Leslie, A. G. W. **1992**. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*. 26.

17. Murshudov, G. N., Vagin, A. A., Dodson, E. J. **1994**. The CCP4 Suite: Programs for protein crystallography. *Acta Cryst. D*. 50, 760-763.

18. Vagin, A., Teplyakov, A. **1997**. MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.

19. Murshudov, G. N., Vagin, A. A., Dodson, E. J. **1997**. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.D.* 53, 240-255.

20. DeLano, W. L. *The PYMOL molecular graphics system (http://www.pymol.org)*, 2003.

21. Lattanzio, F. A., Jr. and Bartschat, D. K. **1991**. The effect of pH on rate constants, ion selectivity and thermodynamic properties of fluorescent calcium and magnesium indicators. *Biochem. Biophys. Res. Comm.* 177, 184-191.

22. Kuzmic, P. **1996**. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* 237, 260-273.

23. Torre, J. G. d. l., Huertas, M. L., B. Carrasco. **2000**. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78, 719-730.

24. Bennett, M. J., Schlunegger, M. P., Eisenberg, D. **1995**. 3D Domain swapping: A mechanism for oligomer assembly. *Protein Sci.* 4, 2455-2468.

25. Liu, Y., Eisenberg, D. **2002**. 3D domain swapping: As domains continue to swap. *Protein Sci.* 11, 1285-1299.

26. Simons, T. J. **1993**. Measurement of free Zn^{2+} ion concentration with the fluorescent probe mag-fura-2 (furaptra). *J. Biochem. Biophys. Methods*. 27, 25-37

27. Golynskiy, M. V., Gunderson, W. A., Hendrich, M. P., Cohen, S. M. **2006**. Metal binding studies and EPR spectroscopy of the manganese transport regulator MntR. *Biochemistry*. 45, 15359-15372.

Chapter 7

Dissertation Conclusions

In traditional inorganic chemistry circles, proteins are considered to simply be the vessel surrounding the invaluable metal ions. In the bioinorganic community, proteins are treated with a little more awe, driving the field of biomimetic model chemistry, which aims to elucidate the inner workings of metal centers without the intricacies of the surrounding protein framework. No matter what the perspective, there is no denying that proteins are sophisticated ligands that guide metals to impressive chemical feats.

When we began this work, we took the notion of "proteins as ligands" literally, asking, can we use individual, folded proteins as ligands to control their self-assembly through metal coordination? Could we then employ this same methodology to build complex multi-protein assemblies like nature does every time it needs to do something more involved than what is achievable with simpler, single protein systems (e.g., photosynthesis, respiration, nitrogen fixation)? Making this concept less far-fetched is the fact that it has been shown that nature has found a way to employ metal ions in exactly this way, as exemplified by the enhancement of Hepatitis B Virus (HBV) capsid formation through the coordination of divalent Ni and, especially, Zn.[1] Thus, the installation of bis-His clamps on to cyt cb_{562}, resulting in the MBPC-1[2] construct capable of forming metal-mediated oligomers, can be seen as a parallel to the natural evolution that has given rise to this viral particle.

The hydrophobic interactions that drive HBV capsid assembly have been estimated to be fairly weak, with an apparent K_d in the low micro-molar range,[3] allowing for the specific formation of the final particle without becoming trapped in intermediate, kinetically stable forms. Metal-Directed Protein Self-Assembly (MDPSA) of the MBPC-1 and 2 mutants are similarly under thermodynamic control, as evidenced by the formation of discreet oligomeric species in solution. The formation of these assemblies, as well as their final overall architectures, are not only dependent on the driving force of the metal coordination and the metal specific stereochemical preferences,[4] but also on the presence of non-optimized, weak "secondary" interactions lining the PPI interfaces.[5]

Fine-tuning of these "secondary" interactions, as implemented in MeTIR re-design of the Zn_4:MBPC-1_4 tetramer, leads to accelerated tetramer formation of RIDC-1.[6] This acceleration is seen to be due to two different factors: 1) a decreased entropic factor due to by the ability of RIDC-1 to pre-assemble in to a dimeric form; and 2) a stabilization of the final Zn-mediated oligomer from increased favorable interactions along the re-designed interface. These interactions then begin to enforce the tetrahedral geometry of the template Zn^{2+} ion not possible with MBPC-1, paralleling the differentiation of proteins with similar folds witnessed within naturally occurring protein families.[7, 8] Crosslinking of the RIDC-1 protein with moieties of varying linker-arm length allowed us to expand the breadth of our synthetic cyt cb_{562} protein family to include two-domain proteins that not only demonstrate a preference for binding Zn^{2+} over other divalent metal ions, but also a possible ability to bind tetrahedral anions. Should this final point bear true, one could imagine creating a large pool of protein-based assemblies with a virtually limitless diversity of function derived from a single parent molecule, much as evolution has been able to develop.

The work summarized above has not only shows us the feasibility of our original goals, but also opened the way for new research prospects, such as the possible roles of metal ions in the evolution of protein folds/complexes and de novo design of protein interfaces. Having increased the range of properties of cyt cb_{562} from monomeric electron transport protein to Zn selective 16-helix assembly, we now look to what we can do to exploit the functionality, and functionalizability, of protein surfaces to rule over the reactivities of these sites. One thing that we have certainly learned along the way is that proteins indeed make complicated ligands, which, more often than not, interact with metals and assemble in ways that are hard to predict. This is the major hurdle that still remains before we can construct a fully functional metalloprotein assembly. However, by expanding our work to include other metal coordination motifs, protein secondary structures, and protein systems with innate internal symmetries, we can further develop rules for how to predict metal-mediated PPIs. With the additional perfect hindsight afforded by structural studies, some inorganic intuition, and a combination of advances in computational protein design and analytical techniques, we are certain that these hurdles will be overcome in time.

Chapter 7 is reproduced in part with permission from: Salgado, E. N., Radford, R. J., Tezcan, F. A. **2010**. Metal-directed protein self-assembly. *Acc. Chem. Res.* 43, 661-672. Copyright 2010 American Chemical Society.

References

1. Stray, S. J., Ceres, P., Zlotnick, A. **2004**. Zinc ions trigger conformational change and oligomerization of Hepatitis B virus capsid protein. *Biochemistry*. 43, 9989-9998.

2. Salgado, E. N., Faraone-Mennella, J., Tezcan, F. A. **2007**. Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* 129, 13374-13375.

3. Zlotnick, A. **2003**. Are weak protein-protein interactions the general rule in capsid assembly? *Virology*. 315, 269-274.

4. Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., Tezcan, F. A. **2009**. Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical assemblies through $\mathrm{Cu}^{\mathrm{II}}$ and $\mathrm{Ni}^{\mathrm{II}}$ coordination. *Inorganic Chemistry*. 48, 2726-2728.

5. Salgado, E. N., Lewis, R. A., Faraone-Mennella, J., Tezcan, F. A. **2008**. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* 130, 6082–6084.

6. Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B., Tezcan, F. A. **2010**. Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. USA*. 107, 1827-1832.

7. Chandra, N. R., Prabu, M. M., Suguna, K., M., V. **2001**. Structural similarity and functional diversity in proteins containing the legume lectin fold. *Protein Eng.* 14, 857-866.

8. Stühmer, W., Ruppersberg, J. P., Schröter, K. H., Sakmann, B., Stocker, M., Giese, K. P., Perschke, A., Baumann, A., Pongs, O. **1989**. Molecular basis of functional diversity of voltage-gated potassium channels in mammalian brain. *EMBO J.* 8, 3235–3244.