UNIVERSITY OF CALIFORNIA SAN DIEGO

Bottom-up design of metalloprotein assemblies with diverse metal coordination motifs and emergent properties

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Chemistry

by

Albert Aris Kakkis

Committee in charge:

Professor F. Akif Tezcan, Chair
Professor Seth Cohen
Professor Neal Devaraj
Professor Ulrich Müller
Professor Dong Wang

2023

The Dissertation of Albert Aris Kakkis is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my friends and family, for their ceaseless devotion to a life of altruism and compassion which has made this thesis possible.

EPIGRAPH

"In chemistry, our theories are crutches; to show that they are valid, they must be used to walk"

Jean-Baptiste-André Dumas

"Life would be indeed easier if the experimentalists would only pause for a little while!"

Rudolph A. Marcus

"Most innovations are not obvious to other people at the time. You have to believe in yourself. If you've got a good idea, follow it even when others say it's not."

Frances Arnold

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Ala/A	alanine
Arg/R	arginine
Asn/N	asparagine
Asp/D	aspartate
Cys/C	cysteine
Glu/E	glutamate
Gln/Q	glutamine
Gly/G	glycine
His/H	histidine
Ile/I	isoleucine
Leu/L	leucine
Lys/K	lysine
Met/M	methionine
Phe/F	phenylalanine
Pro/P	proline
Ser/S	serine

Thr/T	threonine
Trp/W	tryptophan
Tyr/Y	tyrosine
Val/V	valine
ASU	asymmetric unit
bipy	2,2'-bipyridine
BMC	bacterial microcompartment
CD	circular dichroism
CDO	cysteine dioxygenase
CIT	chemically-inducible trimer
DFT	density functional theory
DT	dithionite
DTT	dithiothreitol
EPR	electron paramagnetic resonance
equivalents	equiv.
ESI-MS	electrospray ionization-mass spectrometry
Fe-S	iron-sulfur

FFT	fast fourier transform
FRET	fluorescence resonant energy transfer
FWHM	full width at half maximum
Gdn HCl	guanidine hydrochloride
GEAL	$^{G77/E78/A80/L81}$ncTriCyt3
GESL	$^{G77/E78/S80/L81}$ncTriCyt3
GSH	reduced glutathione
hCAII	human carbonic anhydrase II
His_6	hexahistidine
IPTG	isopropyl β-D-1-thiogalactopyranoside
IW	Irving-Williams
LB	Luria-Bertani
LMCT	ligand-to-metal charge transfer
m/z	mass/charge
MBPC1	metal-binding cytochrome-1
MDPSA	metal-directed protein self-assembly
MeTIR	metal-templated interface redesign

MLCT	metal-to-ligand charge transfer
MWCO	molecular weight cutoff
nc	no chrome
NHase	nitrile hydratase
NMR	nuclear magnetic resonance
ODA	o-dianisidine
PDB	protein data bank
PEG	poly(ethylene glycol)
p-NPA	p-nitrophenyl acetate
PPI	protein-protein interaction
ppm	parts per million
PRE	paramagnetic relaxation enhancement
QM/MM	quantum mechanics/molecular mechanics
REU	rosetta energy units
RIDC1	rosetta interface design 1
SSM	site saturation mutagenesis
SV-AUC	sedimentation velocity-analytical ultracentrifugation

Trispyrazolyl borate	Tp
UV-vis	UV-visible
vdW	van der Waals
z/r	charge/radius
βME	beta-mercaptoethanol

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My intellectual growth as a scientist and my spiritual growth as a man is not possible in a vacuum. After twenty-nine years of life I have come to believe resolutely that no human is purely self-made; we all need helping hands and guiding lights to thrive. This dissertation may have been written through my hands, but it is ultimately the product of many talented, driven, and compassionate individuals. The people acknowledged herein represent just a small fraction of this group.

My love for Chemistry was seeded at Pomona College, under the tutelage of Professors Roberto A. Garza-López and Cynthia Selassie. I appreciate their support of my intellectual wanderings, which ended largely in failure but occasionally led me to joyous moments of success. My appreciation of research as iterative, frustrating, yet worthy of late Friday nights was made possible by my experiences in their labs.

When I left Pomona College, I thought I had a good idea of what it takes to test a hypothesis, to test if ideas sowed in the Earth can bear fruit via experimentation. Stepping into Akif's lab, it turned out I had no idea. Akif showed me that to do great science, one must collectively harness immense imagination, painstaking preparation, and extensive effort. I am grateful to Akif for providing me and my colleagues the resources and encouragement to put crazy ideas to the test, to view Nature as a nanoscopic jungle gym and ourselves as fearless acrobats leaping from branch to branch. I will do my darnedest to maintain that fearless spirit for the rest of my scientific career.

It takes more than a PI to make a lab. For Akif's flights to the unexpected, he needs co-pilots who are eager to share the cockpit. I am grateful to past and present members of the Tezcan Lab for their eye for detail and willingness to share their knowledge. Our lab strikes the perfect balance between intensity and gentility. You will be told when your hypothesis is flawed but

offered an attentive ear as you attempt to grasp something better. You will find a community of devotees with seemingly unbreakable focus but an openness to longer-than-usual Friday morning coffee breaks. I gratefully acknowledge my earliest mentor and late-night-hypothesis sounding board, Professor Eyal Golub. I am grateful for the myriad moments when Eyal tossed me a life preserver as I swirled in a restless sea of scientific unknowns. To this day, I marvel at the fierce compassion he provided to me, and hope that I can repay at least a fraction of it. Also, I must recognize Dr. Julian Esselborn, my late-night lab partner in the early days of my PhD journey. I cherish the memories of our political conversations, particularly our passionate diatribes about unscrupulous governments. Julian has an altruism that is matched by few. In Akif's lab, he devoted an inordinate amount of time and sweat to improve others' hypotheses, data interpretation, and understanding of fundamental scientific principles. He is truly one of a kind.

To my family, especially my mother, father, brother, and sister. My memories of childhood are sepia-toned; they abound with images of basketballs in bushes, tortellini soup and zipolli on kitchen counters, and Christmas music playing through speakers. Every day I reflect with gratitude on that blessed young life and am motivated to provide that same life for children of my own. My love for my family is fierce and unyielding because it so greatly deserves it. Let us make more beautiful memories.

To my Nonna, Nonno, and Yiayia. You set out on the toughest experiment there is: leaving a homeland you love for a new place with unfamiliar faces and an unfamiliar language. Without that brave journey, I wouldn't be sitting here typing these words. That great experiment bore fruit in the lives of three generations, marked by graduations, weddings, and birthday parties. How lucky I was to be raised by my grandparents: an American boy living on a steady diet of melanzana

parmigiana, baklava, and an appreciation for the stories of struggle and triumph that define his ancestors. Vi voglio bene. Σ ‘αγαπώ πάρα πολύ.

Not only was I born into a marvelous family, but I found more members to swell the ranks in all the places I have known. To Adam, Shayan, David, Avi, Franklin, Isaac, Woodson, Sachit, Jewel, Greg, Charles, Pame, José, Erik, Andrea, Giulio, Niloofar, Kazem: from high school to college to graduate school, you have filled my days with joy, creativity, and peace. I am thankful to my found family, and I promise to keep it on my mind and in my heart.

To Nicole: I never expected to find my eternal partner under the fluorescent lights of a sixth-floor Chemistry lab, yet here we are. My graduate life before you was largely comprised of lab work, peanut butter and jelly dinners, and a gnawing sense that my life outside lab can be something more. I am deeply grateful to you for uplifting my taste buds, deepening my appreciation of craftsmanship in all its forms, and providing the spark for forays into Nature to quite literally "smell the roses". To say it simply: you made me whole. I love you always.

This dissertation was funded in part by the NIH (Chemistry-Biology Interface training grant through T32GM112584)

Chapter 2 is reproduced, in part, with permission, from: Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F. A., "A. Metal-Templated Design of Chemically Switchable Protein Assemblies with High-Affinity Coordination Sites", *Angew. Chem., Int. Ed. Engl.* **59**, 21940–21944 (2020). The dissertation author was the primary author on all reprinted materials.

Chapter 4 is reproduced, in part, with permission, from: Kakkis, A.; Golub, E.; Choi, T. S.; Tezcan, F. A., "Redox- and metal-directed structural diversification in designed metalloprotein assemblies", *Chem. Commun.* **58 (49)**, 6958-6961 (2022). The dissertation author was the primary author on all reprinted materials.

VITA

EDUCATION

2016 B.A., Chemistry, Pomona College

2019 M.S., Chemistry, University of California San Diego

2023 Ph.D., Chemistry, University of California San Diego

HONORS AND AWARDS

2017-2019 NIH Chemistry-Biology Interface Training Grant

PUBLICATIONS

Kakkis, A.; Golub, E.; Choi, T. S.; Tezcan, F. A., "Redox- and metal-directed structural diversification in designed metalloprotein assemblies", *Chem. Commun.* **58 (49)**, 6958-6961 (2022)

Zhu, J.; Avakyan, N.; **Kakkis, A.;** Hoffnagle, A.; Han, K.;Li, Y.; Zhang, Z.; Choi, T.S.; Na, Y.; Yu, C.J.; Tezcan, F. A., Protein Assembly by Design. *Chem. Rev.* **2021**, *121* (22), 13701–13796

Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F.A., "A. Metal-Templated Design of Chemically Switchable Protein Assemblies with High-Affinity Coordination Sites", *Angew. Chem., Int. Ed. Engl.* **59**, 21940–21944 (2020)

FIELDS OF STUDY

Major Field: Chemistry

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

ABSTRACT OF THE DISSERTATION

Bottom-up design of metalloprotein assemblies with diverse metal coordination motifs and emergent properties

by

Albert Aris Kakkis

Doctor of Philosophy in Chemistry

University of California San Diego, 2023

Professor F. Akif Tezcan, Chair

Metal ions are indispensable to biological function, as they are utilized for a myriad of processes encompassing signaling, electron transfer, and catalysis. These functions are made possible by exploiting intrinsic properties of the metal ions and the ligands to which they bind, represented by small molecules and metalloproteins. From the perspective of a synthetic chemist, proteins can be conceived as “macromolecular ligands”. Like those of small molecule ligands, the

properties of "macromolecular ligands" can be parameterized at two discrete coordination spheres. The primary sphere is composed of the amino acid residues and exogenous ligands that coordinate the metal ion, and thus has the most direct impact on metal-based function. The secondary sphere is composed of residues that form mainly non-covalent interactions—hydrogen bonding, hydrophobic, and van der Waals (vDW) interactions—with the primary sphere. Although secondary sphere features do not directly bind the metal ion, they play an indispensable role in controlling metal-based reactivity. For example, the oxygen-binding affinity of myoglobin is influenced significantly by hydrogen bonding interactions between a histidine residue in its secondary sphere and the oxygen ligand that binds the heme co-factor. The large functional scope of metalloproteins is a testament to the wide diversity of primary and secondary spheres, and the efficiency with which metalloproteins execute these functions is a testament to the intricate interplay between the spheres. The functional potential of natural metalloproteins has long inspired protein design efforts. Two major metalloprotein design approaches are rational and *de novo* design. In rational design, primary and/or secondary sphere features of a natural protein are repurposed to mediate non-native functions. This approach has culminated in the engineering of new-to-nature functions spanning metal-hydride mediated ketone reduction and carbon-silicon bond coupling. While rational design represents an effective path to engineer metal-based functions, such functions are achieved within rigid and highly evolved protein folds/interfaces. Therefore, important questions remain unanswered: how does metal-based function emerge from an initially nonfunctional metalloprotein, and what are the minimum primary/secondary sphere coordination requirements to achieve metal-based functions? The work described in this dissertation stems from a *de novo* design approach in which both the quaternary structure and metal coordination site(s) of a metalloprotein are designed from scratch. The building block of our

designed metalloprotein structures is cytochrome cb_{562}, a natively monomeric, four-helix bundle protein. With cytochrome cb_{562} as our starting point, we obtain a diverse array of metalloprotein assemblies that serve as platforms for pursuing complex metal-based functions encompassing redox-based signaling and oxygen activation.

In Chapter 2, we describe the sequential design of three metalloprotein trimers, TriCyt1, TriCyt2, and TriCyt3. TriCyt1, our initial trimeric construct, is obtained through a single hydrophobic mutation (G70W) at the interface of a cytochrome cb_{562} variant. Solution experiments indicated that TriCyt1 assembled as a trimer in the presence of all mid-to-late first row transition metals (Mn^{II}-Zn^{II}), with yields ranging from 12% (+Mn^{II}) to 89% (+Co^{II}). Crystal structures of Ni^{II} and Cu^{II}-supplemented TriCyt1 revealed a biologically rare His_6 coordination motif. The only known natural His_6 motif is present in calprotectin, a metal sequestering protein that coordinates Mn^{II} with nanomolar affinity. Motivated by the prospect of obtaining from scratch a Mn^{II}:His_6 site of such high affinity, which had not yet been achieved in protein design, we sought to increase the preorganization of this metal coordination site through redesign of the C_2 and C_3 interfaces of TriCyt1. Computationally prescribed, mostly hydrophobic mutations at the C_3 interface led to TriCyt2, a construct that trimerized with near-quantitative yield in the presence of Mn^{II}. A second round of redesign, in which salt bridges were installed at the C_2 interface, led to a metal-independent, pH-switchable trimer which bound Mn^{II} with ~ 50 nM affinity. While a His_6 primary sphere is suitable for achieving high affinity metal binding sites, our ability to design metalloproteins with broad functional potential hinges on our ability to engineer multiple types of metal binding sites. In Chapter 3, we describe our efforts to diversify the primary and/or secondary spheres of the TriCyt scaffold in pursuit of metalloprotein constructs which could coordinate

lanthanides with high affinity and selectivity, mediate metal-dependent catalysis, and stabilize multinuclear metal coordination sites.

The TriCyt series illustrates metal-ion-identity-independent assembly: the assembly path converges on the same architecture with the same metal coordination motif regardless of the metal ion added. In the TriCyt series there is low cooperativity between metal-ligand and non-covalent interactions, as metal coordination preferences play a minimal role in directing assembly. In Chapter 4, we describe our design and characterization of a variant of Rosetta interface design cytochrome 1 ($^{A74/C96}$RIDC1) in which the cooperative interplay between metal-ligand and non-covalent interactions could be tuned by the redox state of the protein. In the oxidized state, C96-C96 disulfide bonds rigidify a tetrameric architecture ($^{A74/C96}$RIDC1$_4^{ox}$) whose assembly is independent of metal ion identity. Chemical reduction of the disulfide bonds gives rise to $^{A74/C96}$RIDC1^{red}, whose assembly is governed in larger part by the metal coordination preferences of exogenous metal ions. Whereas $^{A74/C96}$RIDC1$_4^{ox}$ can only access two structural states (apo and metal-bound), $^{A74/C96}$RIDC1^{red} can access three distinct, metal-ion-identity dependent structural states with unique coordination environments. With five structural states whose accessibility is dependent on solution redox potential and metal ion identity, $^{A74/C96}$RIDC1 represents a rare protein construct whose oligomerization path is varied through two different types of stimuli. We characterize the oligomeric and conformational states of $^{A74/C96}$RIDC1 under disulfide-oxidized and disulfide-reduced states and in the presence of different mid-to-late first row transition metal ions, both in solution and via X-ray crystallography. We also employ density functional theory (DFT) and Rosetta interface calculations to attempt to quantify the energetic contributions of metal-ligand and non-covalent interactions and rationalize how differences in metal ion identity result in divergent metal-directed assembly paths.

Chapter 1: Emulating and expanding on natural metalloprotein functions via design

1.1 Introduction

It is estimated that over 25% of proteins require a metal ion or metal co-factors to carry out functions critical to Life.[1] The biological importance of metal ions, while long acknowledged and explored by chemists, can also be recognized by the non-scientist, as just a cursory glance at the list of ingredients for a daily multivitamin reveals that a plethora of metal ions are required to ensure healthy bodily function. While many biologically important metal ions reside in the mid-to-late first row of the d-block (Mn^{II}-Zn^{II}), certain alkali metal ions and lanthanides also play important structural and functional roles. While the scope of metalloprotein function is vast–commensurate with the diversity of available metal ions–these functions will be broadly categorized herein as sequestration, hydrolysis, or oxygen activation (**Figure 1.1**). From our perspective, these types of functions represent three levels of complexity that we hope to achieve in our designed metalloprotein assemblies, with oxygen activation representing the pinnacle.

1.2 Natural Metalloprotein Functions

1.2.1 Metal Sequestration

Sequestration is the high affinity coordination of metal ions within a protein scaffold, often in response to a biological event/stimulus (**Figure 1.1a**).[2-5] It was discovered by Corbin *et. al.* that in the event of bacterial infection by *Staphylococcus aureus*, neutrophils release a metalloprotein called calprotectin that inhibits bacterial growth through the chelation of nutrient Mn^{II} near tissue abscesses.[6] This discovery established a new framework for conceptualizing host-pathogen interactions in which host-derived calprotectin engages in a "tug-of-war" with pathogen-derived metal transporters to prevent them from acquiring metal ions critical to their survival.[7, 8] Crystal structures of calprotectin reveal two EF-hand domains, a His_3Asp site, and a His_6 site (**Figure**

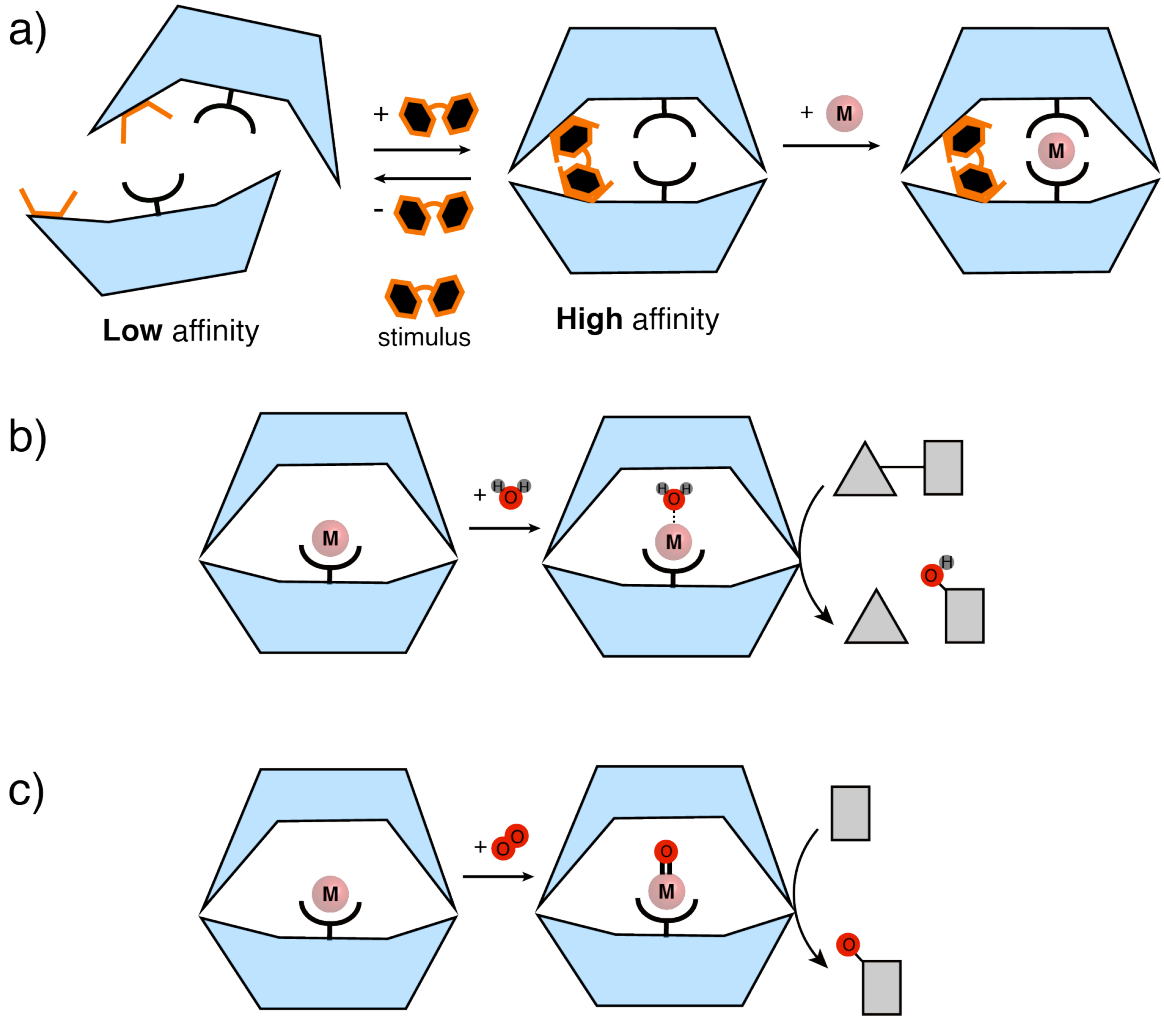

Figure 1.1 | Fundamental metalloprotein functions. (a) Sequestration, (b) Hydrolysis, (c) Oxygen activation. The types of stimuli triggering metal ion sequestration include metal ions, redox state changes, and small molecules. Within each category of metalloproteins, there are many different mechanisms for carrying out the general function.

1.2a).[8-12] Calcium binding through the EF-hand domains triggers oligomerization from a heterodimer, which binds transition metal ions (Mn^{II}, Fe^{II}, Ni^{II}, Zn^{II}) with low affinity, to a heterotetramer that binds these metal ions with at least nanomolar affinity.[10, 11, 13, 14] Thus, while the release of calprotectin by neutrophils is triggered by bacterial invasion, its ability to sequester metal ions from pathogens, and thus its antibacterial function, is stimulated by calcium.[6, 8, 10, 11, 15]

The biological role of calprotectin illustrates the importance of regulating metal localization and concentration in cells. The intracellular concentration of free metal ions is stringently controlled by small molecule and metalloprotein "chelators".[16] Metallothioneins represent a versatile class of metalloprotein chelator, as they can bind a wide range of metal ions such as Cu^{I}, Zn^{II}, Ag^{I}, and Cd^{II} (**Figure 1.2b**). [5, 17] Their abnormally high cysteine content (13-35%, in contrast to 2% across the proteome) allows them to accommodate up to eight metal ions within a single domain. [17, 18] Researchers have established a direct link between metalation state and tertiary structure, with the compactness of metallothionein structures correlating directly with metalation state.[18] Interestingly, the compact, "supermetallated" form of metallothionein is much

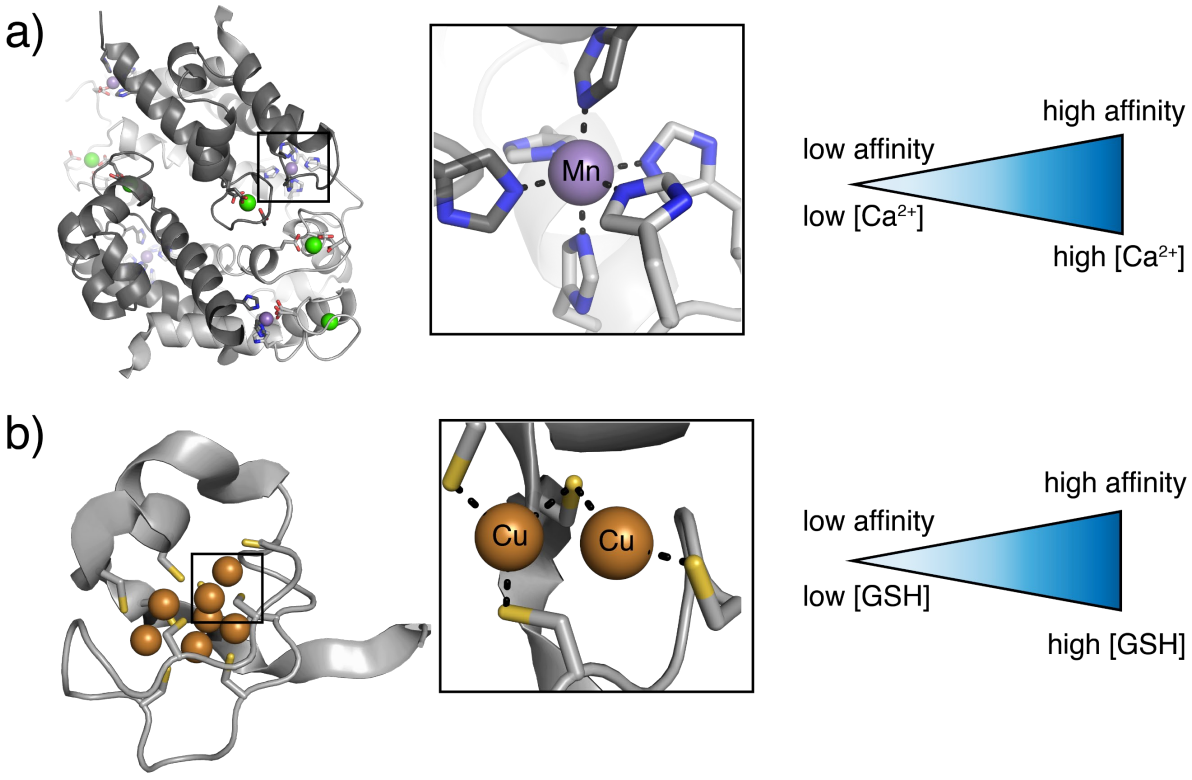

Figure 1.2 | Structural overviews of calprotectin and metallothionein. (a) Calprotectin (PDB ID: 4GGF). The binding of calcium (green spheres) to EF-hand domains pre-organizes the His_6 coordination sites (left). The Mn^{II}-binding affinity of the His_6 site increases by over 3 orders of magnitude in the presence of Ca^{II} (center and right). (b) Cu-metallothionein (PDB ID: 1AQR). The ability of this protein to sequester copper is directly tied to the cellular concentration of reduced glutathione (GSH), which is itself an indicator of the cellular redox state (right).

less prone to degradation than the apo form, suggesting that the intracellular metallothionein concentration is tied directly to the intracellular concentration of free metal ions.[19, 20] The high cysteine content of metallothionein renders it both a metal and redox sensor, as the metal binding affinity of cysteine residues can be reversibly lowered/increased through the formation/cleavage of disulfide bonds.[21, 22] Metallothioneins, with their ability to signal and participate in redox and metal equilibria changes, could represent an important hub in coordinating cellular responses to redox and/or metal-based stress.[23-25]

1.2.2 Hydrolysis

Metal ions can serve both structural roles–as evidenced above–and catalytic roles within their protein hosts. One important class of reactions mediated by metalloproteins is hydrolysis, which is linked to important biological processes spanning cellular signaling, pH regulation, and the degradation of insoluble substrates. [26-28] In hydrolysis, a water molecule is "activated" by the metalloprotein host to cleave a substrate (**Figure 1.1b**). Activating the water molecule is defined as lowering its pKa and thus the thermodynamic barrier for deprotonating it into a nucleophilic hydroxide ion. As exemplified by carbonic anhydrase and nitrile hydratase, two biologically important and extensively studied hydrolytic enzymes, the primary and secondary coordination spheres function in tandem to activate the water molecule and enable catalysis (**Figure 1.3**).[29-33]

Carbonic anhydrase catalyzes the reversible addition of water to carbon dioxide (CO_2) to produce bicarbonate (HCO_3^-), and thus has long been studied for its role in carbon dioxide transport (**Figure 1.3a**).[27, 34, 35] The protein hosts a mononuclear Zn^{II}:His_3 coordination site and an extensive hydrogen bond network that interacts with bound water and Zn^{II} coordinating residues (**Figure 1.3a**).[27, 29, 30] Kreba *et. al.* discovered that Thr-199 of human carbonic anhydrase II (hCAII) plays a critical role in the hydrolytic rate, presumably by stabilizing the Zn^{II}-hydroxide species through

hydrogen bonding to the active site water molecule.[29] Substitutions of Thr-199 with aliphatic residues, which would eliminate hydrogen bonding interactions, destabilized the Zn^{II}-hydroxide species by two pH units.[29] In addition to the electrostatics of the active site water, the electrostatics of the zinc center also influence the catalytic rate of hCAII.[30, 36] Kiefer *et. al.* found that replacing Gln-92 with a glutamate residue increased the pKa of the zinc-bound water by one pH unit,

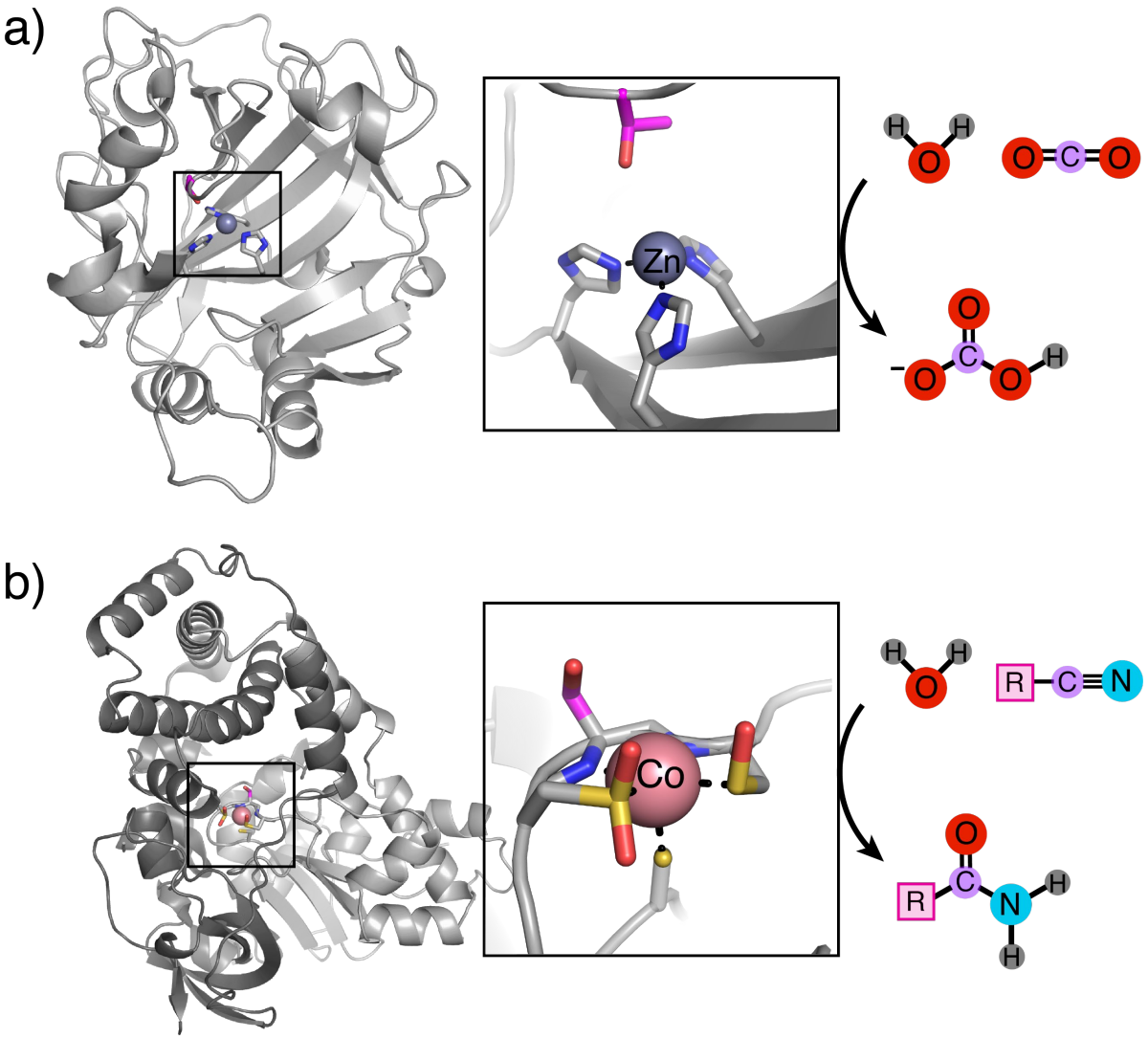

Figure 1.3 | Structural overviews of carbonic anhydrase and nitrile hydratase. (a) Carbonic anhydrase (PDB ID: 1ZNC). Thr199 (highlighted in magenta) has been implicated through mutagenesis studies to play a critical role in stabilizing the Zn^{II}-hydroxide species through hydrogen bonding. (b) Nitrile hydratase (PDB ID: 1UGP). It has been suggested that the serine residue proximal to the coordination site (highlighted in magenta) plays a role in hydrolysis by functioning as a general base.

presumably by increasing the electron density on zinc and thus lowering its Lewis acidity.[30] The decrease in the Lewis acidity of Zn^{II} corresponded to a 9-fold decrease in the CO_2 hydration efficiency (represented by k_{cat}/K_M).[30]

Nitrile hydratase (NHase), like carbonic anhydrase, utilizes both primary and secondary sphere interactions to increase the electrophilicity of its substrates (**Figure 1.3b**). NHase is an αβ heterodimer that catalyzes the hydrolysis of a wide array of nitrile-containing substrates into their amide forms.[32, 33] Given the ubiquity of amide-containing molecules such as acrylamide, nicotinamide, and 5-cyanovaleramide in organic synthesis, NHase has been used at industrial scale for decades.[32, 37] It binds either Fe(III) or Co(III) at a mononuclear, square pyramidal coordination site featuring three cysteine residues, two deprotonated nitrogen atoms from backbone amides, and an activated nitrile substrate in the axial position (**Figure 1.3b**).[33, 38] Fe(III) and Co(III), by virtue of their +3 oxidation states and high charge/radius (z/r) ratios, are strong Lewis acids. Nitrile coordination to the metal center via the nitrogen atom would increase polarization of the CN triple bond and the electrophilicity of the carbon atom, making the nitrile more prone to nucleophilic attack by a water molecule.[39]

The primary sphere of NHase is unlike that of any other enzyme, as coordinating thiols for two of the three cysteine ligands are oxidized to either sulfenic or sulfinic acids.[33, 40] Murakami *et. al.* established that modification of the thiol moiety of βCys114 into sulfenic acid is required for catalysis.[40] Several reaction mechanisms have been proposed wherein sulfenic acid carries out a nucleophilic attack on the carbon atom of the coordinated nitrile, forming a cyclic intermediate which is then cleaved by an activated water molecule.[39, 41] Although NHase and carbonic anhydrase both utilize activated water molecules to carry out catalysis, the coordination environment and catalytic mechanism of NHase is unique.

1.2.3 Oxygen activation

In both carbonic anhydrase and nitrile hydratase, the redox state of the metal ion does not change. However, some biologically important reactions such as alkane hydroxylation and thiol oxidation require that metal ions hosted within metalloprotein scaffolds adopt two or more redox states.[42, 43] Redox state changes of the metal center enable it to activate oxygen, reducing the diatomic molecule to generate metal-superoxide, metal-hydroperoxo, and/or metal-oxo species that ultimately oxidize substrates (**Figure 1.4**). [43-45] Although redox switching and oxygen reduction are universal features of oxygen activating metalloenzymes, their primary spheres, secondary spheres, and activation mechanisms vary widely in accordance with the variety of reactions they catalyze. Herein, we will describe two very different but important oxygen activating metalloenzymes: cytochrome P450 and cysteine dioxygenase (**Figure 1.4**). Unlike the metalloenzymes described thus far, which utilize "bare" metal ions for catalysis, cytochrome P450s utilize an iron-porphyrin metallocofactor to efficiently catalyze a wide range of reactions, including alkane hydroxylation, aromatic oxidation, and olefin epoxidation (**Figure 1.4a**).[46] The large reaction scope is due to the large number of P450 isozymes, each possessing unique protein pockets that select for substrate size, shape, and electronic characteristics.[44, 47] While the deviation of secondary spheres is high, many cytochrome P450s utilize the same primary sphere environment and oxygen activation mechanism to carry out different reactions.[46] The primary sphere is comprised of a porphyrin iron coordinated by four porphyrin-derived nitrogen ligands and an axial cysteine residue (**Figure 1.4a**).[44] In contrast to cytochromes involved only in electron transfer, which are hexacoordinate and thus coordinatively saturated, cytochrome P450s have one open site available to bind oxygen.[46, 48] The high preorganization and electron donating properties

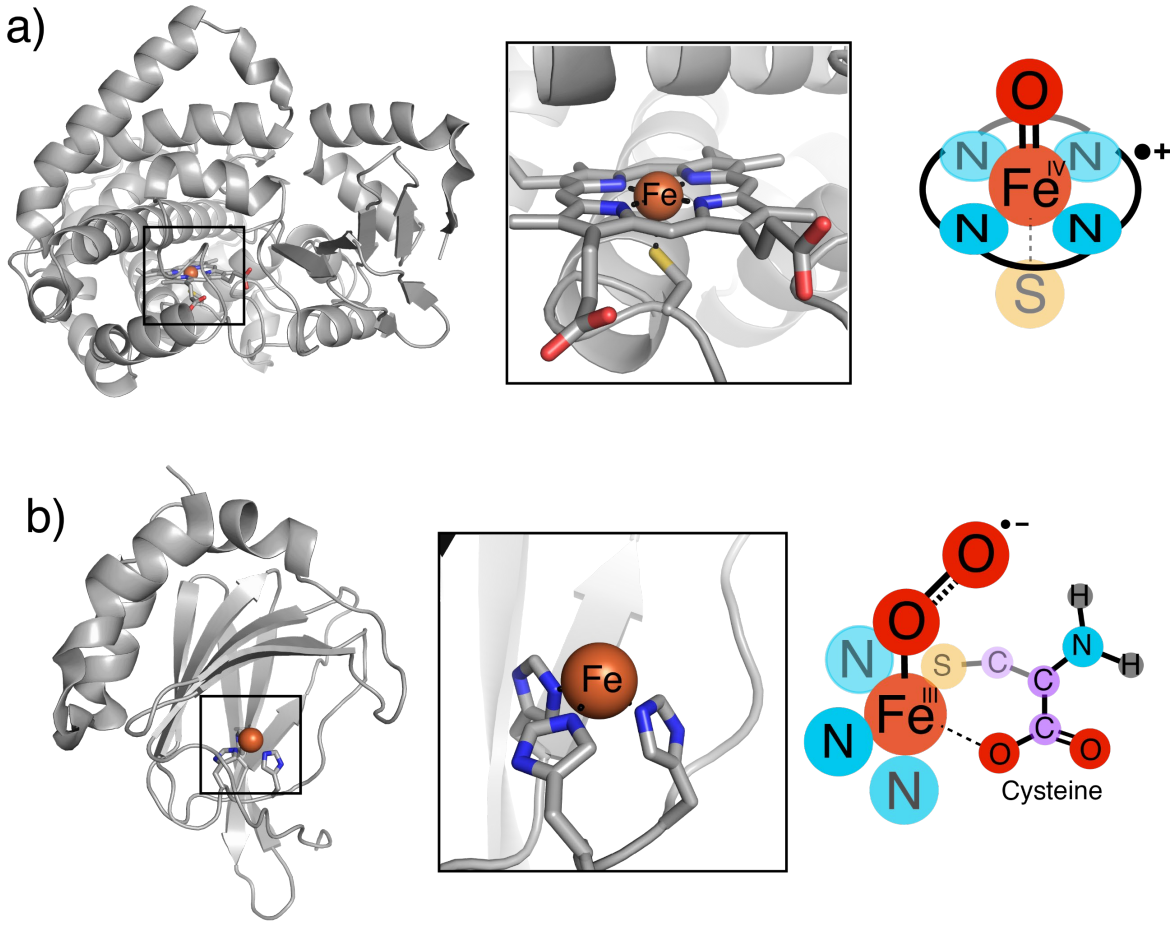

Figure 1.4 | Structural overviews of cysteine dioxygenase and cytochrome P450. (a) Cytochrome P450 (PDB ID: 1N97). The Fe^{IV}-oxo species (termed Compound I) forms without direct coordination by substrate, as the porphyrin ligand can stabilize Fe in multiple oxidation states. (b) Cysteine dioxygenase (PDB ID: 4IEV). The cysteine ligand is excluded in the close-up of the coordination site. The Fe^{III}-superoxide species can only form once the cysteine substrate coordinates the metal ion. It is hypothesized that the binding of cysteine lowers the reduction potential of the Fe^{II} coordination site, increasing the thermodynamic favorability of oxygen binding.

of the primary sphere allow the iron center to access oxidation states as high as +4 over the course of oxygen activation (**Figure 1.4a**).[44]

As the oxidation state of the iron center sequentially increases from +2 to +4, oxygen is sequentially reduced, culminating in the elimination of a water molecule and formation of compound I, an Fe(IV)-oxo species that ultimately carries out catalysis for a number of cytochrome

P450s.[44, 45] The reaction mechanisms of cytochrome P450s are decidedly more complex than those of hydrolytic enzymes due to redox switching of the metal center and multi-step, proton-coupled activation of oxygen. The need to couple oxygen activation to proton transfers makes oxygen activating metalloenzymes particularly difficult to design in protein scaffolds that lack an evolutionary "history" of this function.

Cysteine dioxygenase (CDO), in contrast to cytochrome P450, uses a mononuclear, Fe:His_3 coordination site to carry out catalysis. While its His_3 primary sphere is identical to that of carbonic anhydrase, it catalyzes a very different reaction, namely the oxidation of cysteine into cysteine sulfenic acid (**Figure 1.4b**). This oxidation reaction is a critical step in the biological catabolism of cysteine and thus the regulation of cellular cysteine levels.[49-51] In cysteine dioxygenase, oxygen activation is gated by the bidentate coordination of cysteine to the Fe^{II}:His_3 site, which is believed to lower the redox potential of the Fe^{II} center and consequently the thermodynamic barrier for electron transfer to oxygen.[43] Electron transfer from the Fe^{II} center to oxygen is hypothesized to yield an Fe^{II}-superoxo complex that is stabilized by the protein pocket, a testament to the influence of the secondary coordination sphere in modulating reactivity.[52] In addition to stabilizing the Fe^{II}-superoxo species, the secondary coordination sphere may also impact both the optimal pH for catalysis and substrate positioning.[43, 53, 54] Proximal to its primary sphere, CDO hosts a cysteine residue that is covalently linked to a tyrosine through a thioether bond, an especially rare post-translational modification (**Figure 1.4b**).[53-55] Recent studies by Davies *et. al.* found that this crosslink ensures high catalytic efficiency at pH values above 7.[54] It is hypothesized that the incorporation of the secondary sphere cysteine residue into a covalent crosslink prevents its thiolate functionality from reacting with the cysteine substrate at elevated pH, thereby ensuring that substrate binding is solely coupled to oxygen activation.[53, 54] CDO provides a notable example

of how metalloprotein pockets serve to not only activate the target reaction but also to inhibit unwanted side reactions.

1.3 Metalloenzyme re-engineering

1.3.1 General approaches

The metalloproteins described above represent just a small fraction of functional metalloproteins present in Nature. Metalloenzymes not only catalyze a broad range of reactions but execute these reactions with a catalytic efficiency and selectivity that is difficult to emulate in small molecule systems. When compared to small molecule catalysts, metalloenzymes benefit immensely from a vast primary (amino acid composition), secondary (α-helical/β sheet composition), tertiary (discrete folds), and quaternary (interdomain interactions) structural space. Improving the performance of a small molecule catalyst can require laborious, synthetically challenging modifications. To improve the performance of a metalloenzyme, one can harness evolution and thus more efficiently expand the sequence space. While metalloenzymes are not natural catalysts for every reaction of synthetic interest, their efficiency, selectivity, and evolvability has inspired synthetic chemists to utilize them as macromolecular hosts for new-to-nature reactions. This general principle of "metalloenzyme re-engineering", can be further classified into two approaches, metalloenzyme repurposing and artificial metalloenzyme design (**Figure 1.5**). In the former, both the primary and secondary spheres of the natural metalloprotein are harnessed for catalysis, whereas in the latter, the primary sphere is derived from a synthetic co-factor while the secondary sphere/protein pocket is derived from the protein (**Figure 1.5**). We will describe in detail two well-known examples of metalloenzyme re-engineering: the repurposing of cytochrome P450 to mediate new-to-nature reactions and the utilization of the streptavidin-biotin interactions to construct artificial metalloenzymes.

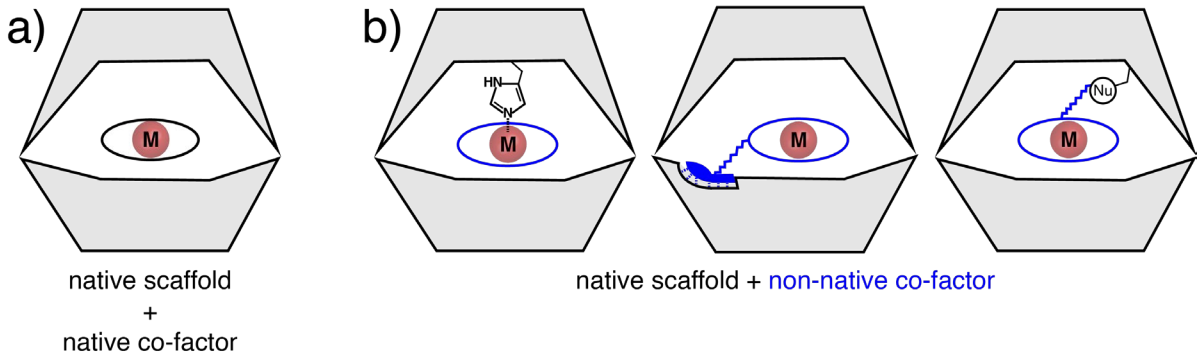

Figure 1.5 | Metalloenzyme re-engineering approaches. (a) Metalloenzyme repurposing. (b) Artificial metalloenzyme design. The co-factor (blue) can be incorporated into protein hosts through a variety of methods, including direct coordination by a protein-derived ligand ("dative anchoring", left), non-covalent interactions between the protein host and a co-factor moiety ("supramolecular anchoring", center), and covalent binding by a protein-derived nucleophile ("covalent anchoring", right).

1.3.2 Repurposing of cytochrome P450 for new-to-nature reactions

As described previously, cytochrome P450s represent a class of enzymes that perform a broad range of reactions while uniformly employing an iron porphyrin as their functional unit. The Arnold Group has harnessed the functional promiscuity of the iron porphyrin co-factor, modifications of the primary sphere residues, and directed evolution of the protein pocket to render cytochrome P450 variants that perform new-to-nature reactions , including carbene and nitrene transfer, with exquisite efficiency and selectivity (**Figure 1.6a**).[56] It was determined by the Arnold Group that the iron porphyrin co-factor can not only bind oxygen to form Fe-oxo species, but also nitrene and carbene precursors to render Fe-nitrene and Fe-carbene species.[56-59] Without modifying the original primary sphere environment of cytochrome P450, these species can mediate a broad range of reactions *in vitro*, including aziridination, amination, and cyclopropanation (**Figure 1.6b**).[57-59] Unlike Fe-oxo, the Fe-nitrene and Fe-carbene are catalytically active when iron is in the +2 oxidation state. [56, 60] In order to maintain this oxidation state *in vivo*, the Arnold Group replaced the axial cysteine residue in the primary sphere with a serine and thus increased the reduction potential from -430 to -293 mV, allowing for the iron center to be reduced by biological

reductants like NADPH ($E°'$= -320 mV).[59, 61, 62] This single mutation, along with other primary sphere mutations, has expanded the biological compatibility and the substrate scope of reactions catalyzed by reengineered cytochrome P450s.[58, 59, 61, 63]

Mutations of the primary sphere are often paired with directed evolution of the protein pocket surrounding the Fe-porphyrin to improve turnover and increase enantioselectivity (**Figure 1.6b**).[56, 64, 65] As a notable example of the reaction novelty achievable in reengineered proteins hosting Fe-porphyrins, the Arnold Group recently evolved a cytochrome *c* variant to promote the formation of carbon-silicon bonds, which have no biological precedence.[65] In addition to cytochrome P450, azurin and carbonic anhydrase have served as natural metalloprotein scaffolds for engineering new-to-nature functions encompassing electron transfer and hydride reduction.[66-70]

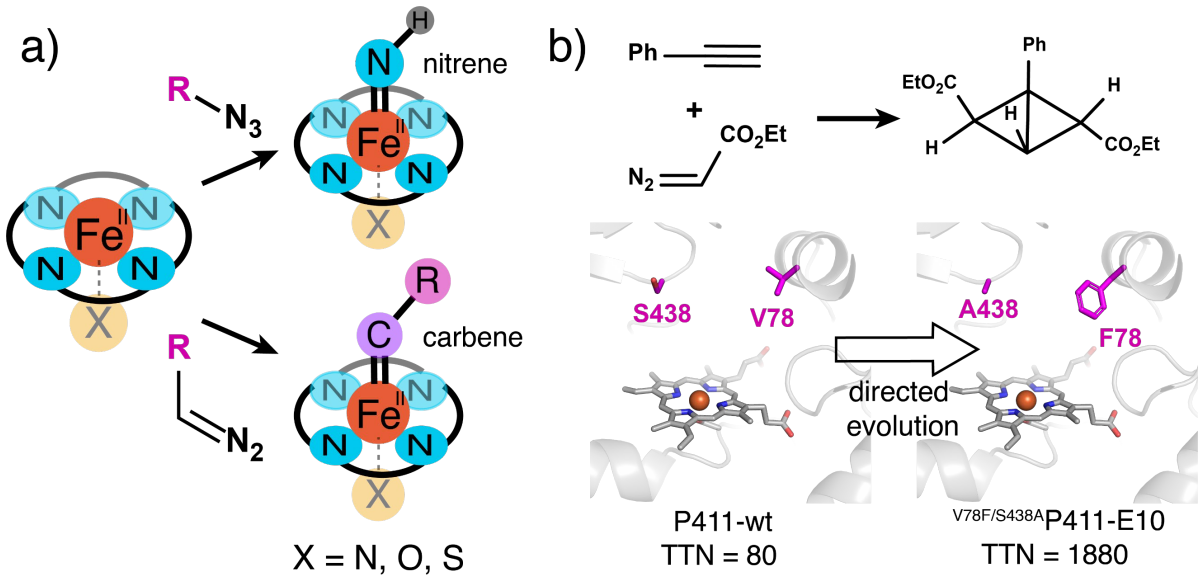

Figure 1.6 | Repurposing of cytochrome P450 for new-to-nature reactions. (a) Reaction schemes for forming Fe^{II}-nitrene (top) and Fe^{II}-carbene (bottom) species in cytochrome P450 using azide and ethyl diazoacetate derivatives. (b) Reaction scheme for the formation of a bicyclic compound mediated by carbene transfer onto a cytochrome P450 variant (top). Directed evolution of wild-type cytochrome P411 (axial cysteine of P450 replaced by serine) led to significantly higher catalytic efficiency for the cyclization reaction, with TTN increasing ~24-fold.

1.3.3 Biotin-streptavidin host-guest pair for artificial metalloenzyme design

In the cytochrome reengineering efforts described above, the primary sphere is derived from the native metalloprotein. With an eye toward expanding the functional scope of metalloenzymes farther, researchers have used proteins as hosts for abiotic metal complexes, with the protein being harnessed to tune reactivity/enantioselectivity through outer sphere interactions, shield metal complexes from aqueous bulk solution, and enable directed evolution of metal-based functions.[71] To bind the metal complexes to protein pockets, researchers synthesize molecules that feature a metal complex moiety and an "anchor" moiety (**Figure 1.5b**).[71] Anchoring is usually achieved through dative coordination of a native metal-binding residue to a coordinatively unsaturated metal complex, non-covalent interactions with the protein, or covalent immobilization proximal to a protein pocket (**Figure 1.5b**).[71] The Ward Group has anchored metal complexes to protein pockets by exploiting high affinity biotin/streptavidin and sulfonamide/Zn^{II} interactions, which has led to the construction of a wide range of artificial metalloenzymes that perform a wide range of reactions both *in vitro* and *in vivo* (**Figure 1.7**). [71-75] The biotin-streptavidin approach has been especially fruitful owing to the high stability of streptavidin and the high affinity ($K_d \sim 10^{-17}$ M) of the interaction, and has given rise to metalloenzymes that perform reactions encompassing imine reduction, metathesis, and C-H activation (**Figure 1.7**).[74] Recently, the Ward Group has identified strategies such as cell surface display and periplasmic localization which have mediated the directed evolution of their artificial metalloenzymes *in vivo*.[76] The biotin-streptavidin and sulfonamide-Zn^{II} strategies are a small fraction of the myriad approaches researchers have adopted to achieve abiotic catalysis within biological protein hosts.[71] Numerous covalent chemistries such

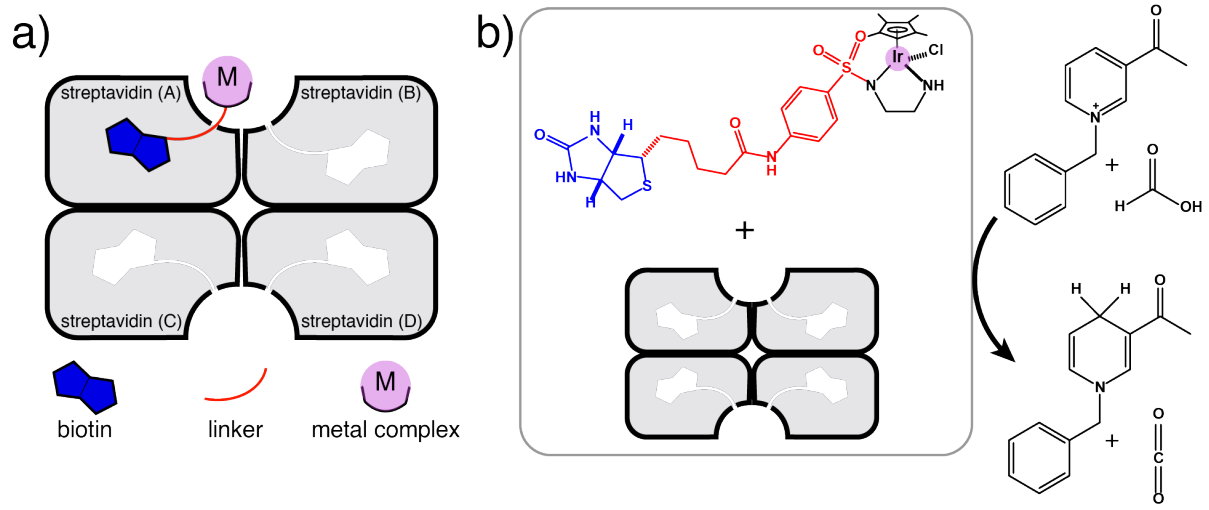

Figure 1.7 | Artificial metalloenzyme design using biotin-streptavidin anchoring strategy. (a) Cartoon scheme of the biotin-streptavidin anchoring strategy, adapted from [76]. (b) Reduction of NAD^+ derivatives mediated by a streptavidin-bound iridium piano stool complex.

as cysteine-maleimide conjugation and the azide-alkyne click reaction have been exploited to immobilize a diverse set of metal complexes into a diverse set of protein hosts.[77-79]

1.4 *De novo* metalloprotein design

The engineering of metal-based functions within natural protein scaffolds has proven to be a relatively straightforward and thus industrially useful approach to design metalloenzymes. However, utilizing complex, highly evolved scaffolds restricts one's ability to assess how metalloprotein structure is coupled to function, and whether complex metal-based functions necessitate complex metalloprotein structures. Constructing a metalloprotein *de novo* offers the potential to establish the minimum requirements of a protein scaffold to mediate metal-based functions, as the building blocks tend to be structurally simple. For example, through *de novo* design a researcher can address whether stabilizing a Zn-His_3 primary sphere within a protein scaffold is the only requirement for hydrolytic activity. Most *de novo* designed metalloprotein constructs are obtained using peptide building blocks that fold into target structures (**Figure 1.8**).[80] The "protein folding problem" has largely limited the structural scope of *de novo* designed

metalloproteins to α-helical bundles and coiled coils, yet these structurally simple constructs have been shown to mediate a wide range of catalytic activities (**Figure 1.8**).[81, 82]

Two prominent structural motifs that have been engineered for metal-based functions are the four-helix bundle and the trimeric coiled coil (**Figure 1.8**).[83] The DeGrado Group observed that Δ^9 ACP desaturase, while on the whole a structurally complex metalloenzyme, hosts its catalytically active di-iron site within a simple four-helix bundle domain.[84] Using computationally guided peptide sequence design, the DeGrado Group recapitulated the di-iron site containing four-

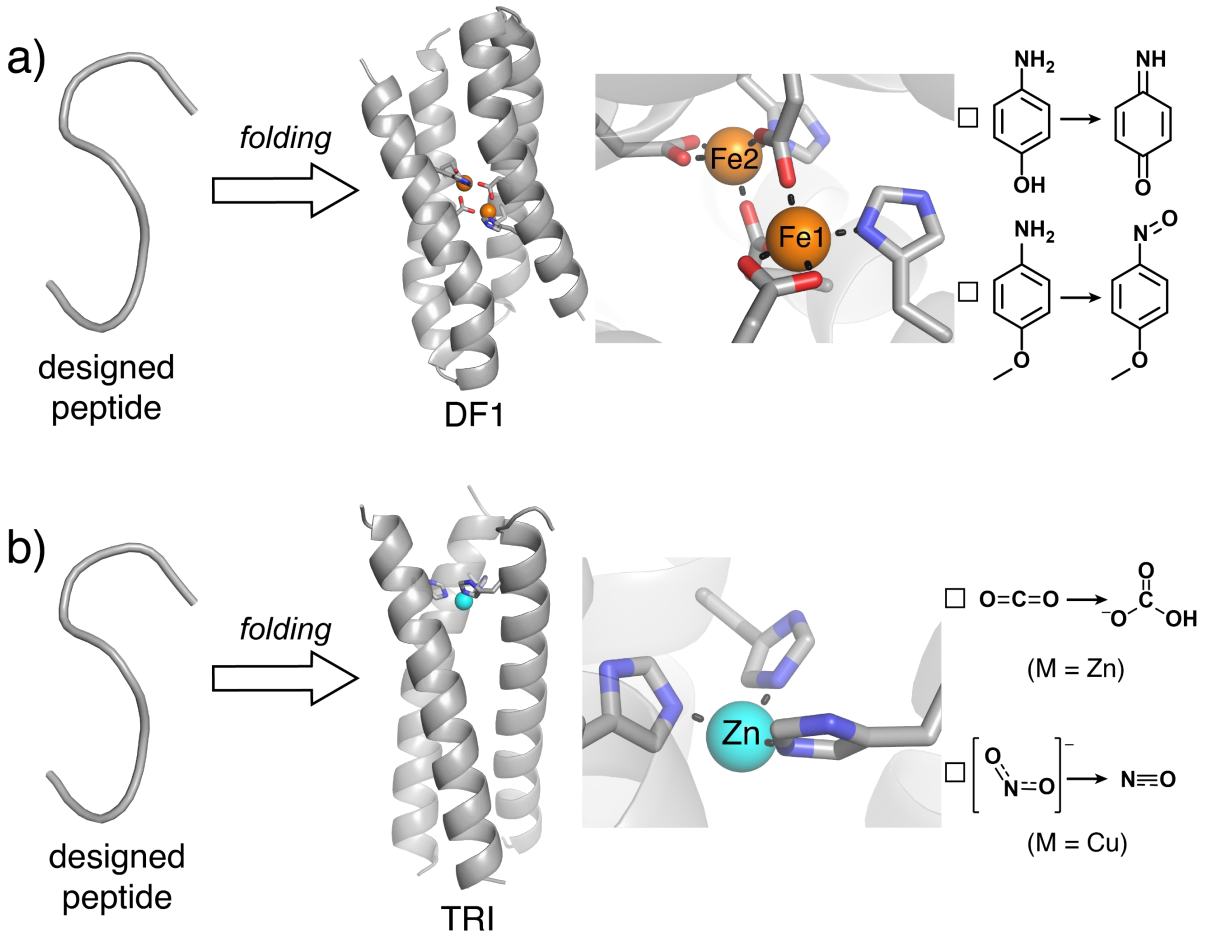

Figure 1.8 | Examples of *de novo* designed metalloenzyme scaffolds. (a) DF1 (PDB ID: 1JMB). The dinuclear site can accommodate metal ions beside Fe, including Zn^{II}. Fe^{II}-bound DF1 variants have been shown to catalyze aminophenol oxidation and p-anisidine oxygenation (right). (b) TRI (PDB ID: 3PBJ). The His_3 site of this scaffold, when bound to Zn^{II} or Cu^{II}, can mediate CO_2 hydration and one electron reduction of nitrite (right).

helix bundle in a new peptide construct, termed "Due Ferri 1" (DF1, **Figure 1.8a**).[84] In subsequent studies, modifications of the primary sphere and peptide pocket gave rise to constructs that mediate amine oxidations, stabilize radical species, and even accommodate tetranuclear zinc clusters.[85-88] In contrast to the monomeric, four-helix bundle generated by the DeGrado Group, the Pecoraro Group developed a coiled-coil peptide that self-assembles into a trimer, α_3D (**Figure 1.4b**).[82] In early studies, the Pecoraro Group demonstrated that this homotrimeric construct can accommodate a Cys_3 coordination site that binds heavy metals such as Hg^{II}, Cd^{II}, and As^{III}.[89, 90] In later studies the thiol-rich site was replaced with a His_3 site which could stably bind Cu^{I}, Cu^{II}, and Zn^{II}, facilitating metal-based functions such as CO_2 hydration and nitrite reduction (**Figure 1.8b**).[91, 92] Recently, the Pecoraro Group successfully carried out quantum mechanics/molecular mechanics (QM/MM)-guided design of an A_2B heterotrimer which had improved p-nitrophenyl acetate (p-NPA) hydrolysis activity compared to the parent homotrimer, illustrating how, as in natural metalloproteins, asymmetry plays a valuable role in mediating efficient catalysis in designed metalloproteins.[93, 94]

1.5 Protein self-assembly in Nature and by design

1.5.1 Natural protein assemblies

While the *de novo* design of metalloproteins using peptide building blocks has yielded several functional constructs, the reliance on such building blocks limits the structural/functional complexity and evolvability of designed metalloproteins, which are key elements of functional metalloenzymes in Nature. In Nature, complex functions can be achieved by evolving new protein-protein interactions (PPIs) at interfaces of simple protein building blocks.[95-97] Four-helix bundle proteins, while on their own structurally simple and functionally limited, can be incorporated through PPIs into complex assemblies with functions encompassing signaling, electron transfer,

and metal storage (**Figure 1.9**).[98-100] Ferritin, a protein cage comprised of 24 four-helix bundle proteins and involved in iron mineralization/storage, is a notable example of how function emerges through protein self-assembly (**Figure 1.9c**).[96] While the four-helix bundle proteins cannot accumulate iron individually, they assemble through extensive non-covalent interactions at six symmetry-related interfaces (one two-fold, two three-fold, three four-fold) into the functional cage.[96] Ferritin is just one example of many natural metalloprotein assemblies (*e.g.*, insulin, hemoglobin, and nitrogenase) that perform metal-based functions critical to Life (**Figure 1.10**).[95]

Each metalloprotein assembly varies markedly with regards to the structure, stoichiometry, and heterogeneity of their protein building blocks.[95] An important example of how structural and

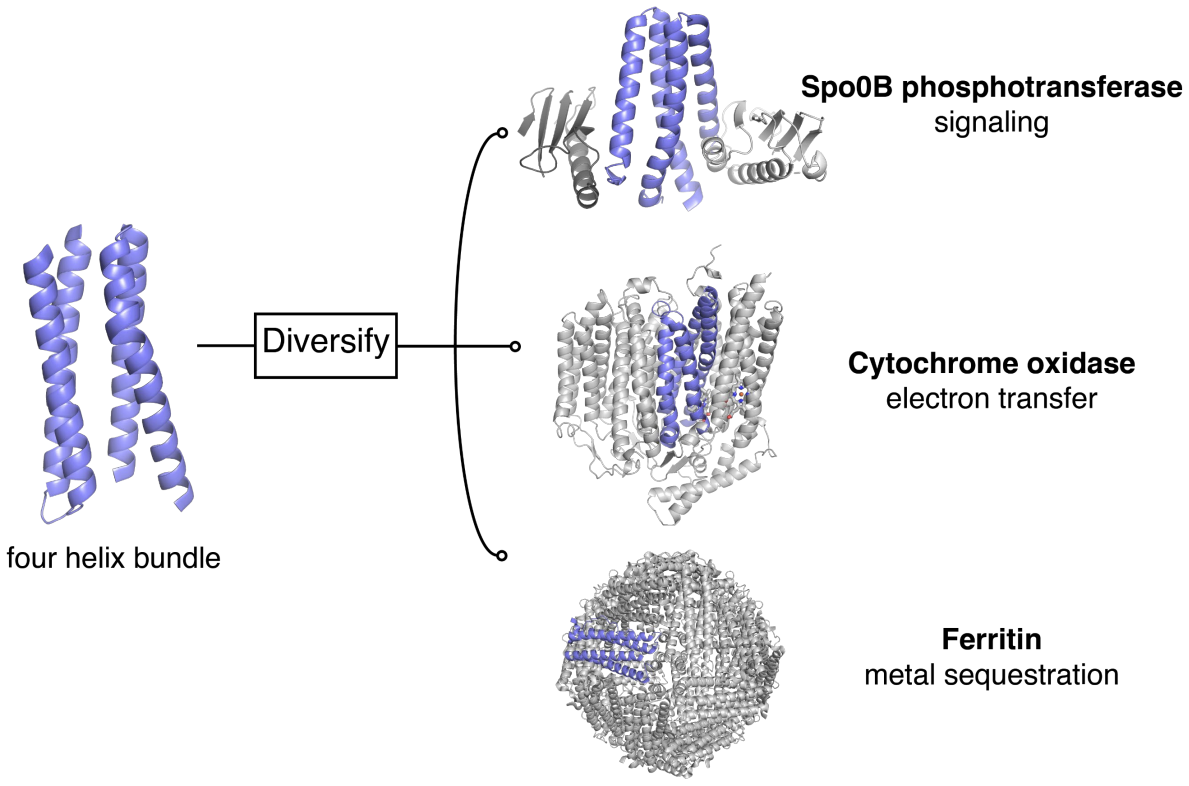

Figure 1.9 | Incorporation of the four-helix bundle motif into functionally diverse oligomers. Four-helix bundle domains are highlighted in blue. PDB IDs: 1IXM (top), 6RX4 (middle), and 6B8F (bottom).

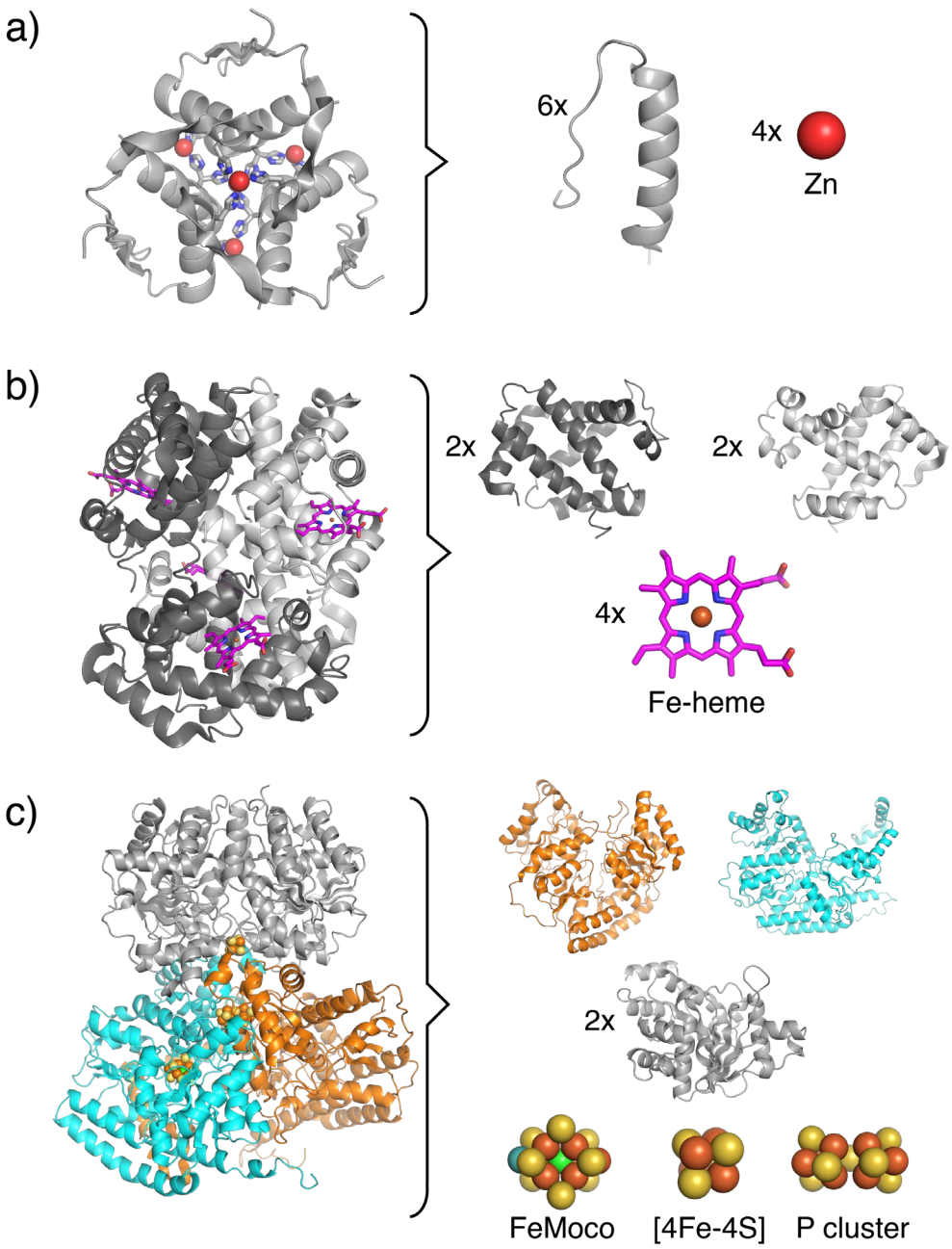

Figure 1.10 | Constituent domains and metal co-factors of natural metalloprotein assemblies. (a) Insulin (PDB ID: 1ZNI), a signaling peptide, assembles into a hexamer in the presence of Zn^{II} (b) Hemoglobin (PDB ID: 1HHO), an oxygen transporter, forms a heterotetrameric assembly that binds Fe-heme co-factors. (c) Nitrogenase (PDB ID: 1M1N), the only enzyme known to reduce N_2 to NH_3, is a complex heteromeric assembly comprised of three unique domains and metal co-factors.

functional complexity are tightly coupled, nitrogenase consists of multiple discrete domains which couple multiple functions (nucleotide binding, electron transfer, nitrogen reduction) within a single

macromolecular machine (**Figure 1.10c**).[95, 101] Given the ubiquitous imprint of protein self-assembly in natural metalloproteins, it represents a promising tool for the *de novo* design of functional metalloproteins. The assembly of folded proteins into quaternary structures and the folding of polypeptide chains into tertiary structures are distinct processes. Compared to protein folding, protein self-assembly is highly concentration-dependent, acutely sensitive to physiological conditions (pH, salt composition/concentration), requires specific interactions between chemically distinct components, and has a "rougher" free energy profile.[95, 102-104] Algorithms like Alphafold have greatly improved researchers' ability to predict protein folds based on primary sequence alone.[105] No analogous algorithm has been developed that can accurately and reproducibly predict self- assembly paths based solely on the sequences and/or structures of folded proteins.[95] Despite the challenges posed by the "protein assembly problem", numerous approaches have been developed to assemble proteins into larger order structures.[95] Herein, we will describe two approaches, computational design and metal-directed self-assembly, as these are most relevant to our designed metalloprotein constructs. An extensive survey of protein assembly design methods is provided by the Tezcan Group in a recent review.[95]

1.5.2 Computational protein assembly design

The computational design of protein assemblies is rooted in protein-protein docking, wherein the structure of a protein-protein complex is predicted based on the structures of the individual proteins.[106, 107] Our advances in protein-protein docking stem from improvements in sampling techniques and scoring functions.[106, 107] To efficiently sample as many conformations as possible, researchers have relied on fast Fourier transform (FFT) docking.[107, 108] FFT docking generates an ensemble of structures that are then scored based on steric and physicochemical complementarity.[107] Scoring steric complementarity requires the prediction of preferred rotamer

conformations for interfacial residues, whereas scoring physicochemical complementarity requires the prediction of favorable interactions (vdW, electrostatic, H-bonding) between those residues (**Figure 1.11a**).[106] Score functions often incorporate physical parameters based on free energy functions and empirical parameters based on high resolution crystal structures of known protein-protein complexes.[106, 109, 110] The pairing of efficient sampling methods with effective energy functions has enabled computational design to be applied beyond the realm of protein dimer prediction and into the realm of *de novo* self-assembly of complex and functional protein architectures (**Figure 1.11b**).[109-111]

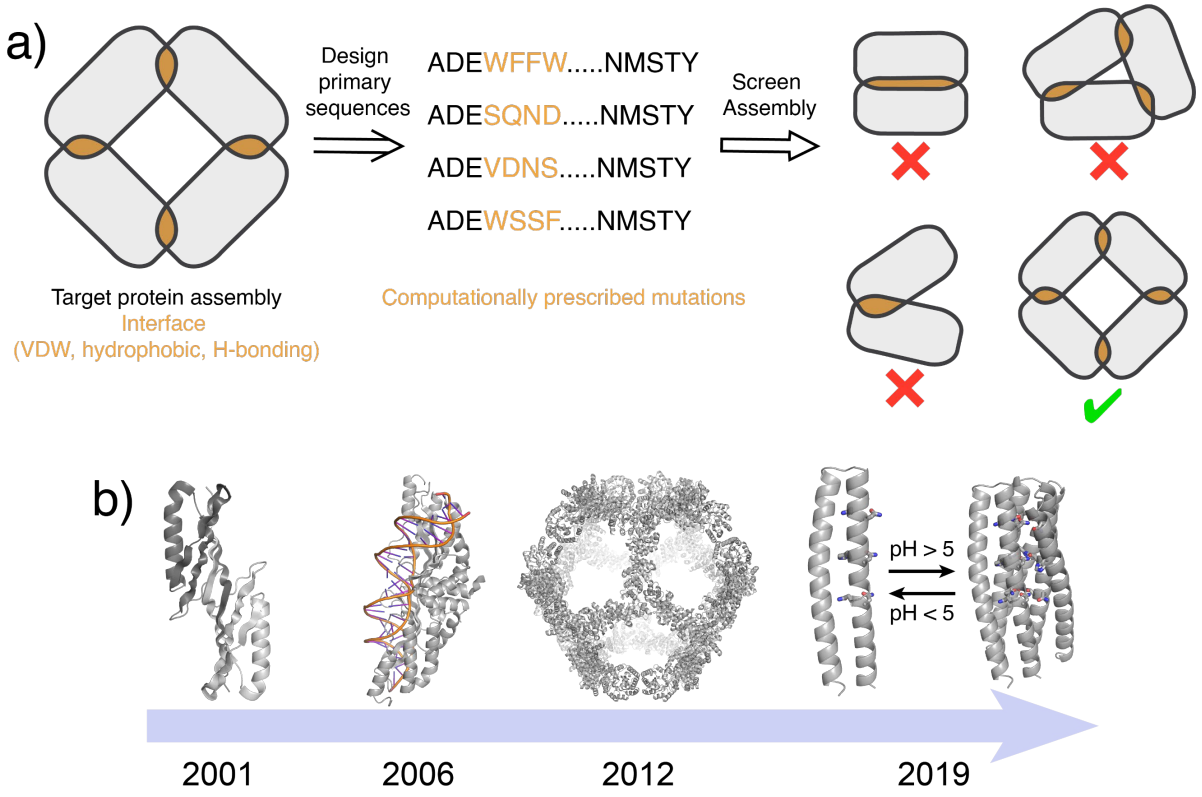

Figure 1.11 | General computational protein design methodology and examples of computationally designed/redesigned protein assemblies. The assemblies include a homodimer (left, PDB ID: 1JML), an endonuclease redesigned for altered DNA cleavage specificity (center left, PDB ID: 2FLD), an octahedral cage protein (center right, PDB ID: 4DDF), and a heterodimer whose assembly is reversibly controlled through pH changes (right, PDB ID: 6MSQ).

A ubiquitous algorithm in *de novo* computational protein design is Rosetta, developed by the Baker Group.[110] With improvements in the Rosetta energy function, the scope of successful designs has expanded from discrete oligomers to extended assemblies.[112, 113] Characterization of many Rosetta-generated designs reveals close correlation to the computationally predicted structure, a testament to the high atomic accuracy attainable through the algorithm. The recent emergence of a new Rosetta-based method for identifying hydrogen bonds, HBnet, has led to the design of pH-switchable assemblies (**Figure 1.11b**) and highly specific hetero-oligomers that function as protein logic gates *in vivo*.[114-116] One significant limitation of the Rosetta energy function is its inability to account for metal-ligand interactions, which, given their relatively high thermodynamic stabilization, would be a useful interaction type for protein self-assembly. The interaction types incorporated into the Rosetta energy function (vdW, hydrogen bonding, hydrophobic) tend to be weaker and of lower specificity (**Figure 1.11**). Consequently, computational protein design incurs a large design footprint, as the generation of a target structure requires extensive mutagenesis of the protein building blocks.[117]

1.5.3 "Metal-directed" protein assembly design

Given their strength, tunability, and geometric specificity, metal-ligand interactions are an attractive interaction type to harness for the assembly of proteins into stable, reversible, and well-defined architectures, all while incurring a smaller design footprint (**Figure 1.12**). In contrast to synthetic chemists, who for many decades utilized metal coordination to construct discrete and extended supramolecular architectures, the explicit utilization of metal-ligand interactions for *de novo* protein self-assembly was only first reported in 2007 by the Tezcan Group.[118-121] They found that the addition of Zn^{II} to a variant of cytochrome cb_{562} bearing two *i/i+4* bis-His motifs (metal binding cytochrome-1, or MBPC1) reversibly triggered its assembly into a tetramer hosting four

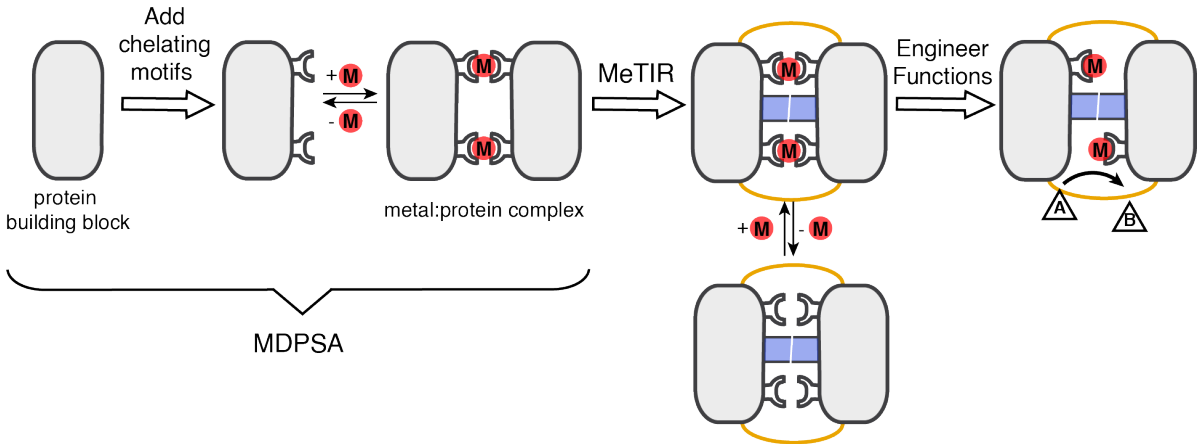

Figure 1.12 | Overview of MDPSA and MeTIR. The protein building block should be natively monomeric. Designed noncovalent and covalent interactions are highlighted in blue and yellow, respectively.

tetrahedral, His_3Asp coordination sites (Zn_4:MBPC1$_4$).[121] The designed assembly only required three mutations on the protein surface, indicative of a small design footprint. A subsequent study by Salgado *et al* revealed that depending on the coordination geometry preferences of exogenously added mid-to-late first row transition metal ions (Ni^{II}: octahedral, Cu^{II}: square planar, Zn^{II}: tetrahedral), MBPC1 crystallized into different oligomers (Ni^{II}: trimer, Cu^{II}: dimer, Zn^{II}: tetramer) (**Figure 1.13**).[122]

In this new protein design method, termed metal-directed protein self-assembly (MDPSA), the metal-ion-identity specific nature of metal coordination enables a single building block to access multiple structural states. This is difficult to achieve via design methods that solely rely on weak, non-covalent interactions, which lack the geometric specificity and diversity of metal-ligand interactions. However, in contrast to computational protein design, the lack of an energy function to approximate metal-ligand interactions renders it very difficult to establish the assembly path *a priori*. Indeed, it is likely that absent a corresponding energy function, MDPSA will remain a less popular tool for *de novo* protein self-assembly compared to Rosetta.

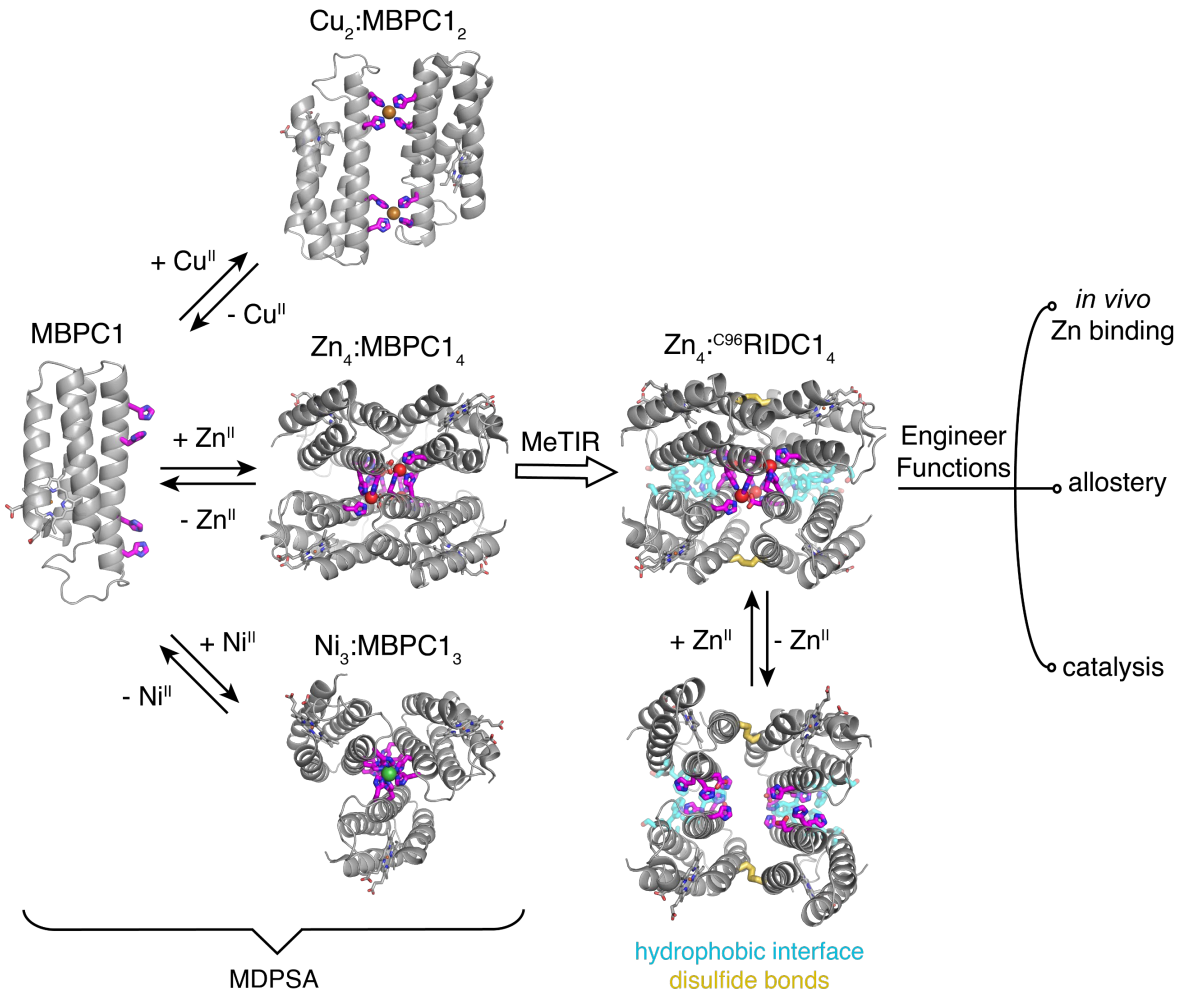

Figure 1.13 | Metalloprotein assemblies obtained via MDPSA and MeTIR. PDB IDs: 3DE8 (Cu_2:MBPC1$_2$), 2QLA (Zn_4:MBPC1$_4$), 3DE9 (Ni_3:MBPC1$_3$), 3IQ6 (Zn_4:^{C96}RIDC1$_4$), 3IQ5 (^{C96}RIDC1$_4$).

As a corollary of the geometric diversity of metal coordination complexes, metal complexes can mediate a wide range of functions spanning sequestration and catalysis. Functional small molecule complexes often feature ligand environments that are rigidified through covalent tethering, as exemplified by the scorpionate ligands.[123, 124] Analogous to these small molecule strategies, the Tezcan Group installed non-covalent and covalent interactions at the interfaces of MBPC1-based assemblies, a strategy termed metal-templated interface redesign (MeTIR), to render more stable architectures that preorganize the metal coordination environments (**Figure 1.12-13**).[95, 125-127] One such assembly, Rosetta Interface Design Cytochrome-1 (RIDC1), was

obtained by using the crystal structure of Zn_4:MBPC1$_4$ as a template for installing auxiliary non-covalent interactions .[125] The Zn-directed RIDC1 tetramer, Zn_4:RIDC1$_4$ has served as a starting point for engineering a variety of functions, including selective Zn binding, *in vivo* Zn sequestration, allostery, and ampicillin hydrolysis (**Figure 1.13**).[126, 128-131] The consolidation of MDPSA and MeTIR into a single workflow exploits the best aspects of metal-directed and computational protein design approaches. In MDPSA, we use rationally guided metal-ligand interactions to obtain a robust initial assembly with minimal sequence design, and with computational protein design we can apply an algorithmic approach to stabilize this emergent assembly in a predictable manner. The interplay between metal-ligand and non-covalent interactions at designed interfaces is a central theme of subsequent chapters, which highlight various cytochrome *cb*$_{562}$ mutants hosting a diverse array of metal coordination environments.

1.6 References

1. Waldron, K. J.; Robinson, N. J., How do bacterial cells ensure that metalloproteins get the correct metal? *Nat. Rev. Microbiol.* **2009,** *7* (1), 25-35.
2. Vijver, M. G.; Van Gestel, C. A.; Lanno, R. P.; Van Straalen, N. M.; Peijnenburg, W. J., Internal metal sequestration and its ecotoxicological relevance: a review. *Environ. Sci. Technol.* **2004,** *38* (18), 4705-12.
3. Phillips, D. J. H.; Rainbow, P. S., Strategies of trace metal sequestration in aquatic organisms. *Mar. Environ. Res.* **1989,** *28* (1-4), 207-210.
4. Pearce, L. L.; Gandley, R. E.; Han, W.; Wasserloos, K.; Stitt, M.; Kanai, A. J.; McLaughlin, M. K.; Pitt, B. R.; Levitan, E. S., Role of metallothionein in nitric oxide signaling as revealed by a green fluorescent fusion protein. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97* (1), 477-82.
5. Coyle, P.; Philcox, J. C.; Carey, L. C.; Rofe, A. M., Metallothionein: the multipurpose protein. *Cell Mol. Life Sci.* **2002,** *59* (4), 627-47.
6. Corbin, B. D.; Seeley, E. H.; Raab, A.; Feldmann, J.; Miller, M. R.; Torres, V. J.; Anderson, K. L.; Dattilo, B. M.; Dunman, P. M.; Gerads, R.; Caprioli, R. M.; Nacken, W.; Chazin, W. J.; Skaar, E. P., Metal chelation and inhibition of bacterial growth in tissue abscesses. *Science* **2008,** *319* (5865), 962-5.
7. Potrykus, J.; Ballou, E. R.; Childers, D. S.; Brown, A. J., Conflicting interests in the pathogen-host tug of war: fungal micronutrient scavenging versus mammalian nutritional immunity. *PLoS Pathog.* **2014,** *10* (3), e1003910.
8. Zygiel, E. M.; Nolan, E. M., Transition Metal Sequestration by the Host-Defense Protein Calprotectin. *Annu. Rev. Biochem.* **2018,** *87*, 621-643.

9. Korndorfer, I. P.; Brueckner, F.; Skerra, A., The crystal structure of the human (S100A8/S100A9)2 heterotetramer, calprotectin, illustrates how conformational changes of interacting alpha-helices can determine specific association of two EF-hand proteins. *J. Mol. Biol.* **2007,** *370* (5), 887-98.
10. Brophy, M. B.; Hayden, J. A.; Nolan, E. M., Calcium ion gradients modulate the zinc affinity and antibacterial activity of human calprotectin. *J. Am. Chem. Soc.* **2012,** *134* (43), 18089-100.
11. Gagnon, D. M.; Brophy, M. B.; Bowman, S. E.; Stich, T. A.; Drennan, C. L.; Britt, R. D.; Nolan, E. M., Manganese binding properties of human calprotectin under conditions of high and low calcium: X-ray crystallographic and advanced electron paramagnetic resonance spectroscopic analysis. *J. Am. Chem. Soc.* **2015,** *137* (8), 3004-16.
12. Nakashige, T. G.; Stephan, J. R.; Cunden, L. S.; Brophy, M. B.; Wommack, A. J.; Keegan, B. C.; Shearer, J. M.; Nolan, E. M., The Hexahistidine Motif of Host-Defense Protein Human Calprotectin Contributes to Zinc Withholding and Its Functional Versatility. *J. Am. Chem. Soc.* **2016,** *138* (37), 12243-51.
13. Nakashige, T. G.; Zhang, B.; Krebs, C.; Nolan, E. M., Human calprotectin is an iron-sequestering host-defense protein. *Nat. Chem. Biol.* **2015,** *11* (10), 765-71.
14. Nakashige, T. G.; Zygiel, E. M.; Drennan, C. L.; Nolan, E. M., Nickel Sequestration by the Host-Defense Protein Human Calprotectin. *J. Am. Chem. Soc.* **2017,** *139* (26), 8828-8836.
15. Hayden, J. A.; Brophy, M. B.; Cunden, L. S.; Nolan, E. M., High-affinity manganese coordination by human calprotectin is calcium-dependent and requires the histidine-rich site formed at the dimer interface. *J. Am. Chem. Soc.* **2013,** *135* (2), 775-87.
16. Finney, L. A.; O'Halloran, T. V., Transition metal speciation in the cell: insights from the chemistry of metal ion receptors. *Science* **2003,** *300* (5621), 931-6.
17. Ziller, A.; Fraissinet-Tachet, L., Metallothionein diversity and distribution in the tree of life: a multifunctional protein. *Metallomics* **2018,** *10* (11), 1549-1559.
18. Sutherland, D. E.; Stillman, M. J., Challenging conventional wisdom: single domain metallothioneins. *Metallomics* **2014,** *6* (4), 702-28.
19. Rigby Duncan, K. E.; Stillman, M. J., Metal-dependent protein folding: metallation of metallothionein. *J. Inorg. Biochem.* **2006,** *100* (12), 2101-7.
20. Chan, J., Studies of metal binding reactions in metallothioneins by spectroscopic, molecular biology, and molecular modeling techniques. *Coord. Chem. Rev.* **2002,** *233-234*, 319-339.
21. Jiang, L. J.; Maret, W.; Vallee, B. L., The glutathione redox couple modulates zinc transfer from metallothionein to zinc-depleted sorbitol dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.*. **1998,** *95* (7), 3483-8.
22. Bell, S. G.; Vallee, B. L., The metallothionein/thionein system: an oxidoreductive metabolic zinc link. *Chembiochem* **2009,** *10* (1), 55-62.
23. Giles, N. M.; Watts, A. B.; Giles, G. I.; Fry, F. H.; Littlechild, J. A.; Jacob, C., Metal and Redox Modulation of Cysteine Protein Function. *Chem. Biol.* **2003,** *10* (8), 677-693.
24. Maret, W., Zinc coordination environments in proteins determine zinc functions. *J. Trace Elem. Med. Biol.* **2005,** *19* (1), 7-12.
25. Subramanian Vignesh, K.; Deepe, G. S., Jr., Metallothioneins: Emerging Modulators in Immunity and Infection. *Int. J. Mol. Sci.* **2017,** *18* (10).
26. Sontag, E., Protein phosphatase 2A: the Trojan Horse of cellular signaling. *Cell. Signal.* **2001,** *13* (1), 7-16.

27. Lindskog, S., Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **1997,** *74* (1), 1-20.
28. Wilson, D. B., Cellulases and biofuels. *Curr. Opin. Biotechnol.* **2009,** *20* (3), 295-9.
29. Krebs, J. F.; Ippolito, J. A.; Christianson, D. W.; Fierke, C. A., Structural and functional importance of a conserved hydrogen bond network in human carbonic anhydrase II. *J. Biol. Chem.* **1993,** *268* (36), 27458-27466.
30. Kiefer, L. L.; Paterno, S. A.; Fierke, C. A., Hydrogen bond network in the metal binding site of carbonic anhydrase enhances zinc affinity and catalytic efficiency. *J. Am. Chem. Soc.* **2002,** *117* (26), 6831-6837.
31. Maret, W.; Li, Y., Coordination dynamics of zinc in proteins. *Chem. Rev.* **2009,** *109* (10), 4682-707.
32. Kobayashi, M.; Shimizu, S., Metalloenzyme nitrile hydratase: structure, regulation, and application to biotechnology. *Nat. Biotechnol.* **1998,** *16* (8), 733-6.
33. Harrop, T. C.; Mascharak, P. K., Fe(III) and Co(III) centers with carboxamido nitrogen and modified sulfur coordination: lessons learned from nitrile hydratase. *Acc. Chem. Res.* **2004,** *37* (4), 253-60.
34. Meldrum, N. U.; Roughton, F. J., Carbonic anhydrase. Its preparation and properties. *J. Physiol.* **1933,** *80* (2), 113-42.
35. Keilin, D.; Mann, T., Carbonic Anhydrase. *Nature* **1939,** *144* (3644), 442-443.
36. Sly, W. S.; Hu, P. Y., Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu. Rev. Biochem.* **1995,** *64*, 375-401.
37. Prasad, S.; Bhalla, T. C., Nitrile hydratases (NHases): at the interface of academia and industry. *Biotechnol. Adv.* **2010,** *28* (6), 725-41.
38. Huang, W.; Jia, J.; Cummings, J.; Nelson, M.; Schneider, G.; Lindqvist, Y., Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold. *Structure* **1997,** *5* (5), 691-699.
39. Hopmann, K. H., Full reaction mechanism of nitrile hydratase: a cyclic intermediate and an unexpected disulfide switch. *Inorg. Chem.* **2014,** *53* (6), 2760-2.
40. Murakami, T.; Nojiri, M.; Nakayama, H.; Odaka, M.; Yohda, M.; Dohmae, N.; Takio, K.; Nagamune, T.; Endo, I., Post-translational modification is essential for catalytic activity of nitrile hydratase. *Protein Sci.* **2000,** *9* (5), 1024-30.
41. Light, K. M.; Yamanaka, Y.; Odaka, M.; Solomon, E. I., Spectroscopic and Computational Studies of Nitrile Hydratase: Insights into Geometric and Electronic Structure and the Mechanism of Amide Synthesis. *Chem. Sci.* **2015,** *6* (11), 6280-6294.
42. Poulos, T. L., Cytochrome P450. *Curr. Opin. Struct. Biol.* **1995,** *5* (6), 767-774.
43. Joseph, C. A.; Maroney, M. J., Cysteine dioxygenase: structure and mechanism. *Chem. Commun. (Camb)* **2007,** (32), 3338-49.
44. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I., Structure and chemistry of cytochrome P450. *Chem. Rev.* **2005,** *105* (6), 2253-77.
45. Rittle, J.; Green, M. T., Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science* **2010,** *330* (6006), 933-7.
46. Meunier, B.; de Visser, S. P.; Shaik, S., Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. *Chem. Rev.* **2004,** *104* (9), 3947-80.
47. Lewis, D. F. V.; Eddershaw, P. J.; Dickins, M.; Tarbit, M. H.; Goldfarb, P. S., Structural determinants of cytochrome P450 substrate specificity, binding affinity and catalytic rate. *Chem. Biol. Interact.* **1998,** *115* (3), 175-199.

48. Gray, H. B.; Winkler, J. R., Electron transfer in proteins. *Annu. Rev. Biochem.* **1996,** *65*, 537-61.
49. Bella, D. L.; Hirschberger, L. L.; Hosokawa, Y.; Stipanuk, M. H., Mechanisms involved in the regulation of key enzymes of cysteine metabolism in rat liver in vivo. *Am. J. Physiol.* **1999,** *276* (2), E326-35.
50. Lee, J. I.; Londono, M.; Hirschberger, L. L.; Stipanuk, M. H., Regulation of cysteine dioxygenase and gamma-glutamylcysteine synthetase is associated with hepatic cysteine level. *J. Nutr. Biochem.* **2004,** *15* (2), 112-22.
51. Stipanuk, M. H.; Ueki, I.; Dominy, J. E., Jr.; Simmons, C. R.; Hirschberger, L. L., Cysteine dioxygenase: a robust system for regulation of cellular cysteine levels. *Amino Acids* **2009,** *37* (1), 55-63.
52. Kumar, D.; Thiel, W.; de Visser, S. P., Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. *J. Am. Chem. Soc.* **2011,** *133* (11), 3869-82.
53. Li, W.; Blaesi, E. J.; Pecore, M. D.; Crowell, J. K.; Pierce, B. S., Second-sphere interactions between the C93-Y157 cross-link and the substrate-bound Fe site influence the O(2) coupling efficiency in mouse cysteine dioxygenase. *Biochemistry* **2013,** *52* (51), 9104-19.
54. Davies, C. G.; Fellner, M.; Tchesnokov, E. P.; Wilbanks, S. M.; Jameson, G. N., The Cys-Tyr cross-link of cysteine dioxygenase changes the optimal pH of the reaction without a structural change. *Biochemistry* **2014,** *53* (50), 7961-8.
55. Martinie, R. J.; Godakumbura, P. I.; Porter, E. G.; Divakaran, A.; Burkhart, B. J.; Wertz, J. T.; Benson, D. E., Identifying proteins that can form tyrosine-cysteine crosslinks. *Metallomics* **2012,** *4* (10), 1037-42, 1008.
56. Yang, Y.; Arnold, F. H., Navigating the Unnatural Reaction Space: Directed Evolution of Heme Proteins for Selective Carbene and Nitrene Transfer. *Acc. Chem. Res.* **2021,** *54* (5), 1209-1225.
57. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H., Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **2013,** *339* (6117), 307-10.
58. McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold, F. H., Enantioselective intramolecular C-H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo. *Angew. Chem. Int. Ed. Engl.* **2013,** *52* (35), 9309-12.
59. Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H., Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015,** *1* (2), 89-93.
60. Jung, S. T.; Lauchli, R.; Arnold, F. H., Cytochrome P450: taming a wild type enzyme. *Curr. Opin. Biotechnol.* **2011,** *22* (6), 809-17.
61. Chen, K.; Zhang, S. Q.; Brandenberg, O. F.; Hong, X.; Arnold, F. H., Alternate Heme Ligation Steers Activity and Selectivity in Engineered Cytochrome P450-Catalyzed Carbene-Transfer Reactions. *J. Am. Chem. Soc.* **2018,** *140* (48), 16402-16407.
62. Hyster, T. K.; Arnold, F. H., P450BM3-Axial Mutations: A Gateway to Non-Natural Reactivity. *Isr. J. Chem.* **2015,** *55* (1), 14-20.
63. Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M., A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **2013,** *9* (8), 485-7.

64. Brustad, E. M.; Arnold, F. H., Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **2011,** *15* (2), 201-10.
65. Kan, S. B.; Lewis, R. D.; Chen, K.; Arnold, F. H., Directed evolution of cytochrome c for carbon-silicon bond formation: Bringing silicon to life. *Science* **2016,** *354* (6315), 1048-1051.
66. Berry, S. M.; Gieselman, M. D.; Nilges, M. J.; van Der Donk, W. A.; Lu, Y., An engineered azurin variant containing a selenocysteine copper ligand. *J. Am. Chem. Soc.* **2002,** *124* (10), 2084-5.
67. Marshall, N. M.; Garner, D. K.; Wilson, T. D.; Gao, Y. G.; Robinson, H.; Nilges, M. J.; Lu, Y., Rationally tuning the reduction potential of a single cupredoxin beyond the natural range. *Nature* **2009,** *462* (7269), 113-6.
68. Clark, K. M.; Yu, Y.; Marshall, N. M.; Sieracki, N. A.; Nilges, M. J.; Blackburn, N. J.; van der Donk, W. A.; Lu, Y., Transforming a blue copper into a red copper protein: engineering cysteine and homocysteine into the axial position of azurin using site-directed mutagenesis and expressed protein ligation. *J. Am. Chem. Soc.* **2010,** *132* (29), 10093-101.
69. Mirts, E. N.; Bhagi-Damodaran, A.; Lu, Y., Understanding and Modulating Metalloenzymes with Unnatural Amino Acids, Non-Native Metal Ions, and Non-Native Metallocofactors. *Acc. Chem. Res.* **2019,** *52* (4), 935-944.
70. Ji, P.; Park, J.; Gu, Y.; Clark, D. S.; Hartwig, J. F., Abiotic reduction of ketones with silanes catalysed by carbonic anhydrase through an enzymatic zinc hydride. *Nat. Chem.* **2021,** *13* (4), 312-318.
71. Davis, H. J.; Ward, T. R., Artificial Metalloenzymes: Challenges and Opportunities. *ACS Cent. Sci.* **2019,** *5* (7), 1120-1136.
72. Pordea, A.; Creus, M.; Panek, J.; Duboc, C.; Mathis, D.; Novic, M.; Ward, T. R., Artificial metalloenzyme for enantioselective sulfoxidation based on vanadyl-loaded streptavidin. *J. Am. Chem. Soc.* **2008,** *130* (25), 8085-8.
73. Monnard, F. W.; Heinisch, T.; Nogueira, E. S.; Schirmer, T.; Ward, T. R., Human carbonic anhydrase II as a host for piano-stool complexes bearing a sulfonamide anchor. *Chem. Commun.* **2011,** *47* (29), 8238-40.
74. Heinisch, T.; Ward, T. R., Artificial Metalloenzymes Based on the Biotin-Streptavidin Technology: Challenges and Opportunities. *Acc. Chem. Res.* **2016,** *49* (9), 1711-21.
75. Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R., Directed evolution of artificial metalloenzymes for in vivo metathesis. *Nature* **2016,** *537* (7622), 661-665.
76. Liang, A. D.; Serrano-Plana, J.; Peterson, R. L.; Ward, T. R., Artificial Metalloenzymes Based on the Biotin-Streptavidin Technology: Enzymatic Cascades and Directed Evolution. *Acc. Chem. Res.* **2019,** *52* (3), 585-595.
77. Grimm, A. R.; Sauer, D. F.; Davari, M. D.; Zhu, L.; Bocola, M.; Kato, S.; Onoda, A.; Hayashi, T.; Okuda, J.; Schwaneberg, U., Cavity Size Engineering of a β-Barrel Protein Generates Efficient Biohybrid Catalysts for Olefin Metathesis. *ACS Catal.* **2018,** *8* (4), 3358-3364.
78. Yang, H.; Srivastava, P.; Zhang, C.; Lewis, J. C., A general method for artificial metalloenzyme formation through strain-promoted azide-alkyne cycloaddition. *Chembiochem* **2014,** *15* (2), 223-7.
79. Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C., Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat. Commun.* **2015,** *6*, 7789.

80. DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A., De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **1999,** *68*, 779-819.
81. Lombardi, A.; Pirro, F.; Maglio, O.; Chino, M.; DeGrado, W. F., De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. *Acc. Chem. Res.* **2019,** *52* (5), 1148-1159.
82. Tebo, A. G.; Pecoraro, V. L., Artificial metalloenzymes derived from three-helix bundles. *Curr. Opin. Chem. Biol.* **2015,** *25*, 65-70.
83. Chalkley, M. J.; Mann, S. I.; DeGrado, W. F., De novo metalloprotein design. *Nat. Rev. Chem.* **2021,** *6* (1), 31-50.
84. Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F., Retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97* (12), 6298-305.
85. Faiella, M.; Andreozzi, C.; de Rosales, R. T.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A., An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **2009,** *5* (12), 882-4.
86. Ulas, G.; Lemmin, T.; Wu, Y.; Gassner, G. T.; DeGrado, W. F., Designed metalloprotein stabilizes a semiquinone radical. *Nat. Chem.* **2016,** *8* (4), 354-9.
87. Zhang, S. Q.; Chino, M.; Liu, L.; Tang, Y.; Hu, X.; DeGrado, W. F.; Lombardi, A., De Novo Design of Tetranuclear Transition Metal Clusters Stabilized by Hydrogen-Bonded Networks in Helical Bundles. *J. Am. Chem. Soc.* **2018,** *140* (4), 1294-1304.
88. Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; DeGrado, W. F., Alteration of the oxygen-dependent reactivity of de novo Due Ferri proteins. *Nat. Chem.* **2012,** *4* (11), 900-6.
89. Farrer, B. T.; McClure, C. P.; Penner-Hahn, J. E.; Pecoraro, V. L., Arsenic(III)-cysteine interactions stabilize three-helix bundles in aqueous solution. *Inorg. Chem.* **2000,** *39* (24), 5422-3.
90. Matzapetakis, M.; Farrer, B. T.; Weng, T. C.; Hemmingsen, L.; Penner-Hahn, J. E.; Pecoraro, V. L., Comparison of the binding of cadmium(II), mercury(II), and arsenic(III) to the de novo designed peptides TRI L12C and TRI L16C. *J. Am. Chem. Soc.* **2002,** *124* (27), 8042-54.
91. Zastrow, M. L.; Peacock, A. F.; Stuckey, J. A.; Pecoraro, V. L., Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **2011,** *4* (2), 118-23.
92. Tegoni, M.; Yu, F.; Bersellini, M.; Penner-Hahn, J. E.; Pecoraro, V. L., Designing a functional type 2 copper center that has nitrite reductase activity within alpha-helical coiled coils. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (52), 21234-9.
93. Tolbert, A. E.; Ervin, C. S.; Ruckthong, L.; Paul, T. J.; Jayasinghe-Arachchige, V. M.; Neupane, K. P.; Stuckey, J. A.; Prabhakar, R.; Pecoraro, V. L., Heteromeric three-stranded coiled coils designed using a Pb(II)(Cys)3 template mediated strategy. *Nat. Chem.* **2020,** *12* (4), 405-411.
94. Kennedy, M., Metalloprotein and redox protein design. *Curr. Opin. in Struct. Biol.* **2001,** *11* (4), 485-490.
95. Zhu, J.; Avakyan, N.; Kakkis, A.; Hoffnagle, A. M.; Han, K.; Li, Y.; Zhang, Z.; Choi, T. S.; Na, Y.; Yu, C. J.; Tezcan, F. A., Protein Assembly by Design. *Chem. Rev.* **2021,** *121* (22), 13701-13796.
96. Wang, W.; Knovich, M. A.; Coffman, L. G.; Torti, F. M.; Torti, S. V., Serum ferritin: Past, present and future. *Biochim. Biophys. Acta* **2010,** *1800* (8), 760-9.

97. Levy, E. D.; Teichmann, S., Structural, evolutionary, and assembly principles of protein oligomerization. *Prog. Mol. Biol. Transl. Sci.* **2013,** *117*, 25-51.
98. Varughese, K. I.; Madhusudan; Zhou, X. Z.; Whiteley, J. M.; Hoch, J. A., Formation of a Novel Four-Helix Bundle and Molecular Recognition Sites by Dimerization of a Response Regulator Phosphotransferase. *Mol. Cell* **1998,** *2* (4), 485-493.
99. Saraste, M., Structural features of cytochrome oxidase. *Q. Rev. Biophys.* **1990,** *23* (4), 331-66.
100. Zhang, Y.; Orner, B. P., Self-assembly in the ferritin nano-cage protein superfamily. *Int. J. Mol. Sci.* **2011,** *12* (8), 5406-21.
101. Nelson, N.; Ben-Shem, A., The complex architecture of oxygenic photosynthesis. *Nat. Rev. Mol. Cell. Biol.* **2004,** *5* (12), 971-82.
102. Dobson, C. M., Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884-90.
103. Dobson, C. M.; Sali, A.; Karplus, M., Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem. Int. Ed. Engl.* **1998,** *37* (7), 868-893.
104. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.
105. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583-589.
106. Keskin, O.; Tuncbag, N.; Gursoy, A., Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem. Rev.* **2016,** *116* (8), 4884-909.
107. Vakser, I. A., Protein-protein docking: from interaction to interactome. *Biophys. J.* **2014,** *107* (8), 1785-1793.
108. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A., Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89* (6), 2195-9.
109. Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L., Jr.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J., The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017,** *13* (6), 3031-3048.
110. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y. E.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popovic, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P., ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth. Enzymol.* **2011,** *487*, 545-74.
111. Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D., Principles for designing ideal protein structures. *Nature* **2012,** *491* (7423), 222-7.

112. King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D., Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **2012,** *336* (6085), 1171-4.
113. Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. A.; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D., A de novo protein binding pair by computational design and directed evolution. *Mol. Cell* **2011,** *42* (2), 250-60.
114. Chen, Z.; Kibler, R. D.; Hunt, A.; Busch, F.; Pearl, J.; Jia, M.; VanAernum, Z. L.; Wicky, B. I. M.; Dods, G.; Liao, H.; Wilken, M. S.; Ciarlo, C.; Green, S.; El-Samad, H.; Stamatoyannopoulos, J.; Wysocki, V. H.; Jewett, M. C.; Boyken, S. E.; Baker, D., De novo design of protein logic gates. *Science* **2020,** *368* (6486), 78-84.
115. Chen, Z.; Boyken, S. E.; Jia, M.; Busch, F.; Flores-Solis, D.; Bick, M. J.; Lu, P.; VanAernum, Z. L.; Sahasrabuddhe, A.; Langan, R. A.; Bermeo, S.; Brunette, T. J.; Mulligan, V. K.; Carter, L. P.; DiMaio, F.; Sgourakis, N. G.; Wysocki, V. H.; Baker, D., Programmable design of orthogonal protein heterodimers. *Nature* **2019,** *565* (7737), 106-111.
116. Boyken, S. E.; Benhaim, M. A.; Busch, F.; Jia, M.; Bick, M. J.; Choi, H.; Klima, J. C.; Chen, Z.; Walkey, C.; Mileant, A.; Sahasrabuddhe, A.; Wei, K. Y.; Hodge, E. A.; Byron, S.; Quijano-Rubio, A.; Sankaran, B.; King, N. P.; Lippincott-Schwartz, J.; Wysocki, V. H.; Lee, K. K.; Baker, D., De novo design of tunable, pH-driven conformational changes. *Science* **2019,** *364* (6441), 658-664.
117. Salgado, E. N.; Radford, R. J.; Tezcan, F. A., Metal-directed protein self-assembly. *Acc. Chem. Res.* **2010,** *43* (5), 661-72.
118. Fiedler, D.; Pagliero, D.; Brumaghim, J. L.; Bergman, R. G.; Raymond, K. N., Encapsulation of cationic ruthenium complexes into a chiral self-assembled cage. *Inorg. Chem.* **2004,** *43* (3), 846-8.
119. Geue, R. J.; Hambley, T. W.; Harrowfield, J. M.; Sargeson, A. M.; Snow, M. R., Metal ion encapsulation: cobalt cages derived from polyamines, formaldehyde, and nitromethane. *J. Am. Chem. Soc.* **2002,** *106* (19), 5478-5488.
120. Zhou, H. C.; Kitagawa, S., Metal-organic frameworks (MOFs). *Chem. Soc. Rev.* **2014,** *43* (16), 5415-8.
121. Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A., Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* **2007,** *129* (44), 13374-5.
122. Salgado, E. N.; Lewis, R. A.; Mossin, S.; Rheingold, A. L.; Tezcan, F. A., Control of protein oligomerization symmetry by metal coordination: C2 and C3 symmetrical assemblies through Cu(II) and Ni(II) coordination. *Inorg. Chem.* **2009,** *48* (7), 2726-8.
123. Trofimenko, S., Recent advances in poly(pyrazolyl)borate (scorpionate) chemistry. *Chem. Rev.* **2002,** *93* (3), 943-980.
124. Ward, M. D.; McCleverty, J. A.; Jeffery, J. C., Coordination and supramolecular chemistry of multinucleating ligands containing two or more pyrazolyl-pyridine 'arms'. *Coord. Chem. Rev.* **2001,** *222* (1), 251-272.
125. Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A., Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (5), 1827-32.

126. Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A., Evolution of metal selectivity in templated protein interfaces. *J. Am. Chem. Soc.* **2010,** *132* (25), 8610-7.
127. Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F. A., Metal-Templated Design of Chemically Switchable Protein Assemblies with High-Affinity Coordination Sites. *Angew. Chem. Int. Ed. Engl.* **2020,** *59* (49), 21940-21944.
128. Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A., In vitro and cellular self-assembly of a Zn-binding protein cryptand via templated disulfide bonds. *J. Am. Chem. Soc.* **2013,** *135* (32), 12013-22.
129. Churchfield, L. A.; Alberstein, R. G.; Williamson, L. M.; Tezcan, F. A., Determining the Structural and Energetic Basis of Allostery in a De Novo Designed Metalloprotein Assembly. *J. Am. Chem. Soc.* **2018,** *140* (31), 10043-10053.
130. Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A., De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J. Am. Chem. Soc.* **2016,** *138* (40), 13163-13166.
131. Song, W. J.; Tezcan, F. A., A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014,** *346* (6216), 1525-8.

Chapter 2: Metal-templated design of chemically switchable protein assemblies with high-affinity coordination sites

2.1 Abstract

To mimic a hypothetical pathway for protein evolution, we previously developed a design strategy (Metal-Templated Interface Redesign), in which a monomeric protein (cytochrome *cb*$_{562}$) was tailored for metal-mediated self-assembly, followed by the re-design of the resulting oligomers for enhanced stability and metal-based functions. Here we show that a single hydrophobic mutation on the cytochrome *cb*$_{562}$ surface can drastically alter the outcome of metal-directed oligomerization to yield a new trimeric architecture, (TriCyt1)$_3$, featuring an unusual hexa-histidine coordination motif. Through computational and rational redesign, this nascent trimer is converted into second and third-generation variants (TriCyt2)$_3$ and (TriCyt3)$_3$ with increased structural stability and preorganization for metal coordination. The three TriCyt variants combined furnish a unique design platform to a) provide tunable coupling between protein quaternary structure and metal coordination, b) enable the construction of metal/pH-switchable protein oligomerization motifs, and c) generate a robust metal coordination site that can accommodate all mid-to-late first-row transition metal ions with high affinity, including Mn(II) with nanomolar dissociation constants, rivaling those of the strongest Mn(II)-binding protein, calprotectin.

2.2 Introduction

Metalloproteins perform countless biological functions despite the fact that they co-opt barely more than a handful of transition metal ions.[1-3] Underlying this functional diversity is a complex interplay between metal coordination/reactivity and protein structure/dynamics.[4-6] Although the metal-protein interplay can often be understood through detailed, top-down studies

of natural metalloproteins, it remains considerably more challenging to build this interplay from scratch in the form of new metalloproteins. Starting with pioneering studies in the 1990's, there have been notable successes in the de novo design of functional metalloproteins, which are predominantly based on four-helix bundle and α-helical coiled-coiled motifs with readily parametrizable structures.[7-10] On the one hand, the fact that diverse bioinorganic functions can be obtained only with a limited set of structural motifs illustrates the versatility of the de novo design approach. On the other hand, it also highlights the challenge of – and the need for – devising alternative strategies and designing new protein architectures for building bioinorganic complexity in a bottom-up fashion.

It has been hypothesized that some modern metalloproteins may have emerged through the metal-nucleated oligomerization of small peptides or protein domains, followed by the evolution of the resulting assemblies into stable, functional architectures.[11-13] Based on this hypothetical trajectory, we previously developed a protein design strategy termed Metal-Templated Interface Redesign (MeTIR),[14-16] primarily using a monomeric, four-helix bundle protein (cytochrome cb_{562}) as a building block.[17] First, we installed two bis-His motifs (H59/H63, H73/H77) on the Helix3 surface of cyt cb_{562} to enable metal coordination.[18] The resulting construct, MBPC1, assembled into different oligomeric states depending on the coordination preferences of nucleating metal ions (Ni^{II}-trimer; Cu^{II}-dimer; Zn^{II}-tetramer) (**Figure 2.1**).[18-20] Given the extensive protein-protein interfaces in the D_2 symmetric Zn_4:$MBPC1_4$ tetramer, this assembly was chosen as a platform for MeTIR.[14] Zn_4:$MBPC1_4$ was elaborated through rational redesign and directed-evolution to build functional architectures that selectively bound metal ions,[21-23] displayed allostery,[24, 25] and performed catalytic reactions *in vivo*.[26, 27] Yet, despite the functional versatility of the Zn_4:$MBPC1_4$ progeny, they are inherently biased by the metal-templating strategy toward

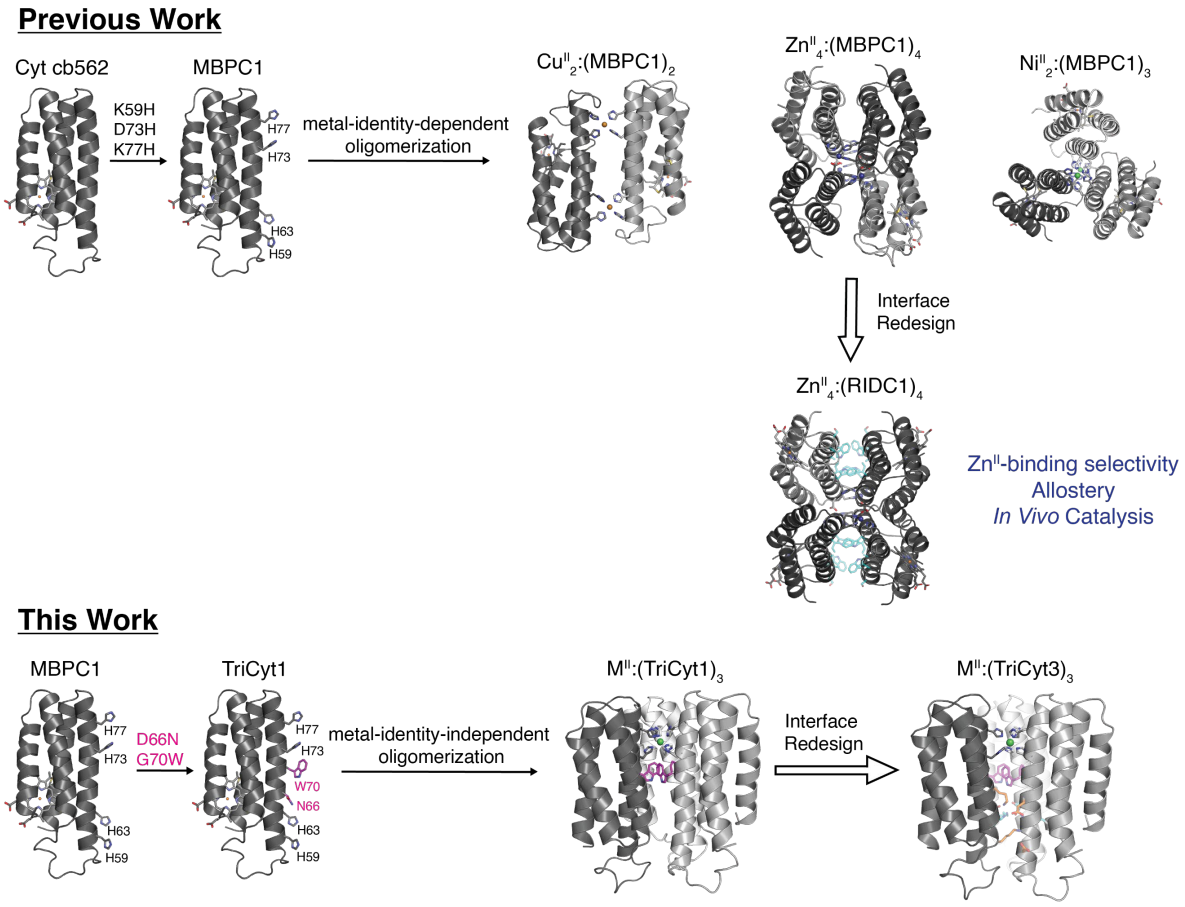

Figure 2.1 | Previous (top) and current (bottom) applications of MeTIR using cytochrome *cb*$_{562}$ as a building block. In the previous route, the K59H/D73H/K77H mutant of cytochrome *cb*$_{562}$ (MBPC1), assembled into a variety of oligomeric states upon the addition of Ni^{II}, Cu^{II}, or Zn^{II}, with the oligomerization state dictated by the geometric coordination preferences of the metal ion. The Zn^{II}-dependent tetramer Zn^{II}_4:(MBPC1)$_4$ was subsequently used as a structural template for computational redesign to generate the variant RIDC1. The Zn^{II}-dependent tetramer of RIDC1, Zn^{II}_4:(RIDC1)$_4$, served as a diversifiable scaffold to obtain Zn^{II} metalloproteins with a wide range of functions. The MeTIR strategy described in this study is based on the D66N/G70W variant of MBPC1 (TriCyt1), which oligomerizes into a trimer regardless of the metal identity, yielding a highly preorganized scaffold for the design of functional metalloproteins using a large variety of metals.

Zn^{II} coordination chemistry. Moreover, because of their D_2 symmetry, they possess at least four copies of each metal center of interest, complicating the examination and modification of the individual metal centers.

Here we aimed to direct the self-assembly of MBPC1 toward more pre-organized architectures in lower oligomerization states, possibly with fewer metal centers. It is well-established that surface-exposed hydrophobic residues can effectively induce protein-protein interactions and aggregation.[28, 29] Thus, we incorporated a Trp residue (W70) onto the MBPC1 Helix3 surface between the H59/H63 and H73/H77 motifs to enable the formation of a hydrophobic core upon metal-mediated oligomerization (**Figure 2.2a and Figure 2.1**). We also mutated the negatively charged Asp66 sidechain on the same surface to Asn to avoid repulsive electrostatic interactions during self-assembly.

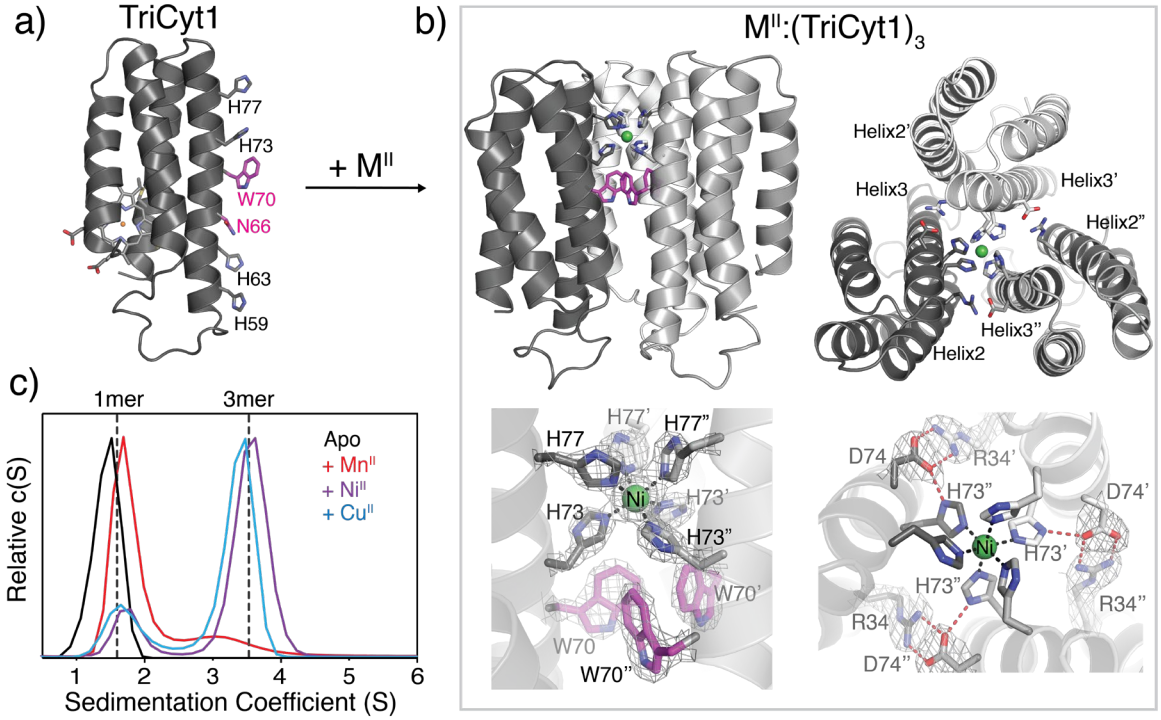

Figure 2.2 | Solution and crystallographic characterization of TriCyt1. (a) Structural model of TriCyt1 monomer. (b) Crystal structure of the Ni^{II}:$(TriCyt1)_3$ trimer (PDB ID: 6WZA), which is essentially identical to that of Cu^{II}:$(TriCyt1)_3$ (PDB ID: 6X8X). The upper panels show the side- and top-views of the trimer and the bottom panels depict the primary and secondary coordination environments that comprise the "core motif". The 2*Fo*–*Fc* maps (grey mesh) are contoured at 1σ. (c) Sedimentation velocity (SV) profiles of TriCyt1 (30 μM monomer) in the absence and presence of 10 μM $MnCl_2$, $NiCl_2$, and $CuCl_2$ (see Figure 2.5 for a complete set)

2.3 Results and Discussion

2.3.1 Assembly of TriCyt1

The metal-dependent assembly of the resulting variant, TriCyt1, was first screened by crystallization in the presence of one equivalent (3 mM) of all mid-to-late first-row transition metal ions (Mn^{II} to Zn^{II}). Regardless of the metal ion identity, we observed crystals with hexagonal morphologies, suggesting that they shared an underlying protein arrangement with three-fold symmetry. We obtained 2.5-Å resolution crystal structures of the Ni^{II}- and Cu^{II}-TriCyt1 complexes (**Figure 2.2b and Figure 2.3**). These structures revealed isostructural trigonal (*P*321) lattices, formed by trimeric TriCyt1 substructures containing a single Ni^{II} or Cu^{II} ion coordinated in a near-octahedral geometry by three pairs of H73/H77 residues. The trimeric substructures feature a close-packed, parallel arrangement of TriCyt1 monomers. Near the center is a hub of T-stacked W70 sidechains that non-covalently buttress the metal-coordinating H73 residues (**Figure 2.2b and Figure 2.4**). Additionally, there are three pairs of intermonomer, salt-bridging interactions

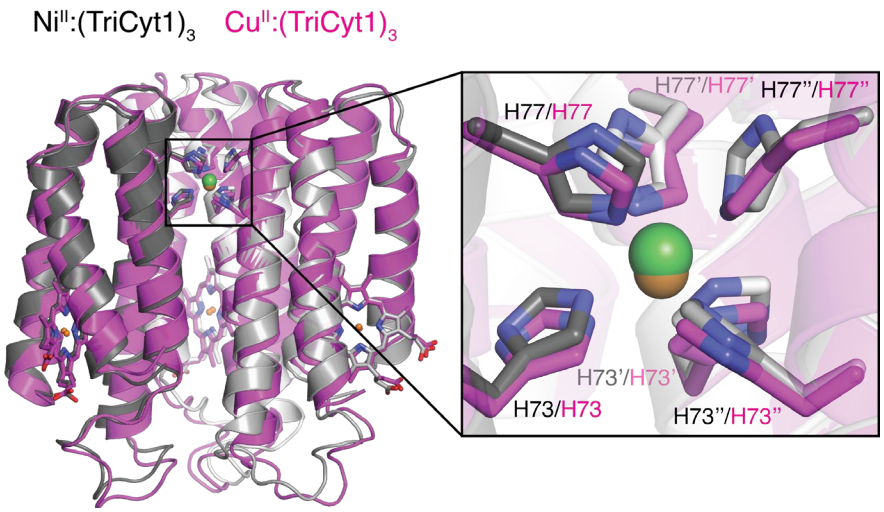

Figure 2.3 | Superposition of Ni^{II}:(TriCyt1)$_3$ (PDB ID: 6WZA) and Cu^{II}:(TriCyt1)$_3$ (PDB ID: 6X8X) structures. The octahedral, hexacoordinate Cu^{II} binding site is unusual given the tendency of octahedral Cu^{II} complexes to undergo Jahn-Teller distortion toward a tetragonal or square planar geometry

between R34 and D74 residues that surround the metal coordination site and the tris-W70 hub. The D74 carboxylates are further H-bonded to the δ-N's of H73 imidazoles from the same monomer, thus completing an extensive network of interactions surrounding the metal coordination site (**Figure 2.2b**). Interestingly, H59/H63 pairs are not coordinated by metal ions.

Collectively, these structural details suggested that the crystallographically observed trimeric TriCyt1 structure possesses a high degree of preorganization, which allows it to accommodate only a single metal ion within the same His_6 coordination motif independently of the metal identity. Indeed, analytical ultracentrifugation (AUC) and size-exclusion chromatography (SEC) experiments showed that TriCyt1 was monomeric in solution, but exclusively formed trimers upon addition of one equiv. of Mn^{II}, Fe^{II}, Co^{II}, Ni^{II}, Cu^{II} and Zn^{II} (**Figure 2.2c, Figure 2.5, and Figure**

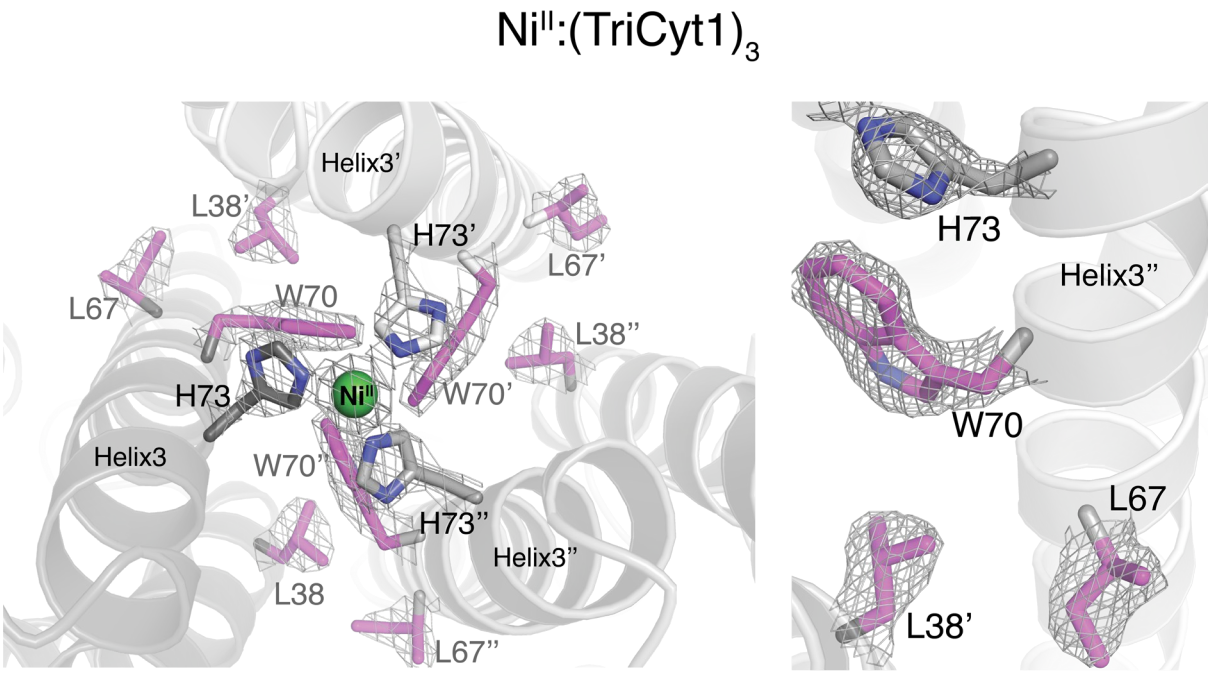

Figure 2.4 | Non-polar residues (purple) positioned at the core interface of Ni^{II}:(TriCyt1)$_3$ (PDB ID: 6WZA). W70 and H73 residues are positioned to form a π-π interaction.

2.6). The yield of trimer formation in solution roughly followed the Irving-Williams (IW) series,[2, 3] ranging from 12% for Mn^{II} to ~95% for Co^{II} (**Figure 2.2c and Figure 2.5**).

2.3.2 Design and assembly of TriCyt2

Metal coordination by a His_6 motif is exceedingly rare in bioinorganic chemistry. The only well-established biological example is found in the immune protein calprotectin,[30, 31] which is involved in the sequestration of metal ions (particularly Mn^{II}) to limit microbial growth and boasts one of the highest Mn^{II} affinities among natural proteins.[32] Given the rarity of the His_6 motif, in addition to the challenge of generating stable Mn coordination sites in proteins, we asked whether the TriCyt1 structure could be subjected to MeTIR to design progressively more stable trimers, which can subsequently bind Mn^{II} (and other divalent ions) with high affinity. The M^{II}:$(TriCyt1)_3$ trimer presents extensive intermonomer interactions (>3000 $Å^2$), which are dominated by a central interface formed by Helices3 of the monomers and peripheral interfaces between neighboring Helices2 and 3 (**Figure 2.2b**). We first undertook a computational redesign of these interfaces, whereby the "core motif" (His_6 site + tris-W70 hub + R34/D74 salt-bridges) and the protein backbone positions were maintained. Through an iterative process involving sidechain and rotamer optimization with Rosetta and visual inspection, we generated the second-generation variant, TriCyt2, which includes six additional surface mutations on TriCyt1 (**Table 2.1**). AUC and SEC experiments showed that TriCyt2 was stable and monomeric in solution but now trimerized in near-quantitative yield upon binding Mn^{II} (as well as the other tested metal ions) (**Figures 2.5- 2.6, Figure 2.7a**). Notably, this represents an 8-fold improvement in Mn^{II}-induced oligomerization.

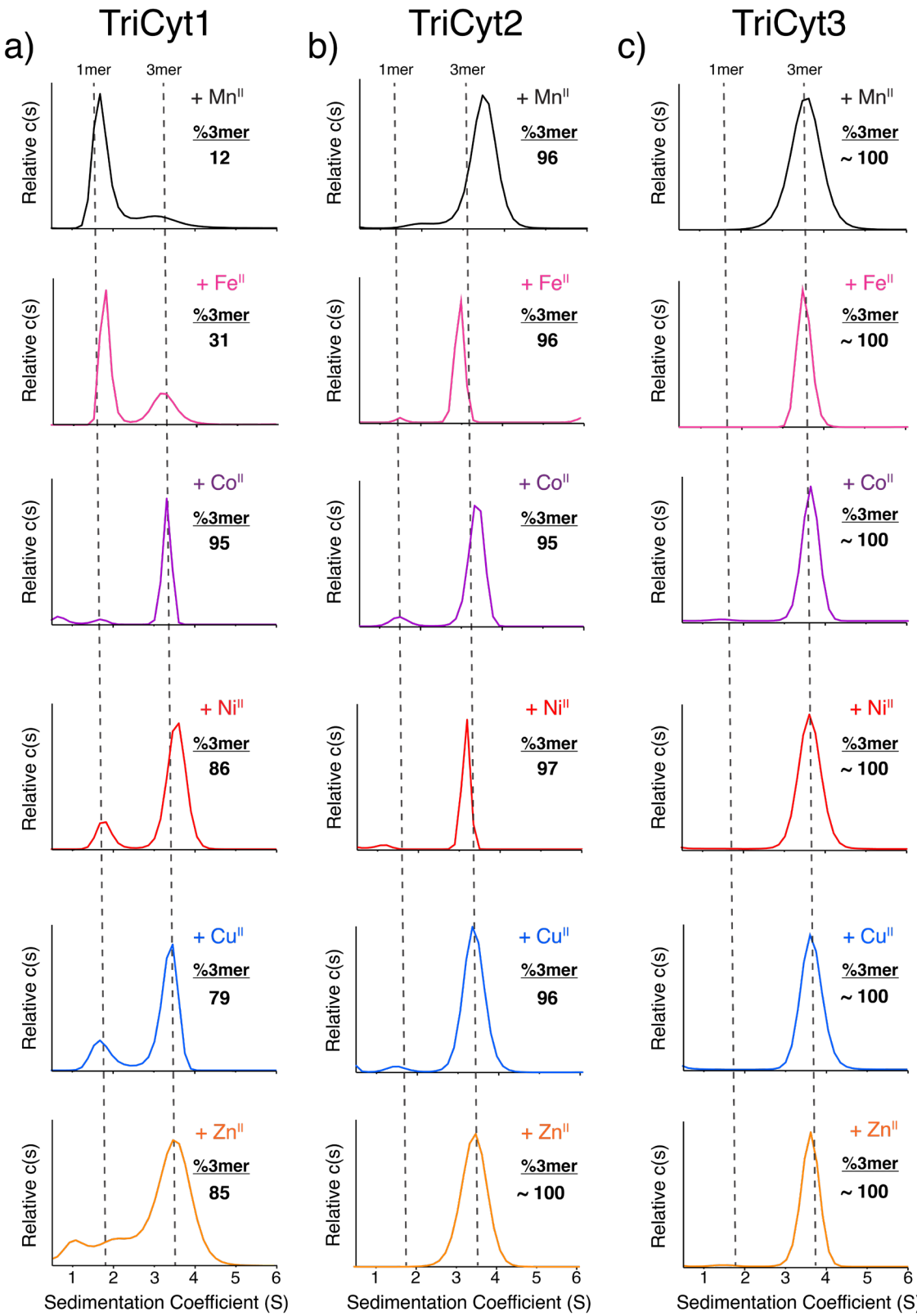

Figure 2.5 | Sedimentation velocity profiles for (a) TriCyt1, (b) TriCyt2, and (c) TriCyt3 in the presence of divalent metal ions.

Table 2.1 | Rosetta-prescribed mutations obtained from interface redesign of TriCyt1 using a rigid backbone protocol.

Residue Position	TriCyt1	Rosetta-prescribed mutations	Mutational Frequency (%)[a]	TriCyt2
34	Arg	Ala Arg	58 29	Arg
38	Leu	Met Leu	58 30	Leu
41	Gln	Lys Ala	96 3	Lys
54	Asp	Ala Ser	96 4	Ala
59	His	Ile Phe	86 5	Ile
63	His	Val Ile	97 3	Val
66	Asn	Asp Leu	96 3	Asp
69	Val	Leu Ala	88 6	Leu

[a] Only two of the highest frequency mutations are shown.

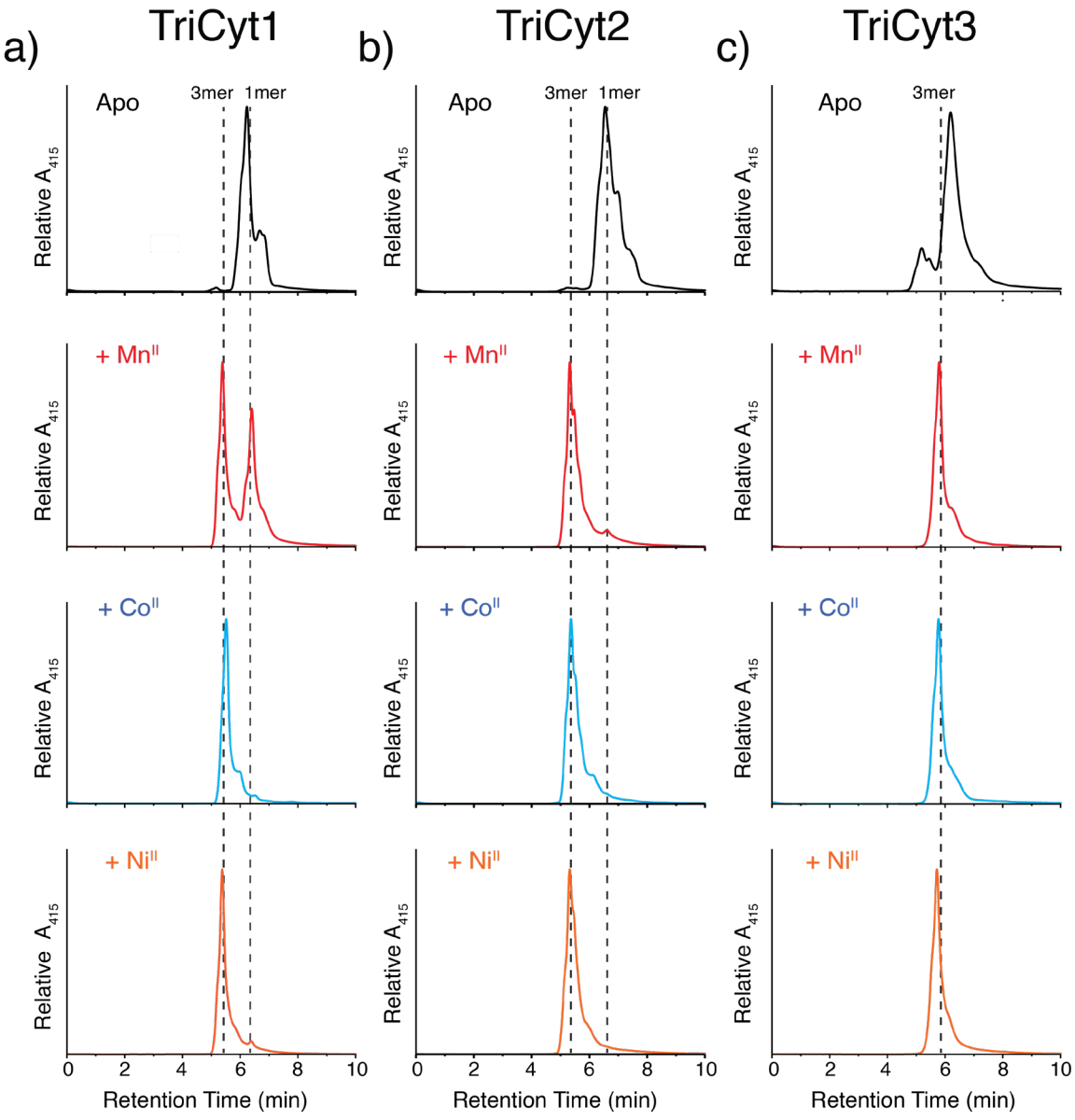

Figure 2.6 | Normalized HPLC chromatograms of (a) TriCyt1, (b) TriCyt2, and (c) TriCyt3 in the presence or absence of divalent metal ions.

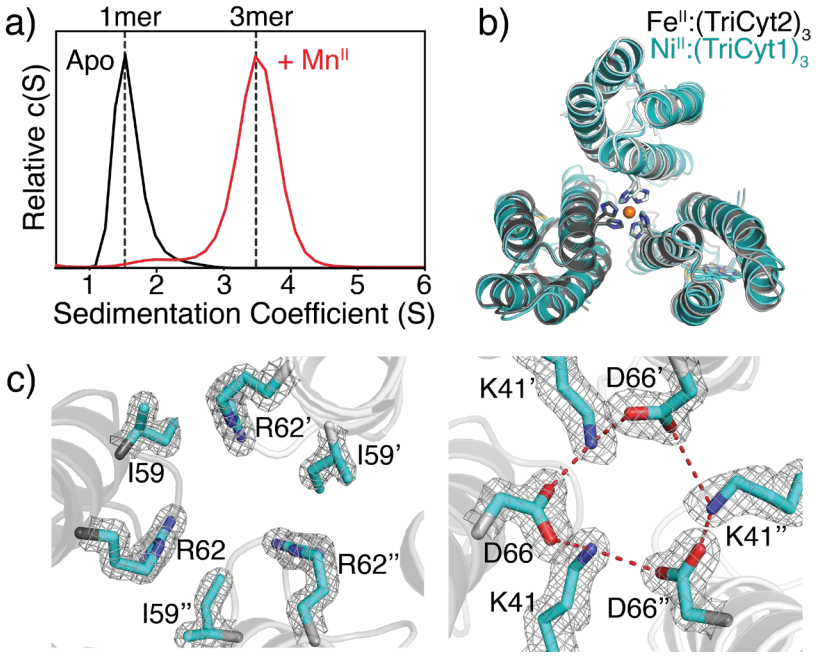

Figure 2.7 | Solution and crystallographic characterization of TriCyt2 (a) SV profiles of TriCyt2 (30 μM monomer) in the absence and presence of 10 μM $MnCl_2$. (b) Structural overlay of Fe^{II}:$(TriCyt2)_3$ (grey, PDB ID: 6WZ0) and Ni^{II}:$(TriCyt1)_3$ (cyan, PDB ID: 6WZA). (c) Hydrophobic packing (left) and H-bonding (right) interactions at the core interface of Fe^{II}:$(TriCyt2)_3$. The 2*F*o–*F*c maps (grey mesh) are contoured at 1σ.

These observations indicate that interface redesign was successful and furnished a unique protein construct that can conditionally assemble into trimers in the presence of all relevant first-row transition metal ions. We determined the 1.7-Å-resolution crystal structure of Fe^{II}-bound TriCyt2 complex, which is nearly isostructural with M^{II}:$(TriCyt1)_3$ complexes (rmsd = 0.86 Å over all 318 α-C's) (**Figure 2.7b**). As designed, the H59I, H63V, V69L mutations contribute to hydrophobic packing in the Helix3 central core while eliminating the non-coordinating H59/H63 motif, whereas the Q41K and N66D substitutions generate a closed network of H-bonding interactions in the same core (**Figure 2.7c**). Additionally, the D54A mutation eliminates the potential repulsive interactions between the Asp54 chains, which causes a slight compaction of the trimer near the 50's loops (**Figure 2.8**). In combination, the six designed mutations yield an

increase in sidechain packing in the TriCyt2 trimer interior (buried surface area or BSA = 3440 Å^2) compared to TriCyt1 (BSA = 3050 Å^2) (**Figure 2.9**), consistent with increased trimer stability. Notably, TriCyt2 forms a trimer in the crystal lattice even in the absence of metal ions, revealing an essentially identical structure (rmsd = 0.37 Å) to the Fe^{II}:$(TriCyt2)_3$ complex that includes a pre-organized His_6 site (**Figure 2.10**).

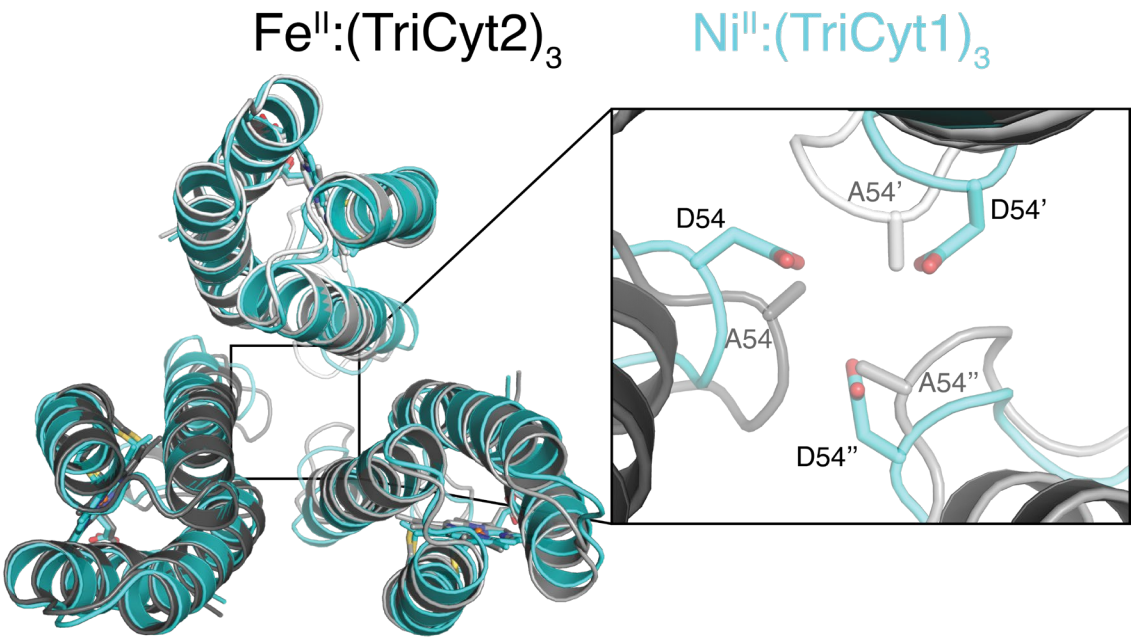

Figure 2.8 | Superposition of Fe^{II}:$(TriCyt2)_3$ (PDB ID: 6WZ0) and Ni^{II}:$(TrCyt1)_3$ (PDB ID: 6WZA) structures. Structural overlays reveal significant conformational changes in the 50's loop region (residues 52-56).

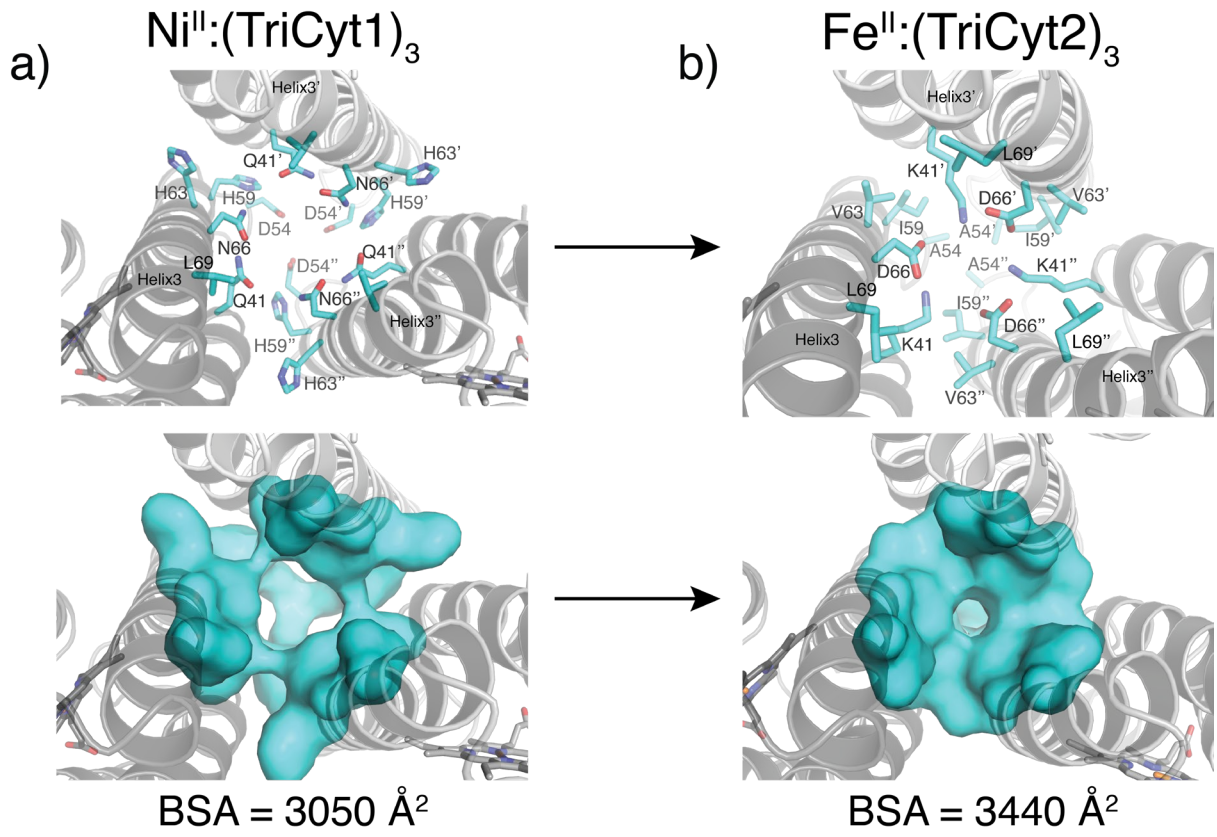

Figure 2.9 | Overview of residues comprising the C_3 interfaces of (a) Ni^{II}:(TriCyt1)$_3$ (PDB ID: 6WZA) and (b) Fe^{II}:(TriCyt2)$_3$ (PDB ID: 6WZ0), shown as sticks (top) or Connolly surfaces (bottom). The buried surface area of the core interface (BSA) increased by 390 Å^2 upon interface redesign.

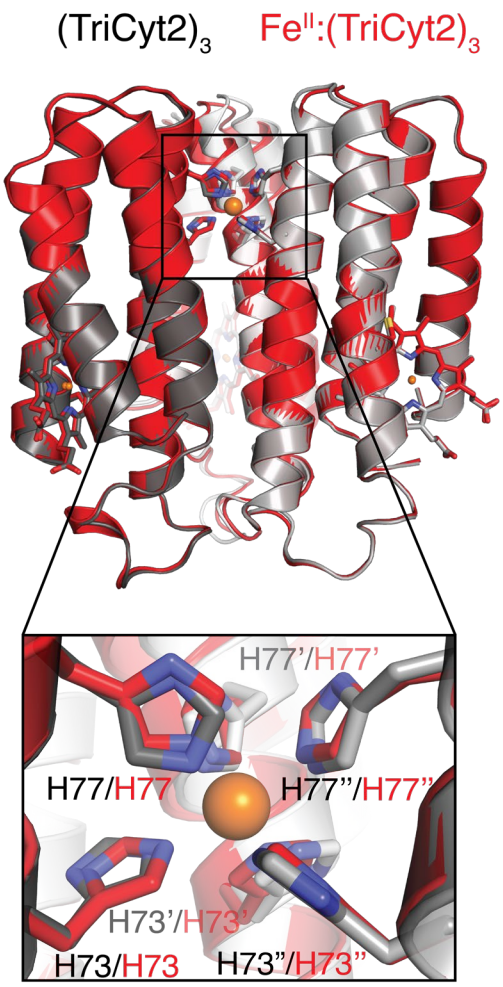

Figure 2.10 | Superposition of metal-free (PDB ID: 6WYU) and Fe^{II}-bound (PDB ID: 6WZ0) $(TriCyt2)_3$ structures.

2.3.3 Design, assembly, and metal-binding characterization of TriCyt3

In the next stage of redesign, we sought to stabilize the peripheral interfaces between Helices2 and 3 from neighboring monomers to obtain a metal-independent trimer in solution. Although the peripheral interfaces are wider and less packed compared to the central interface, they appeared amenable to engineering complementary electrostatic interactions. Accordingly, we incorporated three Lys (T31K, A35K, N80K) and two Glu (I67E, Q71E) residues into TriCyt2 (**Table 2.2**). The resulting third-generation variant, TriCyt3, indeed formed trimers even in the absence of metal ions (**Figure 2.11**). Consistent with their electrostatic stabilization, TriCyt3 trimers reversibly dissociate into monomers upon lowering the solution pH to <4 (likely due to protonation of Glu/Asp residues), even in the presence of tightly binding metal ions such as Cu^{II} (**Figure 2.12**), thus providing a pH-switchable protein assembly platform. We determined the TriCyt3 trimer structure in complex with Mn^{II}, Co^{II}, Ni^{II} and Cu^{II} ions (resolutions ranging from

Table 2.2 | Mutations to convert TriCyt2 into TriCyt3.

Residue Position	TriCyt2	TriCyt3	Purpose
31	Thr	Lys	Salt Bridge
35	Ala	Lys	Salt Bridge
67	Ile	Glu	Salt Bridge
69	Leu	Ala	Steric
71	Gln	Glu	Salt Bridge
76	Leu	Ala	Steric
80	Asn	Lys	Salt Bridge

1.8 to 2.2 Å), which showed little deviation from the TriCyt2 trimers (overall rmsd = 0.37 Å) (**Figure 2.11b and Figure 2.13**).

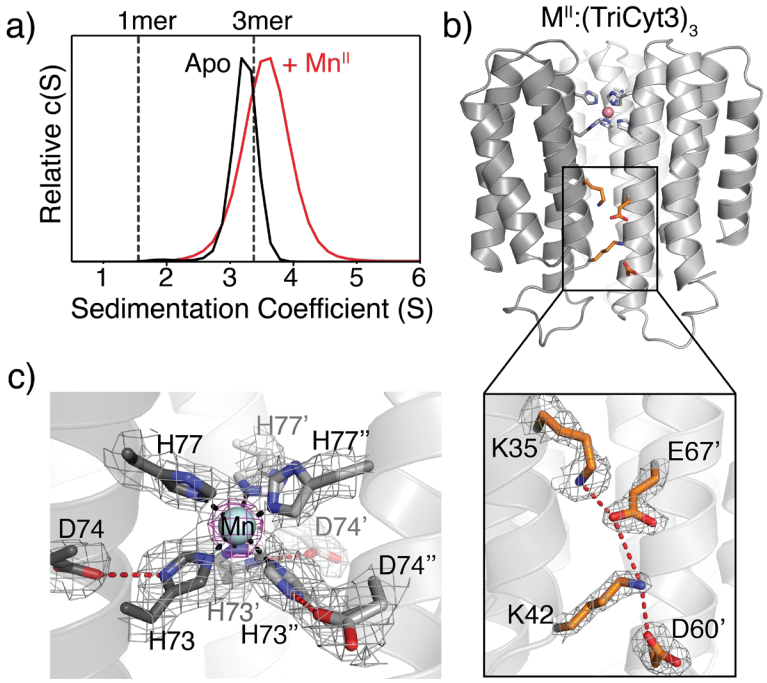

Figure 2.11 | Solution and crystallographic characterization of TriCyt3 (a) SV profiles of TriCyt3 in the absence and presence of $MnCl_2$. (b) Crystal structure of Co^{II}:$(TriCyt3)_3$ (PDB ID: 6WZ2), highlighting engineered H-bonding/electrostatic interactions in peripheral interfaces. (c) His_6-Mn^{II} coordination environment in Mn^{II}:$(TriCyt3)_3$ (PDB ID: 6WZ1). The 2*Fo*–*Fc* (grey) and Mn^{II}-anomalous difference (purple) maps are contoured at 1σ and 5σ, respectively.

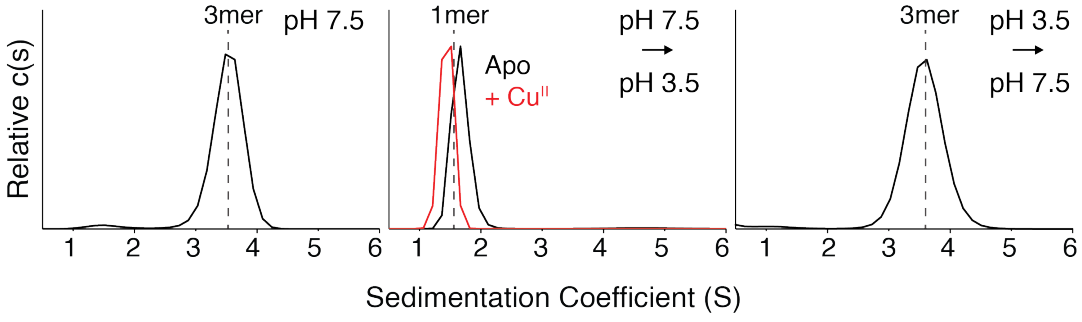

Figure 2.12 | Sedimentation velocity profiles for TriCyt3 at pH 7.5 and pH 3.5, demonstrating pH-reversible assembly/disassembly. TriCyt3 disassembles into monomers even in the presence of Cu^{II} as shown in the middle panel.

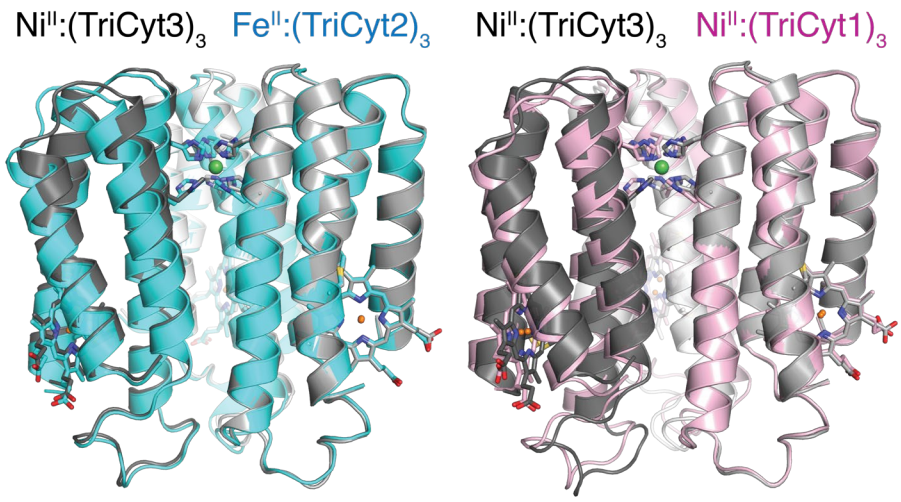

Figure 2.13 | Superposition of Ni^{II}:(TriCyt3)$_3$ (PDB ID: 6WZC), Fe^{II}:(TriCyt2)$_3$ (PDB ID: 6WZ0), and Ni^{II}:(TriCyt1)$_3$ (PDB ID: 6WZA) structures.

The redesigned peripheral interfaces exhibit increased electrostatic complementarity, owing largely to a network of H-bonding/electrostatic interactions involving the Lys and Glu residues (**Figure 2.11b and Figure 2.14**). An examination of metal coordination in the four M^{II}:(TriCyt3)$_3$ structures point to a stable His$_6$ site that can accommodate all tested metal ions in near-octahedral geometries (**Figures 2.11c and Figure 2.15a-c**). Of particular note is the unusual Cu^{II}-His$_6$ coordination, which has–to the best of our knowledge–not been previously observed in a protein scaffold and highlights the ability of the TriCyt$_3$ scaffold to enforce a hexacoordinate geometry (Figure 2.15a, 2.15c). Electron paramagnetic resonance (EPR) spectra of Mn-, Co- and Cu-TriCyt3 complexes are all consistent with metal centers in +2 oxidation states (**Figures 2.17a and Figure 2.18**). Despite the enforcement of His$_6$ binding by the TriCyt$_3$ scaffold, there appears to be some flexibility in metal coordination, as evidenced by a) the relatively high temperature factors of the H77 residues in all structures (**Figure 2.15a**), b) the observation of both Λ and Δ isomers for the Co^{II}-His$_6$ species (**Figure 2.15b**), and c) varying extents of deviation of the

coordination bond angles from perfect octahedral geometry among different metal centers (**Figure 2.16**).

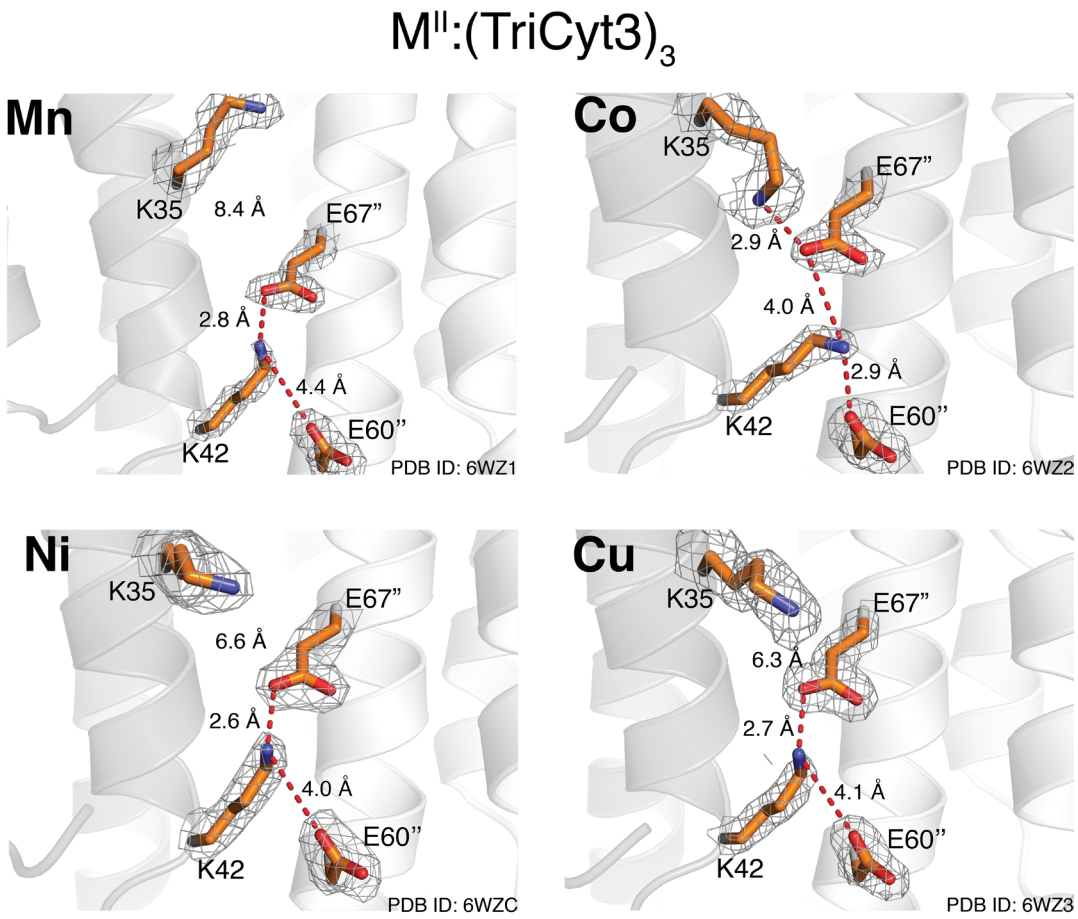

Figure 2.14 | Interfacial salt bridging/H-bonding networks in (M^{II}:TriCyt3)$_3$ assemblies. The conformation of K35 and the intermolecular distances of the E71”/K42/E67” salt bridge vary among different metal-bound structures of TriCyt3.

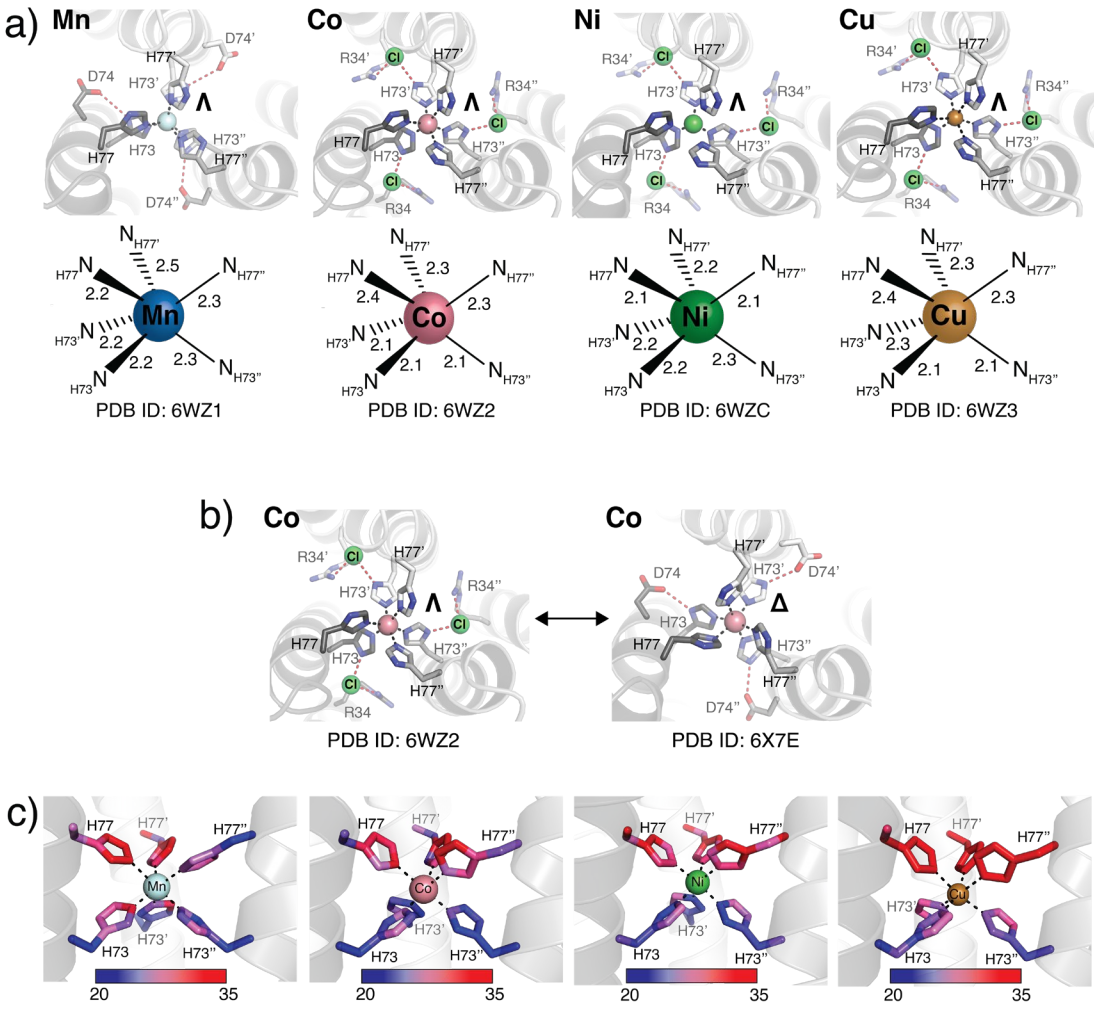

Figure 2.15 | Geometries of the His_6 coordination sites in the metal-bound TriCyt3 trimers. (a) Coordination sites largely featured the Λ configuration, which enables H-bonding interactions between the H73 δ-N's and chloride ions. (b) Co^{II}:$(TriCyt3)_3$ trimers display both Δ or Λ conformations, illustrating fluxionality in the primary coordination sphere. The Δ conformation enables H-bonding interactions between the H73 δ-N's and D74. (c) Metal coordination sites of metal-bound TriCyt3 variants colored according to B-factors, indicating higher mobility/disorder of the H77 residues.

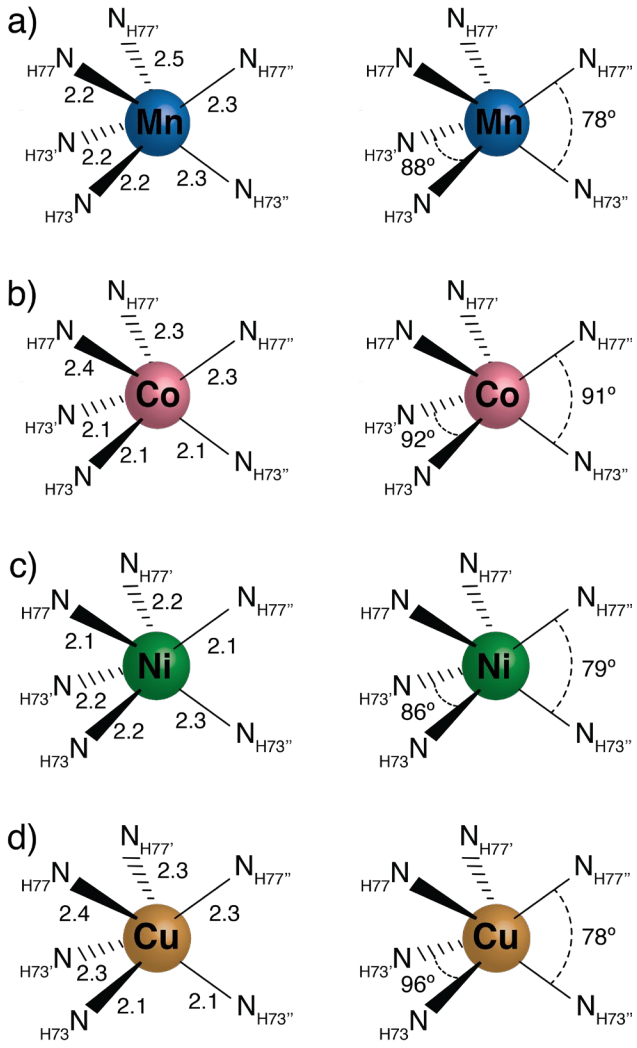

Figure 2.16 | Bond distances and select bond angles of the His_6 coordination sites of (a) Mn^{II}:(TriCyt3)$_3$, (b) Co^{II}:(TriCyt3)$_3$, (c) Ni^{II}:(TriCyt3)$_3$, and (d) Cu^{II}:(TriCyt3)$_3$. Distorted octahedral geometries are observed in all metal-bound TriCyt3 crystal structures.

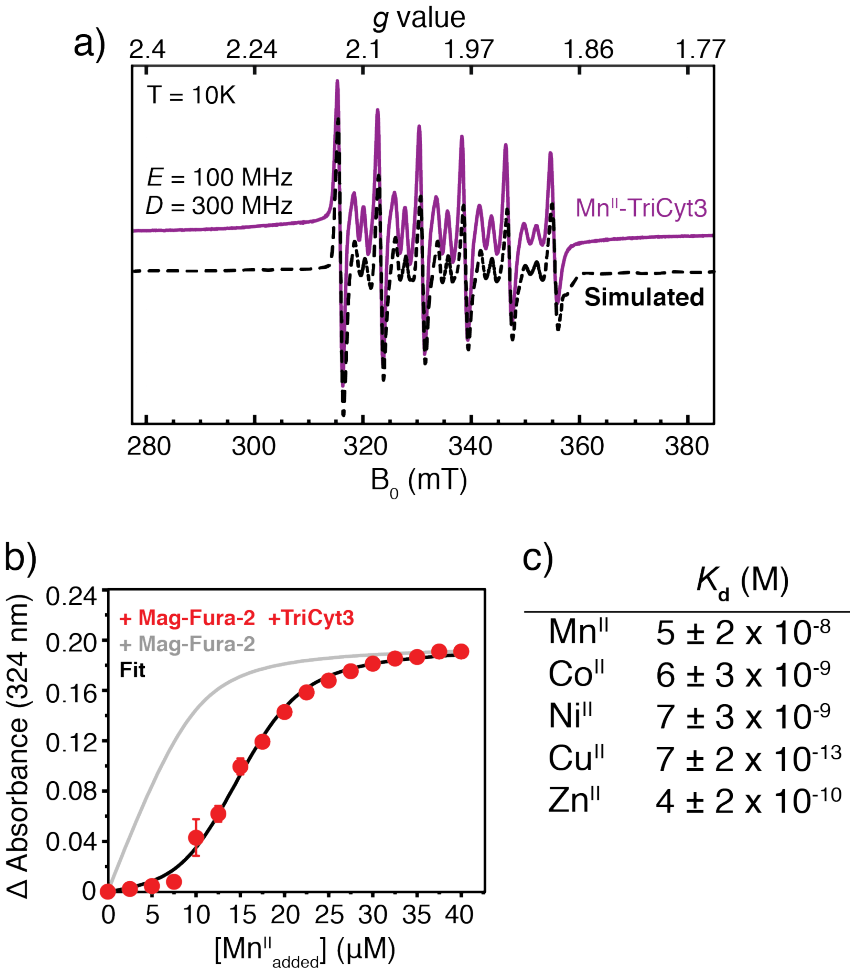

Figure 2.17 | Metal-binding characterization of TriCyt3. (a) X-band EPR spectrum of Mn^{II}:$(TriCyt3)_3$. (b) Mn^{II}-binding isotherm for competitive binding titration of TriCyt3 in the presence of Mag-Fura-2. (c) Dissociation constants for M^{II}:$(TriCyt3)_3$ complexes determined by competition titrations (see also Figures S14 and S15, and Tables S7 and S8). Metal was added in 2.5 μM increments from 2.5 mM metal chloride stock solutions. The standard errors shown are the standard errors of the fits as calculated by the data fitting program DynaFit.

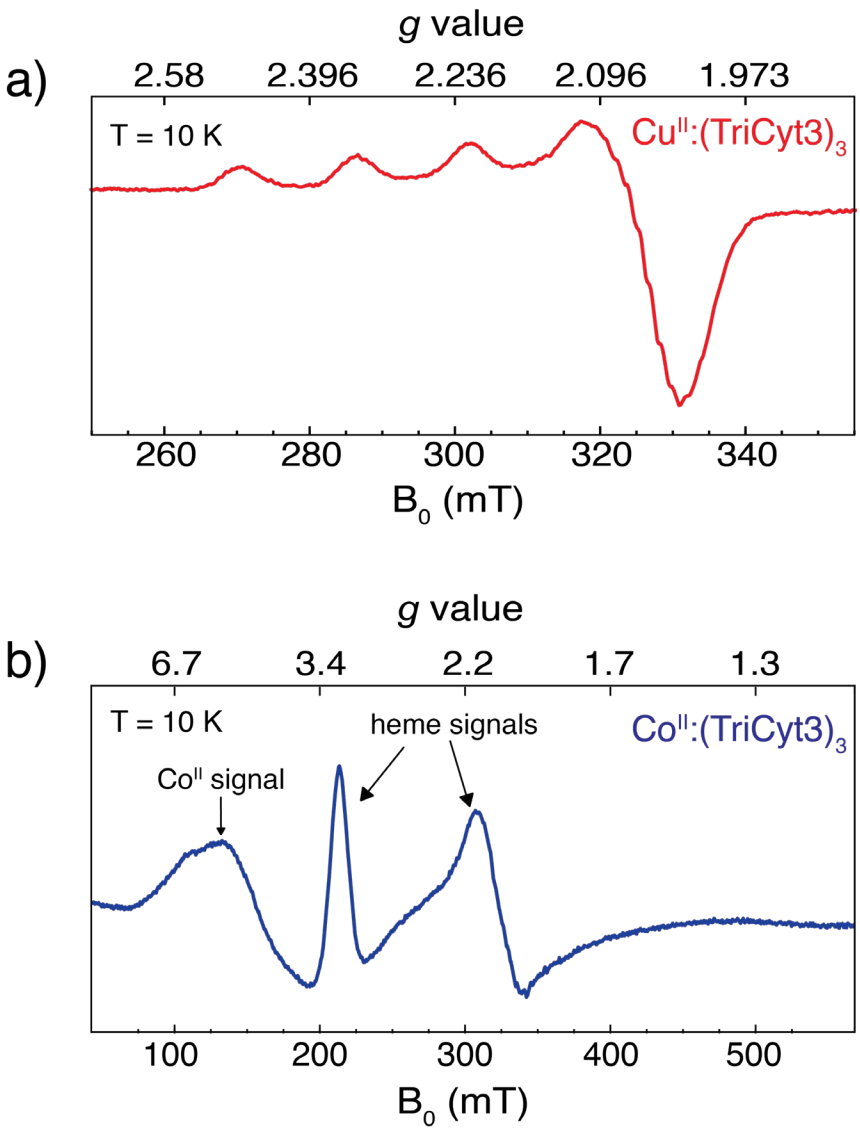

Figure 2.18 | X-band EPR spectra of (a) Cu^{II}-bound TriCyt3 and (b) Co^{II}-bound TriCyt3.

Having uncoupled protein oligomerization from metal binding, we measured the metal binding affinities of $(TriCyt3)_3$ via competitive titrations, using Mag-Fura-2 and Fura-2 as chelating indicators.[23, 33] All titrations were consistent with one M^{II}/one trimer stoichiometry, yielding dissociation constants (K_d) ranging from 50 nM for Mn^{II} to <1 pM for Cu^{II} (**Figures 2.17b, 2.17c, 2.19-2.20, Tables 2.3-2.4**). The Mn^{II} affinity is noteworthy as it approximates the lowest K_d's reported for the Mn-His_6 center of calprotectin (which range from low nM to low μM)[32] and is >1000-fold lower than that for the Mn-regulatory protein, MntR (K_d = 50-160 μM)[34] and >14-fold lower than that for a designed protein with the highest reported Mn^{II}-binding affinity (K_d = 700 nM).[35] Despite the apparent crystallographic disorder in H77 positions, the X-band EPR spectrum of Mn^{II}:$(TriCyt3)_3$ (**Figure 2.16a**) is very similar to that of the His_6-Mn^{II} site in calprotectin.[30] The zero-field splitting of Mn^{II}:$(TriCyt3)_3$ (300 MHz; E/D = 0.30) is in fact lower than that of Mn^{II}-calprotectin (485 MHz; E/D = 0.30), consistent with a highly symmetrical coordination environment (**Table 2.5**).

While the X-band EPR spectrum of Mn^{II}:$(TriCyt3)_3$ indicates a symmetrical coordination environment in solution, the crystal structure reveals that the environment is distorted from an ideal octahedral geometry (**Figure 2.16a**). These distortions are likely enabled by the relatively high flexibility of His77 residues (**Figure 2.15c**), so we aimed to lower their flexibility through the installation of hydrogen bonds to the N_δ's of His77. We surmised that the mutation of Lys80 to Glu80 could lead to hydrogen bonding between Glu80 and the N_δ's of His77 residues, helping to lower the conformational flexibility of His77 and thus enforcing an ideal octahedral geometry at the His_6 coordination site. In addition to the K80E mutation, we mutated Glu81 to Lys81 so that salt bridge interactions between the 80 and 81 residue positions could be retained. The crystal structure of Mn^{II}: $(^{E80/K81}TriCyt3)_3$ revealed a mononuclear Mn^{II}:His_6 primary sphere with an

average bond angle of 91 ± 6.7°, which correlates more closely to an ideal octahedral geometry than that measured in Mn^{II}: $(TriCyt3)_3$ (average bond angle = 94 ± 23°) (**Figure 2.16, 2.21**). Additionally, we observed secondary sphere hydrogen bonds between Glu80 and His77 with an average donor-acceptor distance of 2.7 Å, which corresponds to moderately strong hydrogen bonding interactions. The Mn^{II}-bound crystal structure of $(^{E80/K81}TriCyt3)_3$ demonstrated our ability to modulate primary sphere features of our trimeric architecture through the design of secondary sphere interactions. Facile control over the hydrogen bonding environment could allow us to tune the coordination thermodynamics of the metal binding by tuning the sigma-donor character of N_ε's of His73 and His77.

2.4 Conclusions

In summary, we have reported here the metal-templated design of a series of trimeric protein assemblies (TriCyt1-3), which a) provide tunable coupling between protein quaternary structure and metal coordination, b) furnish metal/pH-switchable protein oligomerization motifs, and c) enable the construction of a robust coordination site for all mid-to-late first-row transition metal ions, including the highest Mn^{II} affinities achieved in an artificial protein. From a practical standpoint, the TriCyt platform offers important advantages for the bottom-up design of functional metalloproteins. Owing to its construction from cytochrome cb_{562} monomers (rather than peptide chains), it is stable and structurally tractable. At the same time, it enables the metal centers to be built in extensive, evolutionarily-naïve interfaces that can be liberally modified to tune the metal-protein interplay and protein oligomerization without affecting protein stability (which stands in contrast to *de novo* designed α-helical metalloproteins).[36, 37] This could, among other possibilities, enable systematic investigations of the redox properties of the unusual Cu-His_6 (and likely also Fe-His_6) coordination motif, as well as the design of coordinatively unsaturated metal centers with

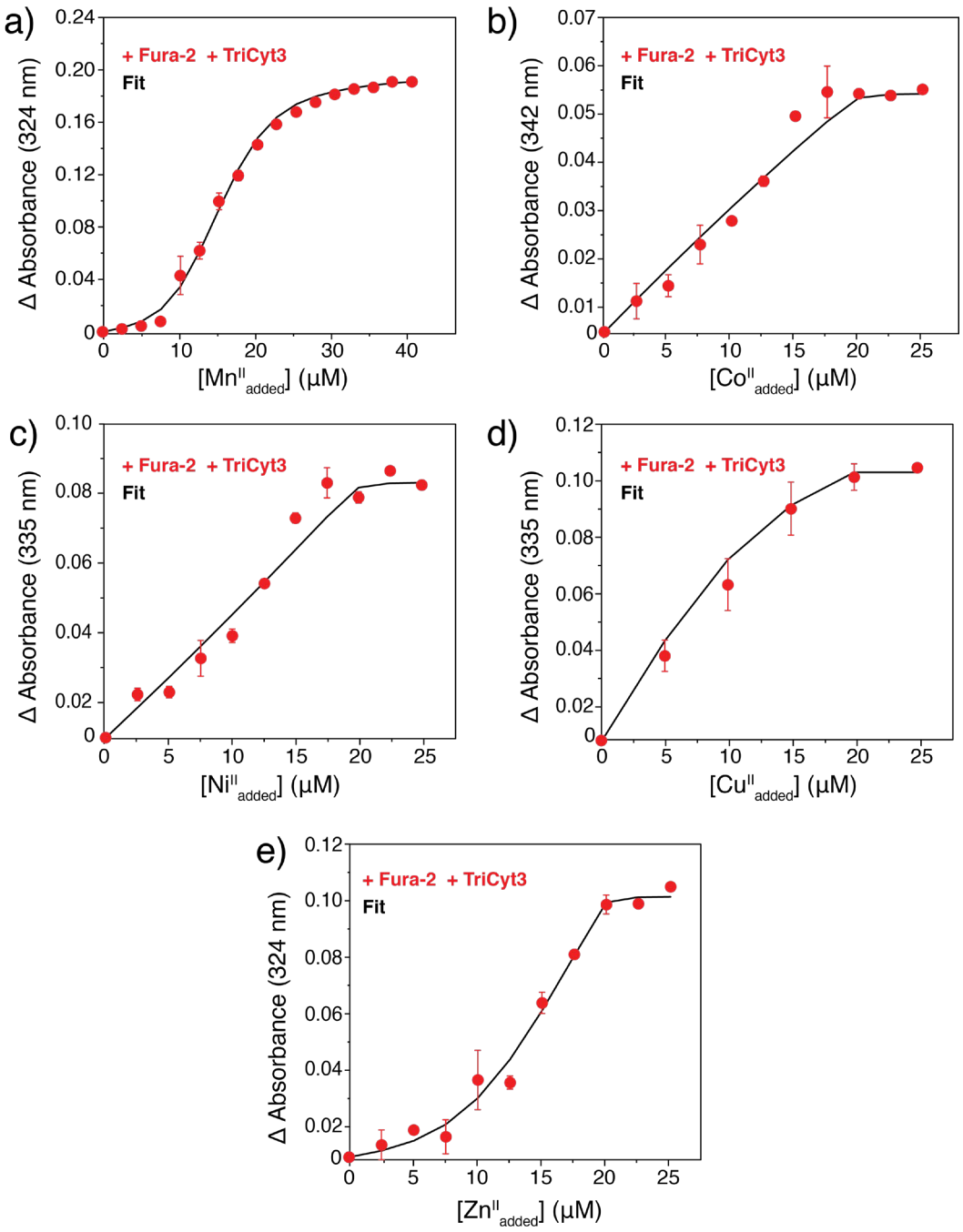

Figure 2.19 | Metal-binding isotherms obtained via competitive binding titrations of TriCyt3 (~10 μM trimer) and Fura-2 (~10 μM) and fitted using a cubic equation. Metals were added sequentially in 1 μL aliquots from either 2.5 mM or 5 mM stock solutions.

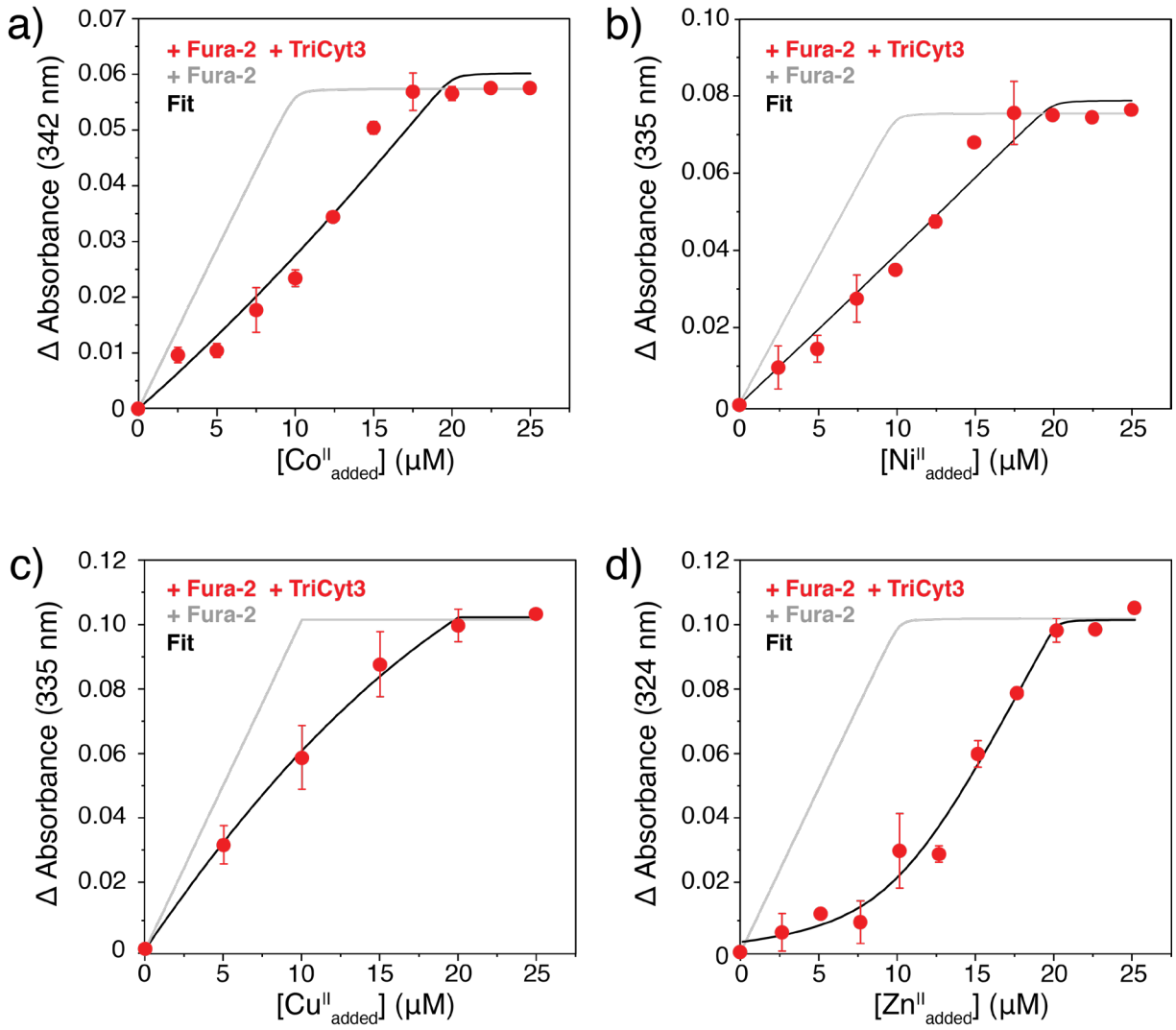

Figure 2.20 | Metal-binding isotherms obtained via competitive binding titrations of TriCyt3 (~10 μM trimer) and Fura-2 (~10 μM) and fitted using DynaFit. Metals were added sequentially in 1 μL aliquots from either 2.5 mM or 5 mM stock solutions.

Table 2.3 | Dissociation constants determined for Metal:TriCyt3 complexes using DynaFit.

Metal	K_d *(TriCyt3)* (M) [a]	K_d *(competing ligand)* (M)	K_d *(TriCyt3)*/K_d *(competing ligand)*
Mn^{II}	$5 \pm 2 \times 10^{-8}$	1×10^{-6} [b]	0.05
Co^{II}	$6 \pm 3 \times 10^{-9}$	9×10^{-9} [c]	0.7
Ni^{II}	$7 \pm 3 \times 10^{-9}$	7×10^{-9} [c]	1.0
Cu^{II}	$8 \pm 7 \times 10^{-13}$	3×10^{-13} [c]	2.0
Zn^{II}	$4 \pm 2 \times 10^{-10}$	6×10^{-9} [c]	0.07

[a] Standard deviations calculated by DynaFit.

[b] Mag-Fura-2

Table 2.4 | Dissociation constants determined for Metal:TriCyt3 complexes using a cubic fit. .

Metal	K_d (TriCyt3) (M) [a]	K_d *(competing ligand)* (M)	K_d *(TriCyt3)*/K_d *(competing ligand)*
Mn^{II}	$5 \pm 1 \times 10^{-8}$	1×10^{-6} [b]	0.05
Co^{II}	$8 \pm 2 \times 10^{-9}$	9×10^{-9} [c]	0.9
Ni^{II}	$9 \pm 1 \times 10^{-9}$	7×10^{-9} [c]	1.3
Cu^{II}	$7 \pm 2 \times 10^{-13}$	3×10^{-13} [c]	2.3
Zn^{II}	$5 \pm 1 \times 10^{-10}$	6×10^{-9} [c]	0.08

[a] Standard deviations calculated based on fits of two separate titrations.

[b] Mag-Fura-2

[c] Fura-2

Table 2.5 | EPR spectroscopic parameters for Mn^{II}:$(TriCyt3)_3$ derived using EasySpin.

Variant	*g*	a_{iso}-^{55}Mn (MHz)	|D| (MHz)	|E| (MHz)	|E/D|
Mn^{II}:$(TriCyt3)_3$	2.002	247	300	100	0.3

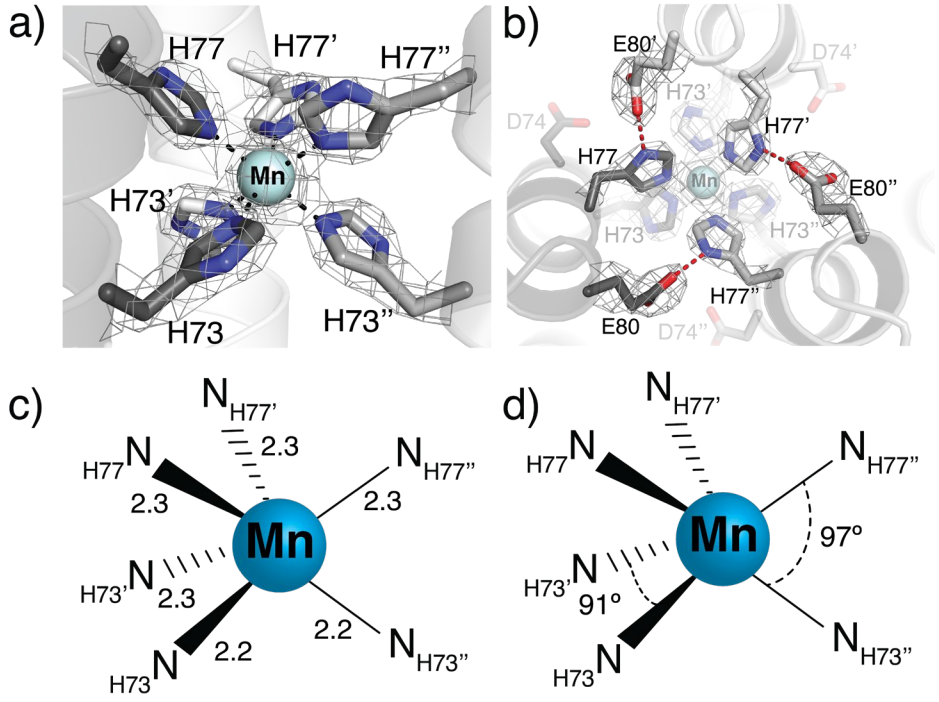

Figure 2.21 | Crystal structure of Mn^{II}: ($^{E80/K81}$TriCyt3)$_3$. (a) The 2Fo-Fc map of the coordination site reveals that all six histidine ligands fully occupy a position in which they coordinate the Mn^{II} center. (b) Topview of the coordination site illustrating H-bonding interactions between E80 and the H77 δ-N's. (c) Select bond lengths and (d) bond angles of the coordination. Both parameters indicate a coordination geometry that more closely approximates an octahedral geometry when compared to those of the coordination site of Mn^{II}:$(TriCyt3)_3$.

potential chemical reactivities. From an evolutionary standpoint, our findings illustrate that even a single mutation on a protein's surface (e.g., W70) can divert protein self-assembly into drastically different pathways, in turn leading to the emergence of new structural motifs with nascent functional sites in protein-protein interfaces.

2.5 Materials and Methods

2.5.1 Protein Mutagenesis, Expression, and Purification

All protein constructs reported in this study are variants of metal-based cytochrome-1 (MBPC1), which is itself a variant of cytochrome cb_{562}.[17] PCR-based site-directed mutagenesis of plasmids encoding MBPC1 (pET20b-[MBPC1]) was performed as previously described (**Table 2.6**).[38] Purified plasmids were transformed into competent BL21 (DE3) *E.coli* cells containing the *ccm* (cytochrome c maturation) cassette plasmid, pEC86.[39] Colonies were grown overnight on LB/agar plates containing ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL). Starter cultures were grown for 16 hours at 37°C in LB media supplemented with the same antibiotic concentrations, diluted 100-fold into fresh, antibiotic supplemented LB media, and then grown for 3-5 hours at 37°C until the OD_{600} reached 0.6-1. Cultures were inoculated into 2.8 L glass flasks containing 1 L of LB media supplemented with antibiotics and shaken at 100 RPM for 20-24 hours at 37°C. Cells were pelleted via centrifugation (5,000 RPM, 4°C, 5 min) and the media discarded. The red cell pellets were resuspended in a 10 mM NaP_i buffer solution (pH 8.0) and vigorously stirred for 30 minutes. The resulting mixture was sonicated for 15 min in pulses of 30 seconds on and 60 seconds off (Qsonica). The lysate was treated with sodium hydroxide to a pH of 10, acetic acid to a pH of 8.0, and then cleared by centrifugation (10,000 RPM, 4°C, 20 min). The cleared lysate was applied to a Q Sepharose Fast Flow resin preequilibrated with a 10 mM NaP_i buffer solution (pH 8.0) and eluted using a step-gradient of 0-500 mM NaCl. The visibly red eluate was

pooled, concentrated, and exchanged into 10 mM sodium acetate buffer solution (pH 3.5). The protein was then loaded onto a CM Sepharose Fast Flow resin preequilibrated with the same buffer solution and eluted using a step-gradient of 0-1 M NaCl. Fractions with Reinheitzahl ratios (A_{415}/A_{280}) above 6 were pooled, concentrated, and exchanged into a 20 mM MOPS buffer solution supplemented with 5 mM EDTA/DPA (pH 7.5). ESI-MS experiments were performed to confirm the identity of purified TriCyt constructs (**Figure 2.22**).

Table 2.6 | Amino acid sequences of the TriCyt series[a]

TriCyt1
ADLEDNMETLNDNLKVIEKADNAAQVKDALTKMRAAALDAQKATPPKLEDKSPDSPEMHDF
FNILVWQIHDALHLANEGKVKEAQAAAEQLKTTCNACHQKYR
TriCyt2
ADLEDNMETLNDNLKVIEKADNAAQVKDALTKMRAAALDA**K**KATPPKLEDKSP**A**SPEM **I**DF
F**D**IL**L**WQIHDALHLANEGKVKEAQAAAEQLKTTCNACHQKYR
TriCyt3
ADLEDNMETLNDNLKVIEKADNAAQVKDAL**K**KMR**K**AALDA**K**KATPPKLEDKSP**A**SPEM**I**DF
VGF**DE**L**A**W**E**IHDAAHLA**K**EGKVKEAQAAAEQLKTTCNACHQKYR

[a] Mutated amino acids in reference to parent TriCyt1 are bolded and underlined.

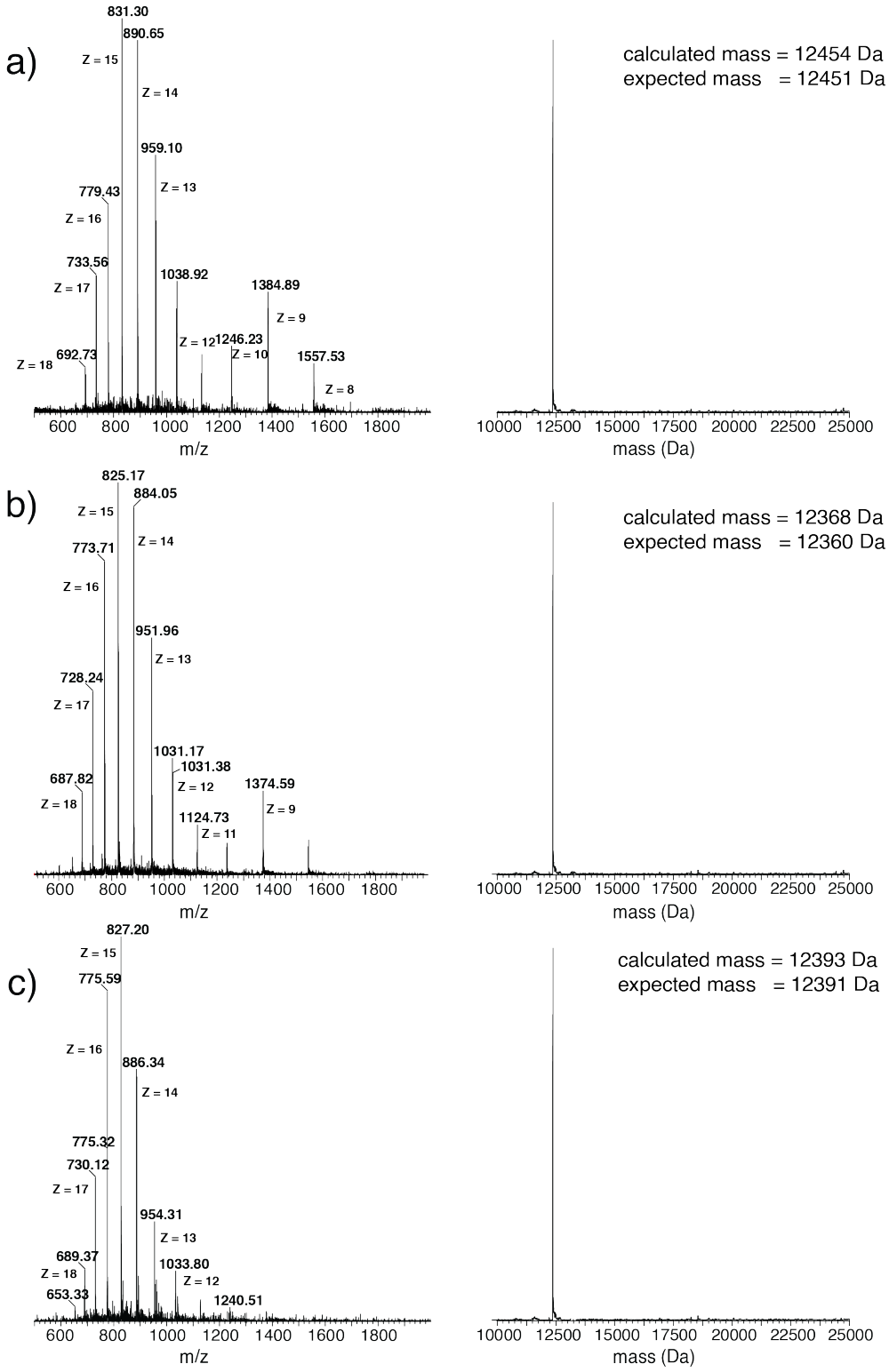

Figure 2.22. | ESI-mass spectra of (a) TriCyt1, (b) TriCyt2, and (c) TriCyt3.

2.5.2 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

Protein solutions (10 μM trimer) were treated with either 1 equivalent of metal chloride salt or 5 mM EDTA/DPA. Sedimentation velocity (SV) measurements were made in a solution of either 20 mM HEPES (pH 7.5) or 20 mM sodium acetate (pH 3.5) at 25°C on a Beckman XL-A instrument equipped with a AN-60 Ti rotor and at 41,000 RPM. Samples were monitored at 420 nm up to 12 h. Scans were processed using SEDFIT software with buffer density (0.99978 g/mL), buffer viscosity (0.01016 poise), and partial specific volume (0.7313 mL/g) parameters calculated by SEDNTERP.[40, 41] SV profiles are shown at a confidence level of 95%. Oligomerization yields were estimated based on Riemann integrations of the peaks of the SV profiles.

2.5.3 Size exclusion chromatography-HPLC (SEC-HPLC)

Solutions of 30 μM protein (10 μM trimer) were treated with 1 equivalent of metal chloride salt. SEC-HPLC analysis was carried out in a solution of 20 mM HEPES (pH 7.5) on a Zenix SEC-300 column (Sepax) at 25°C, operating at a flow rate of 1 mL/min. Samples were monitored at 415 nm and eluted in 20 mM HEPES (pH 7.5). Metal-bound trimeric species eluted in 5.4-5.8 min while monomeric species eluted in 6.2-6.5 min.

2.5.4 Computational redesign of TriCyt1 into TriCyt2 via PyRosetta

The crystal structure of Ni^{II}:$(TriCyt1)_3$ served as the structural model to redesign the interface using the protein design program Rosetta through its Python-based interface (PyRosetta).[42-44] As the metal-binding site and with it the overall conformation was to be kept in the crystallized geometry, no backbone movements were allowed throughout the calculations. Therefore, the nickel atom and heme prosthetic group could be removed prior to calculations. Calculations were performed using a fixed backbone design protocol, which included an initial repacking of sidechains with the natural amino acids followed by sidechain packing and mutation

of 24 amino acids (excluding Cysteine) per monomer through the packer function. The first 1000 trajectories converged to 31 sequences (with >90% of trajectories covered by the 6 most abundant sequences) due to the limited design space afforded by the fixed backbone protocol. Since the differences between the generated models in Rosetta Energy Units (REU) were small, all sequences were aligned and mutational frequencies at all mutated residue positions calculated using UGENE (UniPro).[45] Computationally prescribed mutations with frequencies >80% were considered beneficial and incorporated into TriCyt1 to generate TriCyt2.

2.5.5 X-ray structure determination

Crystals of all TriCyt variants were obtained by sitting-drop vapor diffusion at 25°C (**Table 2.7- 2.8**). To obtain TriCyt1 and TriCyt2 crystals, 1 µL of the protein stock solution (pre-incubated for 1 hour with 1 equivalent of metal per monomer) was mixed with 1 µL mother liquor. To obtain TriCyt3 crystals, 1 µL of a metal-free protein stock solution was mixed with 1 µL mother liquor (pre-incubated with 2.0 equivalents of metal chloride salt per monomer). All crystals appeared within 1 week and were harvested within 1-4 weeks thereafter. Crystals obtained using MPD or PEG400 precipitants were directly frozen in liquid nitrogen, while crystals obtained with other precipitants were transferred into perfluoro polyether (Hampton) for cryoprotection prior to freezing. Diffraction data were collected at 100 K on ALS Beamline 5.0.1, ALS Beamline 5.0.2, ALS Beamline 8.3.1, SSRL 12-2, and SSRL 9-2. Diffraction data were processed and scaled using XDS and XSCALE.[46] Molecular replacement was carried out using Phaser with monomeric cytochrome cb_{562} (PDB: 2BC5) as the search model.[47] Refinement was performed using phenix.refine[48] while model building and placement of metal ions/water was performed using COOT.[49] Electron density maps were generated using Phenix and then converted into CCP4 map files using a Fast Fourier Transform (FFT, CCP4i).[50] All final models and CCP4 electron density

maps were rendered in PYMOL (www.pymol.org). Surface calculations were performed using PISA.

Table 2.7 | Crystallization conditions for TriCyt variants

Protein	Metal	[Protein][a]	[Metal]	Mother liquor
TriCyt1	$NiSO_4$	3 mM	3 mM[b]	22.5% PEG 400, 0.1 M HEPES pH 7.5, 0.2 M NaCl
TriCyt1	$CuSO_4$	3 mM	3 mM[b]	25% PEG 2000, 0.1 M Tris pH 8.5, 0.2 M $CaCl_2$
TriCyt2	None	3 mM	-	43% MPD, 0.1 M Tris pH 8.5, 0.2 M $(NH_4)_2SO_4$
TriCyt2	$FeSO_4$	3 mM	3 mM[b]	25% PEG 400, 0.1 M Tris pH 8.5, 0.2 M $(NH_4)_2SO_4$
TriCyt3	$MnCl_2$	3 mM	3 mM[c]	30% PEG 400, 0.1 M Bis-Tris pH 6.5, 0.2 M $CaCl_2$
TriCyt3	$CoCl_2$[d]	3 mM	3 mM[c]	25% PEG 1500, 0.1 M HEPES pH 7.5, 0.2 M $MgCl_2$
TriCyt3	$CoCl_2$[e]	3 mM	3 mM[c]	25% PEG 1500, 0.1 M Tris pH 8.5, 0.2 M $MgCl_2$
TriCyt3	$NiSO_4$	3 mM	3 mM[c]	25% PEG 2000, 0.1 M Bis-Tris pH 6.5, 0.2 M$CaCl_2$
TriCyt3	$CuSO_4$	3 mM	3 mM[b]	25% PEG 1500, 0.1 M Bis-Tris pH 5.5, 0.2 M $MgCl_2$

[a] [Protein] relates to the concentration of monomeric cytochrome.
[b] Protein solutions were mixed with 3 mM metal ion 1 hour prior to crystallization.
[c] 3 mM metal ion salts were mixed with mother liquor prior to crystallization.

Table 2.8 | X-ray refinement statistics for TriCyt1, TriCyt2, and TriCyt3 crystal structures (numbers in parentheses correspond to values in the highest resolution shell).

Variant	Ni^{II}: $(TriCyt1)_3$	Cu^{II}: $(TriCyt1)_3$	Apo $(TriCyt2)_3$	Fe^{II}: $(TriCyt2)_3$	Mn^{II}: $(TriCyt3)_3$
PDB ID	6WZA	6X8X	6WYU	6WZ0	6WZ1
Data Collection					
Space group	P 3121	P 321	P 1211	P 1211	P 1211
Cell dimensions (Å)	82.12 82.12 137.52	142.16 82.15 48.04	48.26 79.75 50.80	48.20 79.26 49.92	46.76 77.39 47.94
Cell angles (°)	90 90 120	90 90 90	90 105 90	90 107 90	90 111 90
Resolution (Å)	39.35-2.50	39.81-2.51	40.25-1.76	47.78-1.70	34.79-2.00
No. unique reflections	19138 (1865)	12469 (2010)	15659 (3676)	28873 (3931)	21230 (2105)
R_{merge}	0.142 (0.688)	0.038 (0.163)	0.416 (0.416)	0.433 (0.499)	0.079 (0.233)
Multiplicity	19.2 (19.5)	19.5 (20.0)	6.3 (6.3)	6.3 (6.0)	6.6 (6.3)
CC ½	1.0 (0.98)	1.0 (1.0)	0.92 (0.93)	0.91 (0.86)	1.0 (0.98)
$< I / \sigma(I) >$	18.23 (6.38)	53.10 (13.13)	3.05 (2.40)	2.94 (2.11)	18.90 (8.26)
Completeness (%)	99.7(99.9)	99.8 (99.9)	99.8 (100.0)	98.7 (99.7)	98.1 (96.9)
Refinement					
R_{work}/R_{free}	0.240/0.306	0.284/0.344	0.170/0.191	0.182/0.211	0.176/0.229
B-factors ($Å^2$)	34.6	48.1	29.7	29.9	23.1
Protein	34.6	47.9	28.7	29.3	22.9
Ligand/ion	29.6	50.6	27.6	26.4	16.6
Solvent	44.2	50.0	39.4	37.7	29.1
R.m.s deviations					
Bond lengths (Å)	0.009	0.010	0.008	0.007	0.009
Bond angles (°)	1.05	1.15	1.19	0.84	1.04
Clashscore	18.24	12.21	9.81	3.43	8.18
Ramachandran plot (%)					
Favored	98.40	96.15	99.68	99.36	99.36
Outliers	0.00	0.00	0.00	0.00	0.32
Rotamer outliers (%)	0.00	0.00	0.38	0.76	0.77

Table 2.8 (cont.) | X-ray refinement statistics for TriCyt1, TriCyt2, and TriCyt3 crystal structures (numbers in parentheses correspond to values in the highest resolution shell).

Variant	Co^{II}: $(TriCyt3)_3$ (Λ)	Co^{II}: $(TriCyt3)_3$ (Δ)	Ni^{II}: $(TriCyt3)_3$	Cu^{II}: $(TriCyt3)_3$
PDB ID	6WZ2 (Λ)	6X7E	6WZC	6WZ3
Data Collection				
Space group	P 1211	P 1211	P 321	C 121
Cell dimensions (Å)	48.25 78.11 49.56	48.09 78.00 50.19	82.46 82.46 53.12	78.45 81.21 56.43
Cell angles (°)	90 107 90	90 106 90	90 90 120	90 93 90
Resolution (Å)	35.05-2.00	39.70-2.00	35.71-2.20	39.18-1.80
No. unique reflections	22997 (1947)	23245 (2160)	10686 (1068)	31695 (3241)
Rmerge	0.060 (0.425)	0.0418 (0.09344)	0.744 (0.578)	0.034 (0.282)
Multiplicity	6.7 (5.8)	3.4 (3.2)	12.7 (6.6)	3.2 (2.9)
CC ½	1.0 (0.93)	0.998 (0.99)	0.89 (0.89)	1.0 (0.89)
$< I / \sigma(I) >$	20.34 (3.64)	23.76 (9.79)	2.96 (1.89)	22.13 (3.16)
Completeness (%)	96.6 (82.1)	96.34 (90.83)	99.6 (97.4)	97.8 (98.9)
Refinement				
Rwork/Rfree	0.166/0.211	0.189/0.240	0.183/0.218	0.186/0.236
B-factors ($Å^2$)	23.5	21.8	31.3	32.6
Protein	26.2	21.84	30.6	31.9
Ligand/ion	21.9	19.16	30.4	29.6
Solvent	35.5	24.64	38.8	40.1
R.m.s deviations				
Bond lengths (Å)	0.820	0.008	0.910	0.007
Bond angles (°)	0.008	0.97	0.008	0.80
Clashscore	5.00	3.05	3.37	11.22
Ramachandran plot (%)				
Favored	100	99.04	100	100
Outliers	0.00	0.32	0.00	0.00
Rotamer outliers (%)	0.78	0.77	1.12	2.21

2.5.6 Spectroscopic characterization methods

X-band continuous wave (CW) EPR experiments were carried out on a Bruker E500 spectrometer equipped with a Bruker superhigh QE (SHQE) resonator operating in perpendicular mode at 9.4 GHz. Manganese, cobalt, and copper spectra were collected at 10 K using an Oxford ESR900 liquid helium cryostat. Samples were prepared at 1.2 mM protein (400 μM trimer) and 320 μM $MnCl_2$, $CoCl_2$, and $CuCl_2$ (0.8 equiv. metal/trimer) in a 20 mM MOPS buffer solution with 150 mM NaCl (pH 7.5). Spectra were simulated using EasySpin.[51]

2.5.7 Determination of metal-binding affinities of TriCyt3 via competition assays

Competition assays using Mag-Fura-2 and Fura-2 were performed to determine the dissociation constants for binding of divalent metal ions. All buffer solutions were pre-treated with Chelex 100 overnight. Mag-Fura-2 and Fura-2 stocks (1 mM) were prepared in HPLC-grade H_2O. To a 1-mL solution of 20 mM MOPS and 150 mM NaCl (pH 7.5), ~10 μM Mag-Fura-2 or Fura-2 or was added. To this solution, ~30 μM of metal-free protein (~10 μM trimer) was added from a concentrated stock solution. Following a 10-minute equilibration period, 1 μL aliquots of either 2.5 mM or 5 mM $MnCl_2$, $CoCl_2$, $NiCl_2$, $CuCl_2$, or $ZnCl_2$ stock solutions were added. Irreversible electron transfer between free ferrous ions and the ferric-heme co-factors of TriCyt3 precludes the accurate determination of a dissociation constant (K_d) for Fe^{II} binding, therefore Fe^{II} binding titrations were not performed. Metal binding was monitored via UV-visible spectroscopy 10-15 minutes after each metal addition. For titrations with Mag-Fura-2, an increase in absorbance at 324 nm correlated with an increase in the concentration of metal-bound Mag-Fura-2. For titrations with Fura-2, an increase in absorbance at 324 nm (Zn^{II}), 335 nm (Ni^{II} and Cu^{II}), or 342 nm (Co^{II}) correlated with an increase in the concentration of metal-bound Fura-2. In all cases, the metal-binding isotherms were fit using two different methods. In the first method, the average of two

metal-binding isotherms for a given metal was fit to a one-site binding model using DynaFit as previously described.[52] The standard deviations of the calculated K_d values were determined by DynaFit. In the second method, each metal-binding isotherm was fit to a cubic equation as previously described.[53] Notably, the data fitting methods yielded very similar K_d values (**Table 2.3 and 2.4**). Dissociation constants of metal:Fura-2 and metal:Mag-Fura-2 complexes were obtained from previous studies.[33] A DynaFit script and a description of the cubic fitting method can be found in **Appendix 1**.

2.6 Acknowledgements

We thank R. Subramanian and T. Choi for helpful discussions. This work was funded by the NSF (CHE1607145), by the NIH (a CBI traineeship to A.K. through T32GM112584 and R01GM138884 to F.A.T), by NASA (80NSSC18M0093; ENIGMA: Evolution of Nanomachines in Geospheres and Microbial Ancestors (NASA Astrobiology Institute Cycle 8)). Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource at the Stanford Linear Accelerator Center and the Advanced Light Source at the Lawrence Berkeley National Laboratory, which are supported by the DOE, Office of Science, Office of Basic Energy Sciences under contracts DE-AC02-76SF00515 and DE-AC02-05CH11231, respectively. Coordinate and structure factor files for the crystal structures have been deposited into the Protein Data Bank (www.rcsb.org) with the following accession codes: 6WZA (Ni^{II}:TriCyt1$_3$), 6X8X (Cu^{II}:TriCyt1$_3$), 6WYU (TriCyt2$_3$), 6WZ0 (Fe^{II}:TriCyt2$_3$), 6WZ1 (Mn^{II}:TriCyt3$_3$), 6WZ2 (Co^{II}:TriCyt3$_3$, Λ isomer), 6X7E (Co^{II}:TriCyt3$_3$, Δ isomer), 6WZC (Ni^{II}:TriCyt3$_3$), and 6WZ3 (Cu^{II}:TriCyt3$_3$).

Chapter 2 is reproduced, in part, with permission, from: Kakkis, A Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F.A., "Metal-Templated Design of Chemically Switchable Protein

Assemblies with High-Affinity Coordination Sites", *Angew. Chem., Int. Ed. Engl.* **59**, 21940–21944 (2020). The dissertation author was the primary author on all reprinted materials.

2.7 References

1. Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., Metalloproteins and metal sensing. *Nature* **2009,** *460* (7257), 823-830.
2. Irving, H.; Williams, R. J. P., Order of Stability of Metal Complexes. *Nature* **1948,** *162* (4123), 746-747.
3. Frausto da Silva, J. J. R.; Williams, R. J. P., *The biological chemistry of the elements*. Oxford University Press: Oxford, 2001.
4. Cook, S. A.; Hill, E. A.; Borovik, A. S., Lessons from Nature: A Bio-Inspired Approach to Molecular Design. *Biochemistry* **2015,** *54* (27), 4167-4180.
5. Churchfield, Lewis A.; George, A.; Tezcan, F. A., Repurposing proteins for new bioinorganic functions. *Essays Biochem.*. **2017,** *61* (2), 245-258.
6. Nastri, F.; D'Alonzo, D.; Leone, L.; Zambrano, G.; Pavone, V.; Lombardi, A., Engineering Metalloprotein Functions in Designed and Native Scaffolds. *Trends Biochem. Sci.* **2019**.
7. Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valentine, J. S., *Biological Inorganic Chemistry, Structure & Reactivity*. University Science Books: Sausalito, 2007.
8. Lippard, S.; Berg, J., *Principles of Bioinorganic Chemistry*. University Science Books: Mill Valley, 1994.
9. Holm, R. H.; Kennepohl, P.; Solomon, E. I., Structural and functional aspects of metal sites in biology. *Chem. Rev.* **1996,** *96* (7), 2239-2314.
10. Holm, R. H.; Solomon, E. I., Introduction: Bioinorganic Enzymology II. *Chem. Rev.* **2014,** *114* (7), 3367-3368.
11. Eck, R. V.; Dayhoff, M. O., Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. *Science* **1966,** *152* (3720), 363-366.
12. Armstrong, R. N., Mechanistic diversity in a metalloenzyme superfamily. *Biochemistry* **2000,** *39* (45), 13625-13632.
13. Bergdoll, M.; Eltis, L. D.; Cameron, A. D.; Dumas, P.; Bolin, J. T., All in the family: Structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly. *Prot. Sci.* **1998,** *7* (8), 1661-1670.
14. Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A., Metal-Templated Design of Protein Interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, 1827-1832.
15. Salgado, E. N.; Radford, R. J.; Tezcan, F. A., Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **2010**.
16. Churchfield, L. A.; Tezcan, F. A., Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **2019,** *52* (2), 345-355.
17. Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R., Stability and folding kinetics of structurally characterized cytochrome cb_{562}. *Biochemistry* **2006,** *45* (35), 10504-10511.
18. Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A., Controlling Protein-Protein Interactions through Metal Coordination: Assembly of a 16-Helix Bundle Protein. *J. Am. Chem. Soc.* **2007,** *129* (44), 13374-13375.

19. Salgado, E. N.; Lewis, R. A.; Faraone-Mennella, J.; Tezcan, F. A., Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* **2008,** *130* (19), 6082-6084.
20. Salgado, E. N.; Lewis, R. A.; Mossin, S.; Rheingold, A. L.; Tezcan, F. A., Control of Protein Oligomerization Symmetry by Metal Coordination: C_2 and C_3 Symmetrical Assemblies through Cu^{II} and Ni^{II} Coordination. *Inorg. Chem.* **2009,** *48* (7), 2726-2728.
21. Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A., In Vitro and Cellular Self-Assembly of a Zn-Binding Protein Cryptand via Templated Disulfide Bonds. *J. Am. Chem. Soc.* **2013,** *135* (32), 12013-12022.
22. Salgado, E. N.; Brodin, J. D.; To, M. M.; Tezcan, F. A., Templated Construction of a Zn-Selective Protein Dimerization Motif. *Inorg. Chem.* **2011,** *50* (13), 6323-6329.
23. Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A., Evolution of Metal Selectivity in Templated Protein Interfaces. *J. Am. Chem. Soc.* **2010,** *132* (25), 8610-8617.
24. Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A., De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J. Am. Chem. Soc.* **2016,** *138* (40), 13163-13166.
25. Churchfield, L. A.; Alberstein, R. G.; Williamson, L. M.; Tezcan, F. A., Determining the Structural and Energetic Basis of Allostery in a De Novo Designed Metalloprotein Assembly. *J. Am. Chem. Soc.* **2018,** *140* (31), 10043-10053.
26. Song, W. J.; Tezcan, F. A., A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014,** *346* (6216), 1525-1528.
27. Song, W. J.; Yu, J.; Tezcan, F. A., Importance of Scaffold Flexibility/Rigidity in the Design and Directed Evolution of Artificial Metallo-β-lactamases. *J. Am. Chem. Soc.* **2017,** *139* (46), 16772-16779.
28. Fink, A. L., Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold. Des.* **1998,** *3* (1), R9-R23.
29. Garcia-Seisdedos, H.; Empereur-Mot, C.; Elad, N.; Levy, E. D., Proteins evolve on the edge of supramolecular self-assembly. *Nature* **2017,** *548* (7666), 244-247.
30. Gagnon, D. M.; Brophy, M. B.; Bowman, S. E. J.; Stich, T. A.; Drennan, C. L.; Britt, R. D.; Nolan, E. M., Manganese Binding Properties of Human Calprotectin under Conditions of High and Low Calcium: X-ray Crystallographic and Advanced Electron Paramagnetic Resonance Spectroscopic Analysis. *J. Am. Chem. Soc.* **2015,** *137* (8), 3004-3016.
31. Hayden, J. A.; Brophy, M. B.; Cunden, L. S.; Nolan, E. M., High-Affinity Manganese Coordination by Human Calprotectin Is Calcium-Dependent and Requires the Histidine-Rich Site Formed at the Dimer Interface. *J. Am. Chem. Soc.* **2013,** *135* (2), 775-787.
32. Zygiel, E. M.; Nolan, E. M., Transition Metal Sequestration by the Host-Defense Protein Calprotectin. *Annu. Rev. Biochem.* **2018,** *87* (1), 621-643.
33. Grynkiewicz, G.; Poenie, M.; Tsien, R. Y., A New Generation of Ca-2+ Indicators with Greatly Improved Fluorescence Properties. *J. Biol. Chem.* **1985,** *260* (6), 3440-3450.
34. Golynskiy, M. V.; Gunderson, W. A.; Hendrich, M. P.; Cohen, S. M., Metal binding studies and EPR spectroscopy of the manganese transport regulator MntR. *Biochemistry* **2006,** *45* (51), 15359-15372.
35. Rittle, J.; Field, M. J.; Green, M. T.; Tezcan, F. A., An efficient, step-economical strategy for the design of functional metalloproteins. *Nat. Chem.* **2019,** *11* (5), 434-441.

36. Lombardi, A.; Pirro, F.; Maglio, O.; Chino, M.; DeGrado, W. F., De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. *Acc. Chem. Res.* **2019,** *52* (5), 1148-1159.
37. Tebo, A. G.; Pecoraro, V. L., Artificial metalloenzymes derived from three-helix bundles. *Curr. Opin. Chem. Biol.* **2015,** *25*, 65-70.
38. H. Liu, J. N., An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **2008,** *8* (91).
39. E. Arslan, H. S., R. Zufferey, P. Kü, L. Thöny-Meyer, Overproduction of the Bradyrhizobium japonicum c-Type Cytochrome Subunits of the cbb 3 Oxidase in Escherichia coli. *Biochem. Biophys. Res.* **1998,** *251* (3), 744-747.
40. Schuck, P., Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **2000,** *78* (3), 1606-1619.
41. J. Lebowitz, M. L., P. Schuck, Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci.* **2009,** *11* (9), 2067-2079.
42. C.A. Rohl, C. E. M. S., K.M.S. Misura, D. Baker, Protein Structure Prediction Using Rosetta. *Meth. Enzymol.* **2004,** *383*, 66-93.
43. R.F. Alford, A. L.-F., J.R. Jeliazkov, M. J. O'Meara, F.P. DiMaio,H. Park, M.V. Shapovalov. D. Renfrew, V.K. Mulligan, K. Kappel, J.W. Labonte, M.S. Pacella, R. Bonneau, P. Bradley, R.L. Dunbrack, Jr.,R. Das, D. Baker, B. Kuhlman,T. Kortemme, J.J. Gray, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017,** *13* (6), 3031-3048.
44. S. Chaudhury, S. L., J. J. Gray, PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *J. Bioinform.* **2010,** *26* (5), 689-691.
45. K. Okonechnikov, O. G., M. Fursov, Unipro UGENE: a unified bioinformatics toolkit. *J. Bioinform.* **2012,** *28* (8), 1166-1167.
46. Kabsch, W., XDS. *Acta Crystallogr. D* **2009,** *66*, 125-132.
47. Project, C. C., The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D Biol. Cryst.* **1994,** *50*, 760-763.
48. P.D. Adams , P. V. A., G. Bunkóczi , V.B. Chen,I.W. Davis , N. Echols , J.J. Headd , L.W. Hung, G.J. Kapral , R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson ,T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66*, 213-221.
49. P. Emsley, K. C., Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D.* **2004,** *60*, 2126-2132.
50. R. Read, A. S., A Phased Translation Function. *J. Appl. Crystallogr.* **1988,** *21*, 490-495.
51. S. Stoll , A. S., EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **2006,** *178* (1), 42-55.
52. JD Brodin, A. M.-M., T. Ni, EN Salgado, XI Ambroggio, F.A. Tezcan, Evolution of metal selectivity in templated protein interfaces. *J. Am. Chem. Soc.* **2010,** *132* (25), 8610-8617.
53. Wang, Z.-X., An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.* **1995,** *360* (2), 111-114.

Chapter 3: Primary and outer sphere diversification of TriCyt in pursuit of metal-based functions

3.1 Abstract

In pursuit of a protein architecture that could bind all mid-to-late first row transition metals with nanomolar affinity at a mononuclear coordination site, we previously redesigned the metal-directed trimer M:(TriCyt1)$_3$ (M = Mn^{II}-Zn^{II}) into second and third-generation variants (TriCyt2 and TriCyt3, respectively) which feature increased preorganization of a hexahistidine (His_6) coordination site. The culmination of our redesign efforts, TriCyt3, is a constitutive trimer which binds Mn^{II} with nanomolar affinity ($K_d \sim 50$ nM) at the His_6 site.[1] While TriCyt3 in its first iteration functions as a high affinity binder of mid-to-late first row transition metals, we aimed to establish whether this metalloprotein could undergo further modification to mediate metal-based functions encompassing lanthanide coordination, catalysis, and iron-sulfur (Fe-S) cluster binding. Since the functional role of a metalloprotein is inextricably linked to its primary and outer coordination sphere characteristics, we describe herein the diversification of primary and outer sphere features of TriCyt3 in pursuit of a diverse array of metal-based functions.

3.2 Introduction

If one were to build a functional metalloprotein from scratch, what would be the key components? To start, we require ligands to coordinate the metal ion and furnish a primary coordination sphere. The primary sphere features should be compatible with the desired function. For example, if we were interested in designing a hydrolytic metalloenzyme we would avoid utilizing coordinatively saturated metal binding sites, as such sites preclude a direct interaction between the hydroxide nucleophile and the metal ion. Beyond the primary sphere is the outer sphere, which is composed of residues that interact either directly or indirectly with the primary

sphere and/or substrate. Metalloprotein function hinges on an interplay between the outer and primary spheres.[2] For example, substituting Thr-199 of carbonic anhydrase with an alanine residue and thus eliminating a hydrogen bonding interaction near the active site lowers the enzyme's catalytic efficiency (k_{cat}/K_M) almost 60-fold.[3] To emulate the interplay between primary and outer spheres present in natural metalloproteins, our design approach involves modification of the primary sphere to generate the desired metal coordination environment (**Figure 3.1**). If there is evidence for an emergent function, we then carry out additional modifications around the metal coordination site to improve the function.

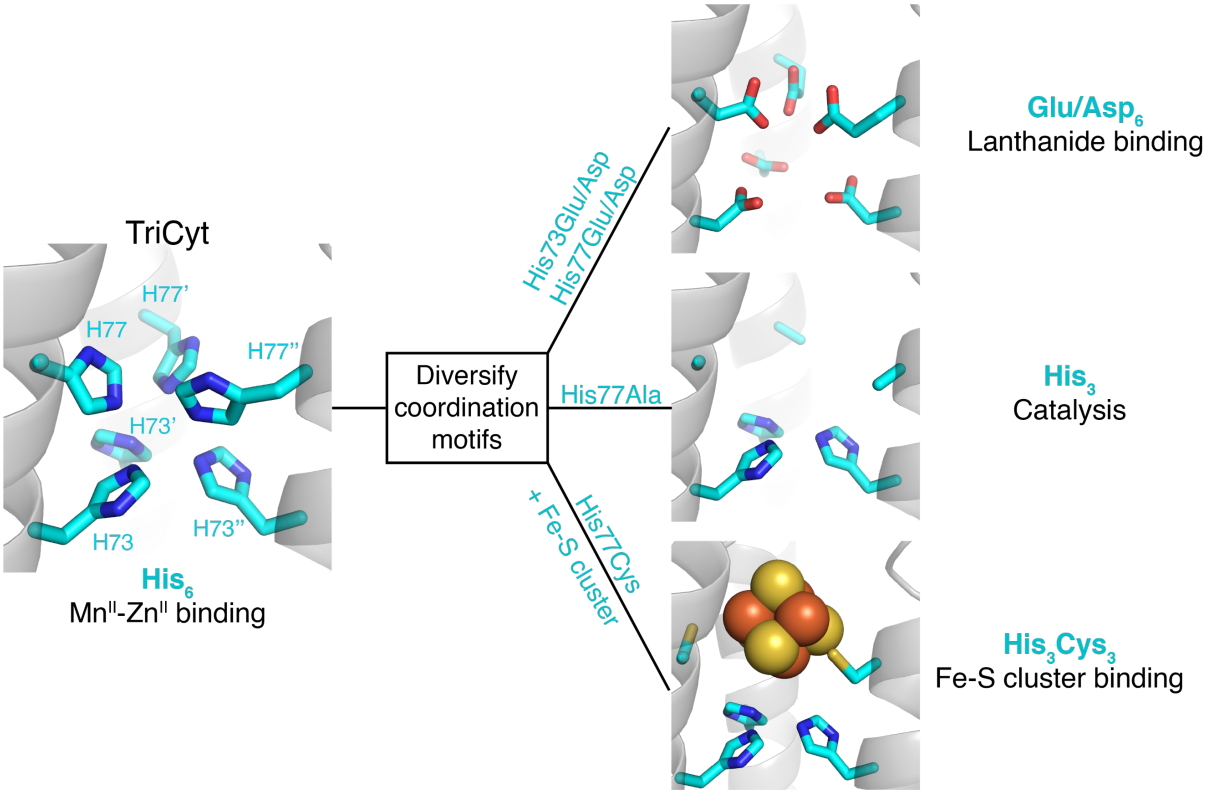

Figure 3.1 | Proposed mutations of the His_6 coordination site of TriCyt to render new coordination environments. Each new coordination environment (left) represents a platform to pursue a wide array of metal-based functions. Excluding the His_3 model, all other TriCyt models with modified primary spheres were manually generated in PYMOL and not based on crystal structures.

3.3 Results and Discussion

3.3.1 High-throughput screening of lanthanide coordination in TriCy2

Lanthanides represent an important class of rare earth elements (REEs) for utilization in a broad range of technologies, including batteries, MRI contrast agents, and single molecule magnets.[4-6] However, their efficient extraction and separation are key challenges in their technological use. It has been estimated that 90% of the global production of REEs can be attributed to major ore deposits in China, highlighting the need to identify new extraction methods and thus ensure wider geographical distribution of this critical resource.[7] Biology offers an attractive means to extract lanthanides in a sustainable manner. In recent years, the Cotruvo Lab discovered and characterized the protein lanmodulin, which uses a modified EF hand motif to bind lanthanide ions (Ln^{III}) with picomolar affinity and 10^8 fold selectivity over Ca^{II}.[8-10] In contrast to its Ln^{III}/Ca^{II} selectivity, the range of apparent K_d for Ln^{III} coordination (La^{III}-Y^{III}) only spanned two orders of magnitude, with moderate selectivity for smaller lanthanide ions (La^{III}-Sm^{III}).[8]

Given the unique spectroscopic properties and technological applications of metal ions within the lanthanide series,[11, 12] metalloprotein constructs with high selectivity for individual lanthanide ions are an attractive protein design target. Recently, Slope et. al. demonstrated that the size selectivity of Ln^{III} binding within a peptide trimer could be systematically tuned based on the location of the Asn_3Asp_3 coordination site along the C_3 symmetry axis.[13] Notably, placement of the coordination site in more structurally flexible regions of the peptide resulted in lower size selectivity.[13] Though this study provides insight into engineering size discrimination in lanthanide binding proteins through pre-organization, its scope is limited by the use of a peptide scaffold and only two primary sphere environments (Asp_3 and Asn_3/Asp_3). Peptides are often difficult to crystallize, which limits the structural insight that can be gained regarding size selective metal

coordination, while Asp_3 and Asn_3/Asp_3 represent only a small fraction of hard, oxygen-rich coordination environments that could comprise a lanthanide binding site. In addressing these limitations, we propose a high-throughput screening approach to quickly identify TriCyt variants which bind Ln^{III} ions and establish whether this binding behavior is size selective (**Figure 3.2**).

In this approach, we would generate our library using VAN codon randomization at residue positions 73 and 77 of TriCyt (**Figure 3.2a**). Notably, the VAN codon encompasses all O-based ligands present in landmodulin (Asn, Asp, Glu). As TriCyt is expressed in the periplasm, chloroform "shock" would serve as a rapid method to extract variants from the cellular milieu

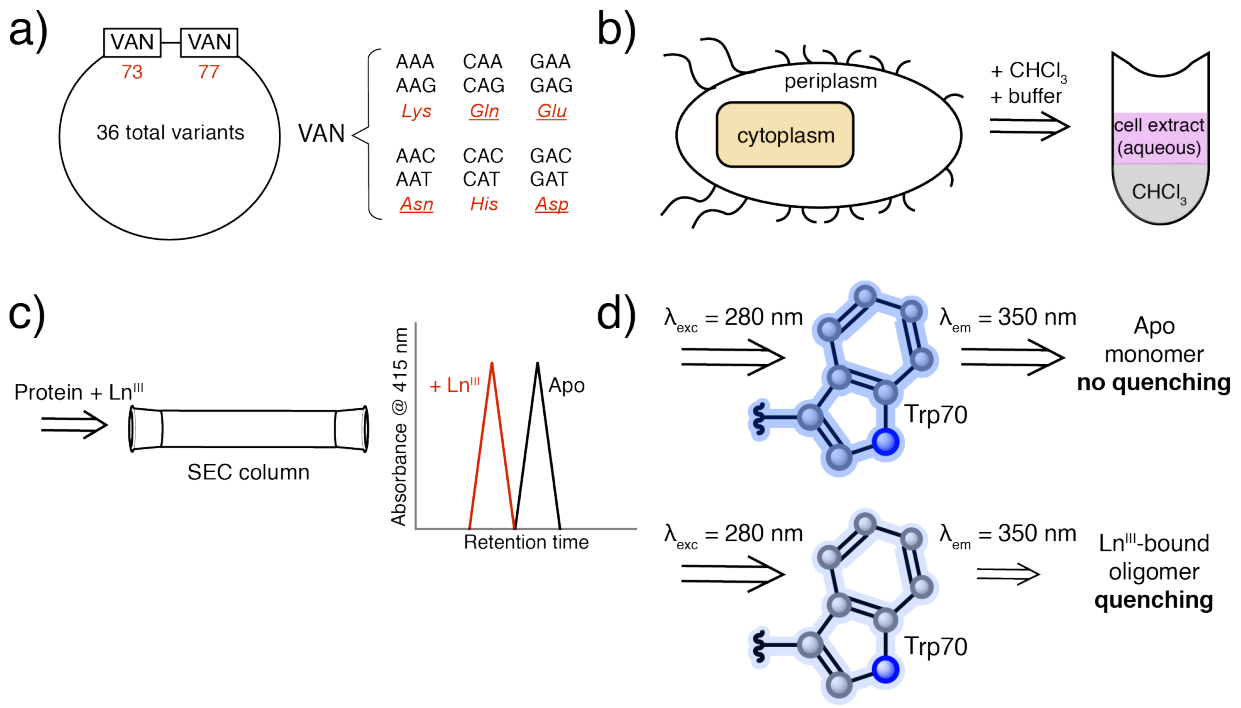

Figure 3.2 | High-throughput screening approach to identify lanthanide binding variants of TriCyt2. (a) Library generation. The VAN codon encompasses all amino acid residues with O-based ligands, excluding Tyr. V: A/C/G, N: A/C/G/T. (b) Periplasmic extraction using chloroform "shock". The "aqueous" phase is composed of 20 mM HEPES/MOPS (pH 7.5). (c) Proposed method to assess oligomerization state of protein variants in the absence and presence of Ln^{III} via SEC-HPLC. (d) Proposed method to assess oligomerization state of protein variants in the absence and presence of Ln^{III} via fluorescence quenching experiments. Given that Trp70 is located ~7.7 Å away from the metal coordination site, fluorescence energy transfer (FRET) to metal ions is another possible quenching mechanism in addition to protein oligomerization.

(**Figure 3.2b**). Once extracted, the protein variants could be screened for lanthanide coordination using two different approaches (**Figure 3.2c-d**). In the first method, protein variants would be mixed with Ln^{III} ions at 0.3 equivalents (equiv.) Ln^{III}/monomer and the resulting oligomerization states assessed via size exclusion chromatography-HPLC (SEC-HPLC, **Figure 3.2c**). In this method, the absorption of the heme co-factor at 415 nm (A_{415}) would serve as the spectroscopic handle for oligomerization, as interference from other absorbing species present in the extracts is more likely when measuring A_{280}. In the second method, significant changes in the fluorescence emission of Trp70 upon metal addition could indicate metal-directed oligomerization, as the solvent environments around the tryptophan residues should shift upon the conversion of monomer to trimer (**Figure 3.2d**).[14] We anticipated that TriCyt2 would be the ideal scaffold for our high-throughput screening methodology, as its metal-dependent trimerization behavior makes it amenable to the oligomerization screening approaches described above.

Before implementing a full screen of our proposed protein library, we carried out experiments to test the viability of each aspect of the high-throughput screening method (**Figure 3.2**). Sequencing following VAN codon randomization of residue positions 73 and 77 of TriCyt2 revealed low sequence diversity, with bias toward His73/Asp77 and His73/Glu77 coordination sites (**Figure 3.3a**). Additional rounds of randomization would be needed to increase the sequence diversity. In parallel, we tested the efficiency of cytochrome extraction from the periplasm via chloroform "shock". From a 20 mL growth of TriCyt3 we extracted about 0.4 mL of 25 μM protein monomer (estimate of [protein] based on A_{415}), enough for an SV-AUC measurement. The major oligomeric species identified was a trimer (**Figure 3.3b**), demonstrating that chloroform "shock" is an effective method for extracting TriCyt variants from cells without lowering their propensity to self-assemble.

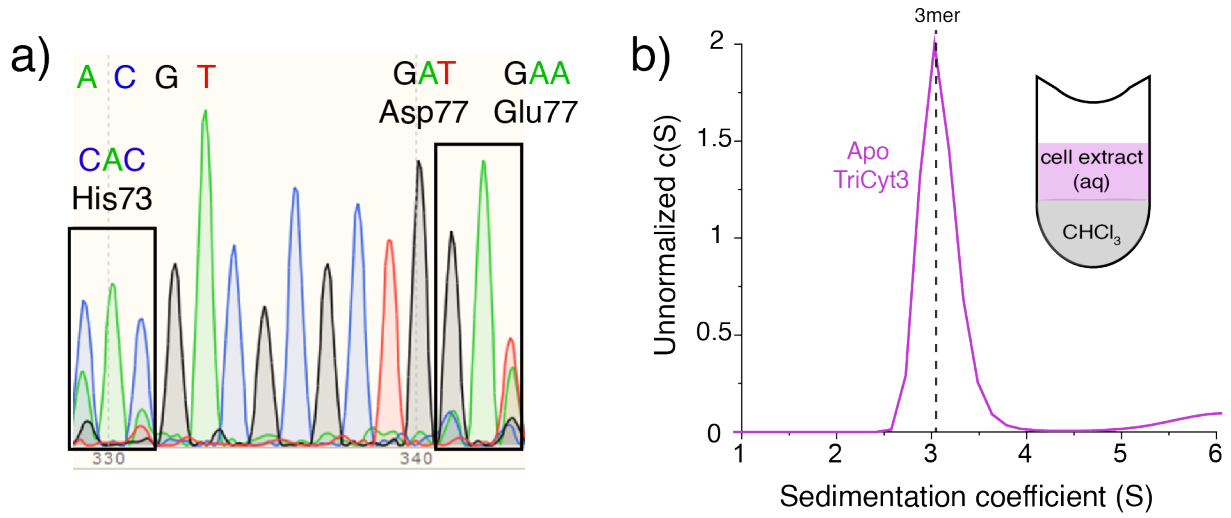

Figure 3.3 | Sequence chromatogram following VAN codon randomization and assembly properties of periplasmic extract of TriCyt3. (a) Sequence chromatogram following VAN codon randomization of residue positions 73 and 77 of TriCyt2. (b) SV-AUC of TriCyt3 following chloroform-mediated extraction from the periplasm. [TriCyt3] = 25 μM

Having established a suitable method for periplasmic extraction of TriCyt, we proceeded to extract TriCyt2 and assess its oligomerization behavior by SEC-HPLC. To our surprise, TriCyt2 extract in the absence of exogenously added metal ions had a retention time of ~6.5 minutes when eluting from the column, while extract mixed with 10 mM EDTA/DPA had a retention time of ~8 minutes (**Figure 3.4**). The retention time of apo TriCyt2 extract correlated closely with that of Ni^{II}-supplemented extract, suggesting that TriCyt2 could acquire trace metal ions from the periplasm or buffer solutions to form metal-directed assemblies. For high-throughput experiments, we would need to incubate extracts with excess EDTA/DPA overnight and then perform buffer exchange into strictly metal-free buffer prior to metal addition. Such steps would lower the potential for "false positives" when screening for lanthanide-directed oligomerization. While we have demonstrated that SEC-HPLC can be used to assess oligomerization states of nominally impure TriCyt extracts when monitoring A_{415} and is thus a promising high-throughput screening approach, it is possible that lanthanide-bound protein complexes with lower affinity than M^{II}:$(TriCyt2)_3$ (M = Mn-Zn) would not be isolable via HPLC.

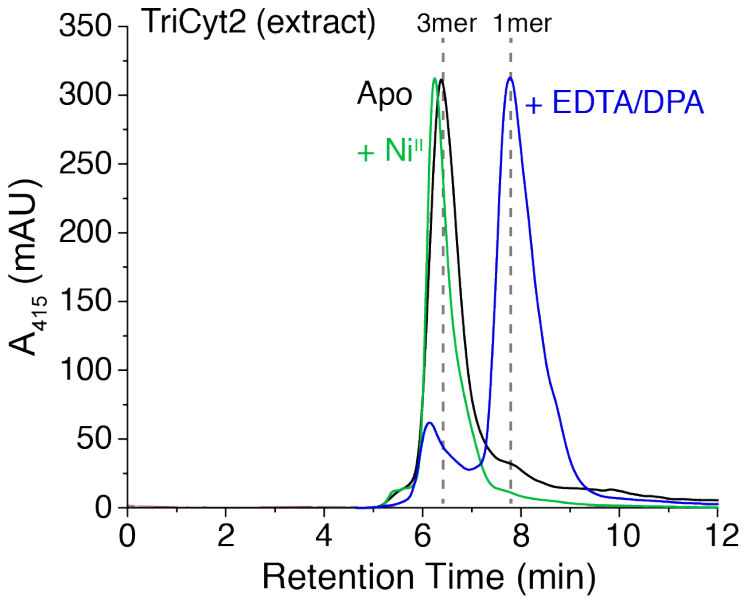

Figure 3.4 | Scaled SEC-HPLC chromatograms of TriCyt2 extracts following chloroform "shock". Apo (black), + 0.3 equivalents Ni^{II}/protein monomer (green), + 10 mM EDTA/DPA (blue). The running buffer is 20 mM MOPS (7.5). Protein extract was incubated with metal for at least one hour prior to loading onto column.

Tryptophan fluorescence experiments represent an alternative screening approach, whereby Trp70, ubiquitous in the TriCyt series, would serve as a fluorescent reporter for oligomerization state changes upon the addition of metal ions (**Figure 3.2d**). As a proof-of-concept, we incubated a purified TriCyt2 variant (^{C15}TriCyt2) with mid-to-late first row transition metals (Mn^{II}, Co^{II}, Ni^{II}, Cu^{II}, Zn^{II}) which are known to direct self-assembly and monitored the change in fluorescence emission at 350 nm (λ_{exc} = 280 nm) relative to apo sample. The addition of these metal ions led to an average change in the fluorescence emission of -2211 ± 619 AU relative to apo protein (**Figure 3.5a**). The addition of Mg^{II} and La^{III}, which is not known to direct assembly of TriCyt, induced fluorescence emission changes almost 10-fold lower (~305 AU and ~-230 AU, respectively). Notably, although the addition of all mid-to-late first row transition metal ions at 30 μM TriCyt2 monomer leads to near-quantitative trimer yield (**Figure 2.5**), metal addition to ^{C15}TriCyt2 induces changes in fluorescence emission that vary by as much as 1350 AU (**Figure**

3.5a). This large difference in fluorescence emission at 350 nm may be due to fluorescence resonance energy transfer (FRET) from Trp70 to the exogenously added metal ions. As described by the Zagotta Group, the magnitude of transition metal ion FRET would depend on the absorptivity of each metalloprotein complex at 350 nm, which should be metal-ion-identity dependent.[15, 16] Encouraged by the fluorescence emission data obtained for a purified TriCyt2 variant, we conducted fluorescence experiments for periplasmic extracts of TriCyt2 and a ΔHis_6 variant, $^{A73/A77}$TriCyt2. Experiments with $^{A73/A77}$TriCyt2 were intended to control for other potential mechanisms of tryptophan fluorescence enhancement/quenching, including FRET from Trp70 to surface-bound metal ions and enhancement/quenching by other periplasmic proteins. Among three separate periplasmic extracts of TriCyt2, the average change in fluorescence emission upon the addition of Ni^{II} was -1238 ± 653 AU (**Figure 3.5b**). The average value was >2-fold lower than that obtained with Ni^{II}-supplemented ^{C15}TriCyt2, while the percent error (53%)

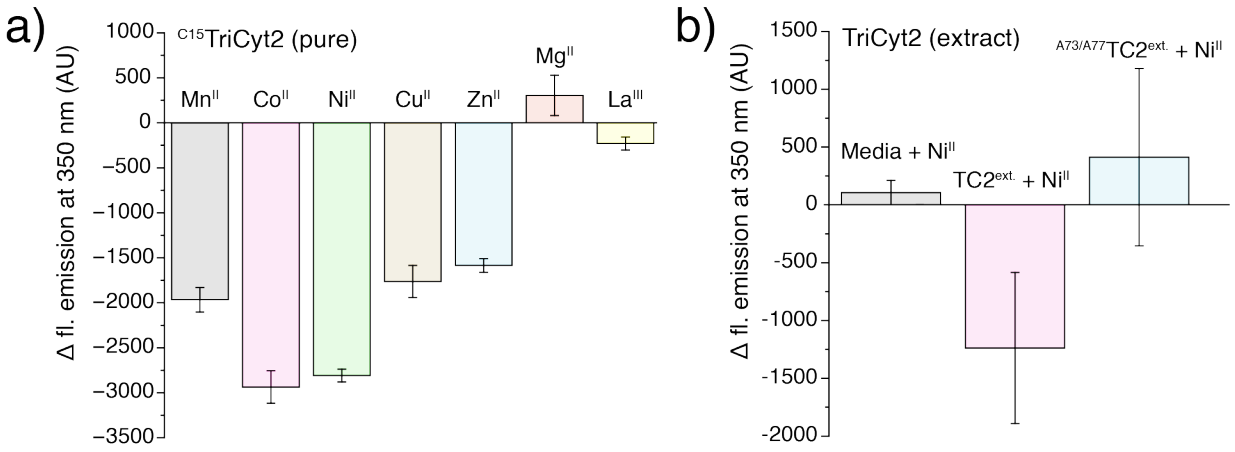

Figure 3.5 | Tryptophan fluorescence emission experiments with *in vitro* purified and chloroform extracts of TriCyt2 variants. (a) Change in fluorescence emission at 350 nm (relative to apo protein) upon metal addition to pure ^{C15}TriCyt2. (b) Change in fluorescence emission at 350 nm (relative to apo protein) upon metal addition to periplasmic extracts of TriCyt2 variants. All measurements were performed in triplicate and at 21-30 μM protein monomer. In all experiments, protein was mixed with 0.3 equivalents metal ion/monomer.

was >20-fold higher. Replicate measurements of Luria-Bertani (LB) media and $^{A73/A77}$TriCyt2 (**Figure 3.5b**) yielded an average fluorescence emission change upon Ni^{II} addition of 106 ± 106 AU and 413 ± 767 AU, respectively.

The very high percent errors, which likely reflect the heterogeneity of the sample matrix, could complicate the interpretation of high-throughput screening data, especially for metal-directed assemblies with below 90-100% trimer yield. Future implementation of this methodology to screen for lanthanide binding variants of TriCyt2 may require variants of higher purity. This could be achieved through a purification step such as affinity chromatography, though the affinity tag itself may interfere with protein self-assembly. An alternative, less laborious means to obtain purer variants would be to express TriCyt2 in the media, precluding the need for an extraction step while placing protein variants in a less heterogenous sample matrix. Since heme incorporation into cytochrome cb_{562} takes place in the periplasm[17], media expression would require the design of a stable, heme-less version of the protein. As described in subsequent sections, this has recently been achieved in our group through the design of "no chrome" (**Chapter 3.4-3.5**).[18] While media expression could better facilitate high-throughput screens, the need to use a heme-less building block prevents us from using heme absorption as a spectroscopic handle for SEC-HPLC measurements. If we can demonstrate that tryptophan fluorescence is a reliable indicator of TriCyt2 assembly in the media, then we could foreseeably rely on fluorescence experiments to screen for lanthanide-directed assembly of TriCyt variants in a high-throughput manner. Yet another alternative is mass-spectrometry, which, while lower throughput, could provide us with a less ambiguous readout for the formation of protein-lanthanide complexes.

3.3.2 Structural characterization and functional screening of a His_3 site in the TriCyt3 scaffold

A significant fraction of metalloenzyme functions rely on the presence of at least one coordinatively unsaturated metal binding site. Coordinative unsaturation at the primary sphere is critical for the binding and activation of substrates and nucleophiles for a wide range of chemical transformations (**Figure 3.6**). Evidently, coordinative unsaturation is just one requirement for efficient catalysis, as metalloenzymes which mediate distinct chemical transformations often utilize identical primary sphere features. For example, carbonic anhydrase and cysteine

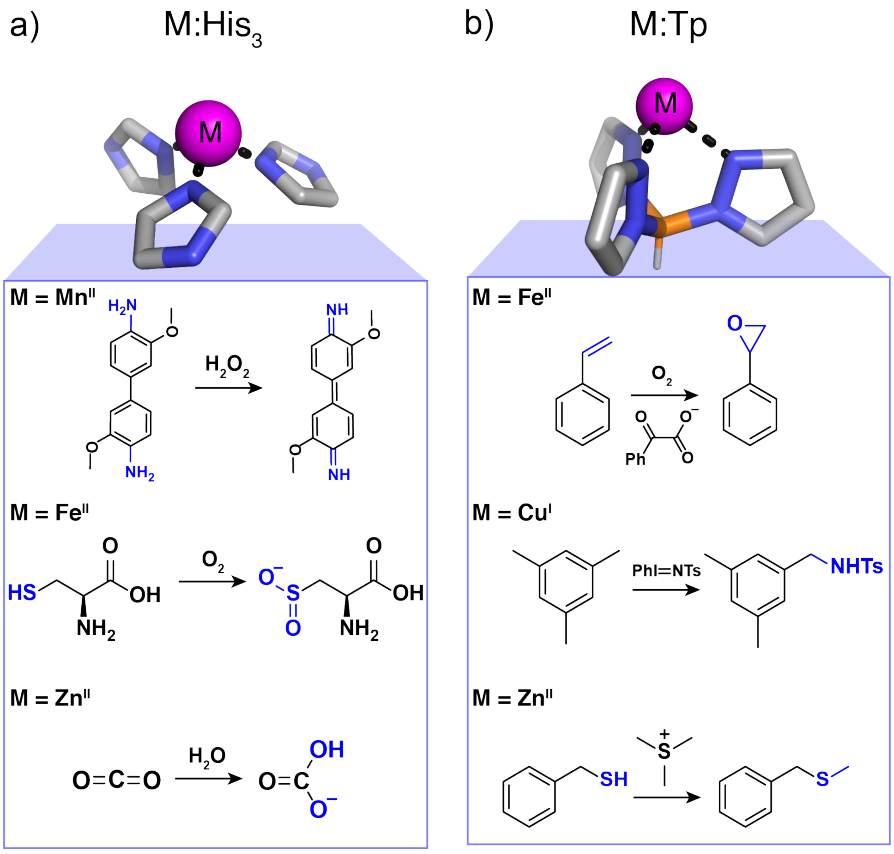

Figure 3.6 | His_3 and Tp ligand frameworks. (a) Examples of reactions mediated by M:His_3 sites in proteins: peroxidation (top), oxygen insertion (middle), and hydrolysis (bottom). Mn^{II} catalyzed oxidation was achieved through repurposing of the His_3 site of carbonic anhydrase. (b) M:Tp complexes have been harnessed for epoxidation (top), amination (middle), and methylation (bottom). The boron atom is colored orange.

dioxygenase, which catalyze different classes of reactions, utilize the same mononuclear, His_3 coordination site (**Figure 3.6a**). Notably, the His_3 coordination site and its close synthetic analogue, trispyrazolyl borate (Tp), have been harnessed for a diverse array of reactions (**Figure 3.6a-b**).[3, 19-24] Through the installation of a His_3 site in TriCyt3 followed by the design of outer sphere interactions, we aspire to mediate one or more of these functions. In this section, we describe our efforts to achieve catalysis in TriCyt3 variants bearing a His_3 primary sphere, with particular focus on peroxidase (**Figure 3.7a-c**) and hydride transfer reactions (**Figure 3.7d**).

Monitoring product formation via UV-vis spectroscopy is a valuable means to screen for catalysis, especially in high-throughput settings. However, the high absorptivity of the heme co-factor in the visible range (ε_{415} = 148,000 M^{-1} x cm^{-1})[17] and its intrinsic reactivity hinder our ability to characterize the catalytic activity of TriCyt3 variants. Recently, we applied Rosetta computational design[25] to obtain a heme-less variant of TriCyt3, termed no chrome TriCyt3 (ncTriCyt3) (**Figure 3.8a**).[18] Like TriCyt3, ncTriCyt3 assembles quantitatively into a trimer at mid μM concentrations and in the absence of metal ions (**Figure 3.8b**). Crystal structures reveal a

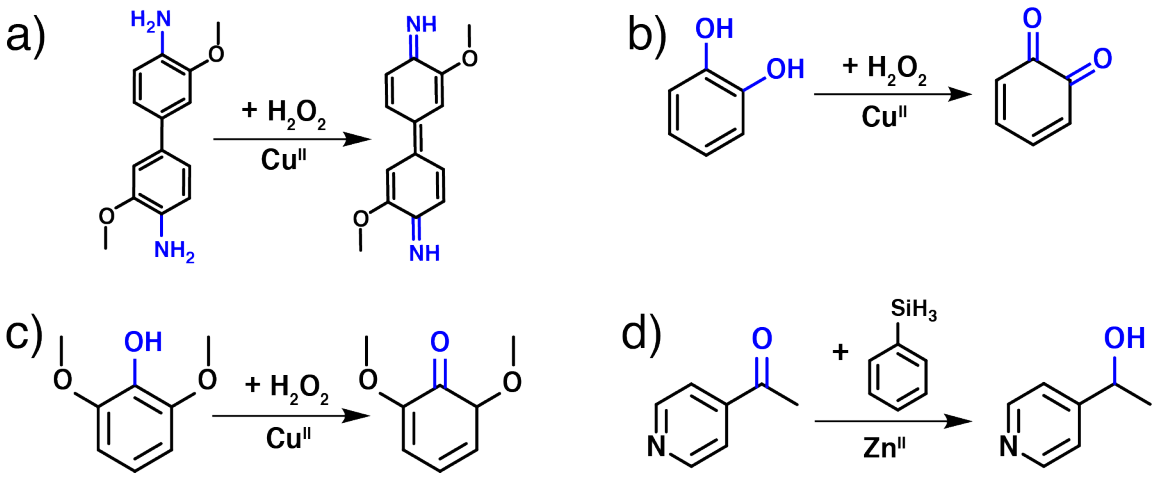

Figure 3.7 | Reactions for screening catalytic activity in TriCyt3 variants hosting His_3 coordination sites. (a-c). H_2O_2-mediated oxidation reactions with *o*-dianisidine (ODA), catechol, and syringol substrates. (d) Hydride transfer.

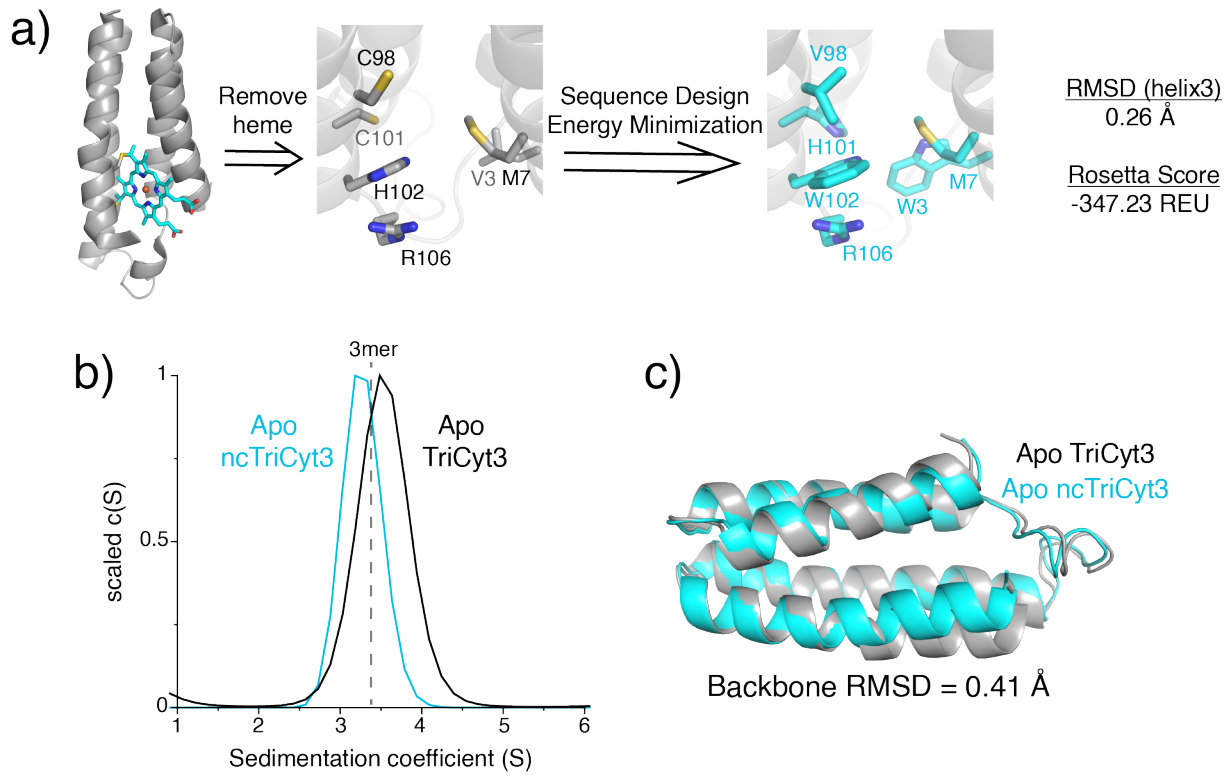

Figure 3.8 | Design and characterization of ncTriCyt3. (a) Summary of design workflow. Iterative cycles of sequence design and energy minimization were employed before final scoring. The main score metrics were 1. RMSD between helix3 of the design model and TriCyt3 and 2. Rosetta score. Helix3 contains most mutations that are required for self-assembly, and so a low RMSD in this region should correlate with similar oligomerization behavior. Ultimately, the model with the lowest helix3 RMSD relative to TriCyt3 was selected for further experimental characterization. (b) SV-AUC demonstrating trimeric assembly of apo ncTriCyt3 and TriCyt3. SV-AUC measurements were performed at 60 μM monomer. (c) Structural overlay of apo ncTriCyt3 and TriCyt3.

backbone RMSD between apo TriCyt3 and ncTriCyt3 of only 0.41 Å (**Figure 3.8c**), illustrating that removal of the heme and redesign of the pocket did not significantly distort the trimeric architecture.[18]

Having isolated and characterized ncTriCyt3, we proceeded to make systematic modifications to the outer sphere while retaining the His_3 coordination site. One outer sphere

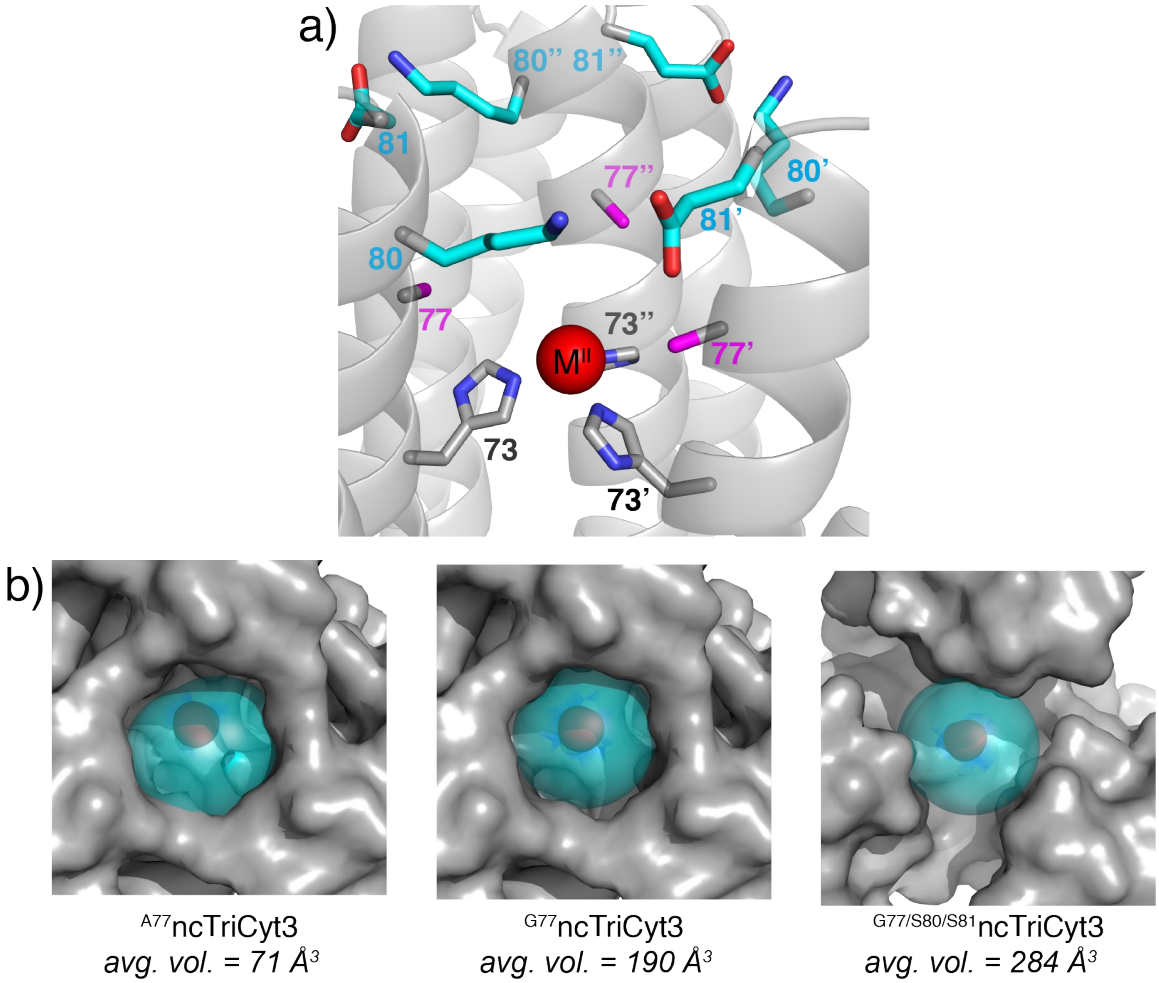

Figure 3.9 | Modeling cavity size of ncTriCyt3 mutants. (a) Model of ^{A77}ncTC3, with residue position 77 highlighted in magenta and positions 80 and 81 highlighted in cyan. Residue positions 80 and 81 mark the apex of the cavity. (b) Topview of various ncTriCyt3 models and predicted cavities (represented by cyan spheres). The active site metal ion is shown as a red sphere.

parameter that is universally important for catalytic function is the cavity space around the active site.[26] We surmised that ncTriCyt3 mutants with varying levels of steric occlusion around the coordination site, and thus varying degrees of substrate accessibility, would catalyze reactions with different efficiencies. To systematically tune the cavity size of ncTriCyt3, we carried out mutagenesis of residue positions 77, 80, and 81. Collectively, these positions mark the two "layers" along the three-fold interface separating the His73 coordination site from bulk solution (**Figure 3.9a**). With ^{A77}ncTriCyt3 as a starting point, we modeled different mutations at residue positions

77, 80, and 81 and then carried out calculations of the cavity size using Caver.[27] To our surprise, mutating Ala77 into Gly77, which ultimately removes only three methyl groups from the pocket, led to a 2.7-fold increase in average cavity volume (**Figure 3.9b**). Further mutations of Lys80 and Glu81 into serine residues led to an overall increase in volume of 213 Å^3 (**Figure 3.9b**). Serine substitutions were selected as they are small but also polar, giving rise to mutants that may be less prone to aggregation compared to mutants that feature more hydrophobic residues. With molecular models in hand, we expressed, purified, and analyzed the oligomerization behavior of all three pocket mutants via SV-AUC (**Figure 3.10**).

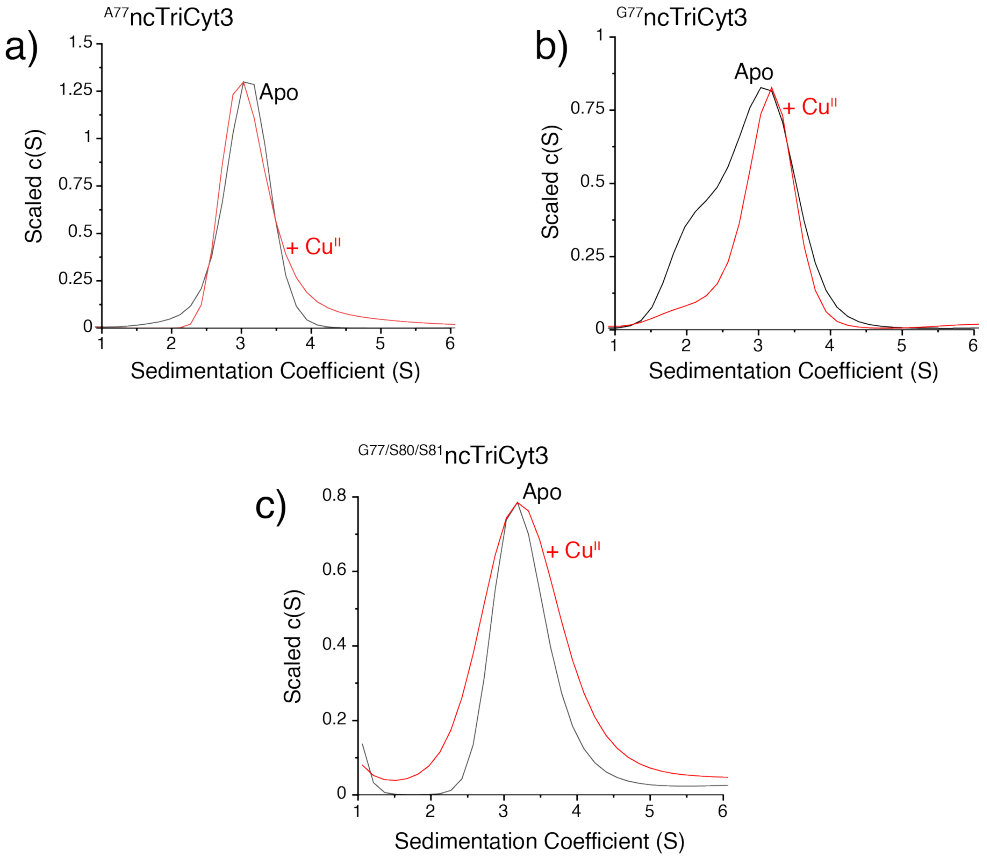

Figure 3.10 | SV-AUC distributions of (a) ^{A77}ncTriCyt3, (b) ^{G77}ncTriCyt3, (c) $^{G77/S80/S81}$ncTriCyt3. Measurements were performed with 60 µM protein monomer and 16 µM $CuSO_4$.

Interestingly, while ^{A77}ncTriCyt3 and $^{G77/S80/S81}$ncTriCyt3 were both almost quantitatively trimeric in the absence of metal, the distribution of apo ^{G77}ncTriCyt3 revealed a significant population of monomeric species (~30% based on relative c(S) values) (**Figure 3.10b**). The addition of 0.24 equiv. Cu^{II}/monomer to ^{G77}ncTriCyt3 increased the relative trimer population to ~82%, while Cu^{II} addition to $^{G77/S80/S81}$ ncTriCyt3 symmetrically broadened the distribution around S = 3.18 relative to apo, suggesting an increase in the flexibility of the architecture (**Figure 3.10b-c**). No symmetric broadening is observed upon the addition of Cu^{II} to ^{A77}ncTriCyt3 (**Figure 3.10a**). Taken together, the SV-AUC distributions illustrate how just a small number of mutations can significantly influence the assembly of TriCyt3. Each variant in this trio represents a potential starting point for screening and eventually improving catalytic functions, such as peroxidation.

In all our peroxidase screens, H_2O_2 is present at 10 mM. Among the peroxidase substrates selected for our screens (**Figure 3.7a-c**), syringol is expected to be the easiest to oxidize owing to its small size and the presence of two electron donating substituents positioned *ortho* to the hydroxyl group, which should lower its reduction potential (**Figure 3.7c**). The syringol oxidation

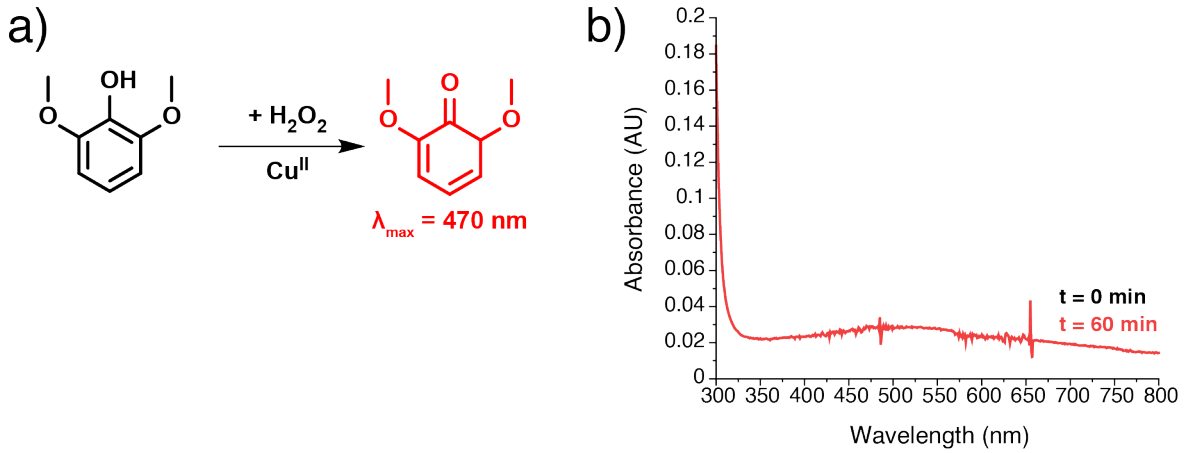

Figure 3.11 | Syringol oxidation activity of ^{A77}ncTriCyt3. (a) Reaction scheme, (b) UV-vis spectroscopy following incubation of Cu^{II}-loaded A77ncTriCyt3 with 1 mM syringol and 10 mM H_2O_2 for 0 minutes (black trace), and 60 minutes (red trace). The black and red traces fully overlap, indicating no product formation.

product has an absorbance maximum at 470 nm, which allows us to assay for activity using UV-vis spectroscopy (**Figure 3.11a**) We initially tested the syringol oxidation activity of Cu^{II}-loaded ^{A77}ncTriCyt3, employing the same protein/Cu^{II} concentrations used for SV-AUC (**Figure 3.11b**). We observed that over the course of 60 minutes there was no increase in the absorbance at 470 nm, indicating no catalytic activity (**Figure 3.11b**).

By contrast, incubating syringol with Cu^{II}-bound ^{G77}ncTriCyt3 under the same reaction conditions led to the formation of oxidation product within 10 minutes (**Figure 3.12a**). Monitoring the change in absorbance at 470 nm every 10 minutes, we observed a rapid increase in A_{470} up to 30 minutes followed by a decrease which we attribute to the formation of precipitate (**Figure 3.12a**). While the chemical identity of this precipitate has not been determined, oxidative polymerization of syringol is a possible rationale, as that would yield polymers with lower aqueous solubility than the monomeric oxidation product (**Figure 3.12a**).[28, 29] Ultimately, precipitation made it difficult to obtain an accurate estimate of the reaction rate (expressed as ΔA_{470} x s^{-1}) of

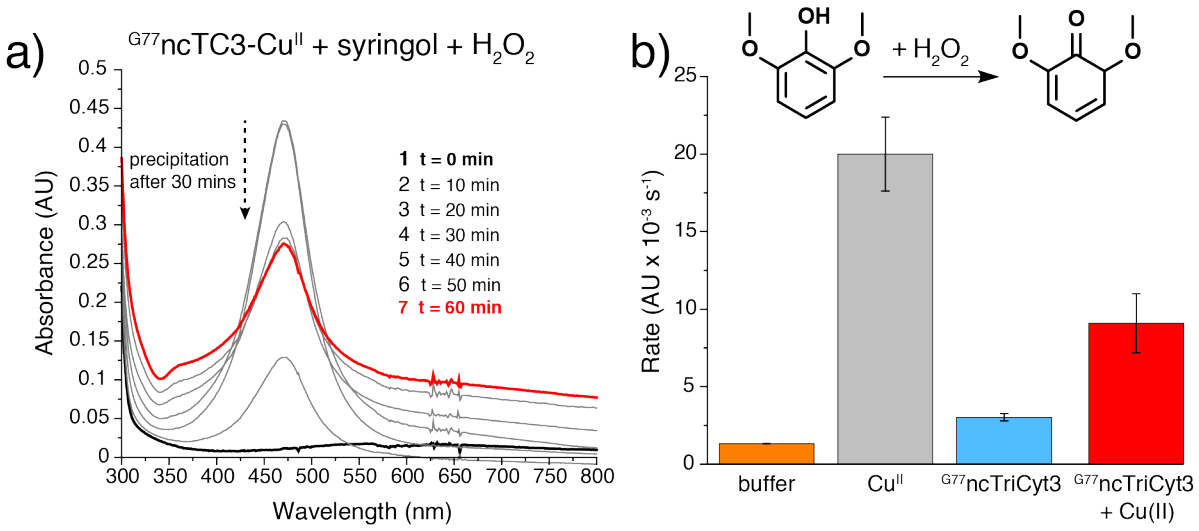

Figure 3.12 | Syringol oxidation activity of ^{G77}ncTriCyt3. (a) UV-vis spectroscopy following incubation of Cu^{II}-loaded ^{G77}ncTriCyt3 with 1 mM syringol and 10 mM H_2O_2. Spectra were recorded every 10 minutes over the course of 1 hour. (b) Summary of syringol oxidation rates under different reaction conditions.

Cu^{II}-loaded ^{G77}ncTriCyt3 over the initial 60 minutes of the reaction. The reaction rate, while about 3.0-fold higher than the reaction rate of apo ^{G77}ncTriCyt3, was about 2.0-fold lower than that of free Cu^{II} (**Figure 3.12b**). Interestingly, no precipitation was observed under catalyst conditions in which both protein and Cu^{II} were present, suggesting either a distinct mechanism for oxidation under these conditions or Cu^{II}-induced precipitation of the protein. Ultimately, the observation that Cu^{II}-loaded ^{G77}ncTriCyt3 had higher peroxidase activity than apo protein encouraged us to screen for peroxidase activity of this variant with additional substrates. When using catechol and *o*-

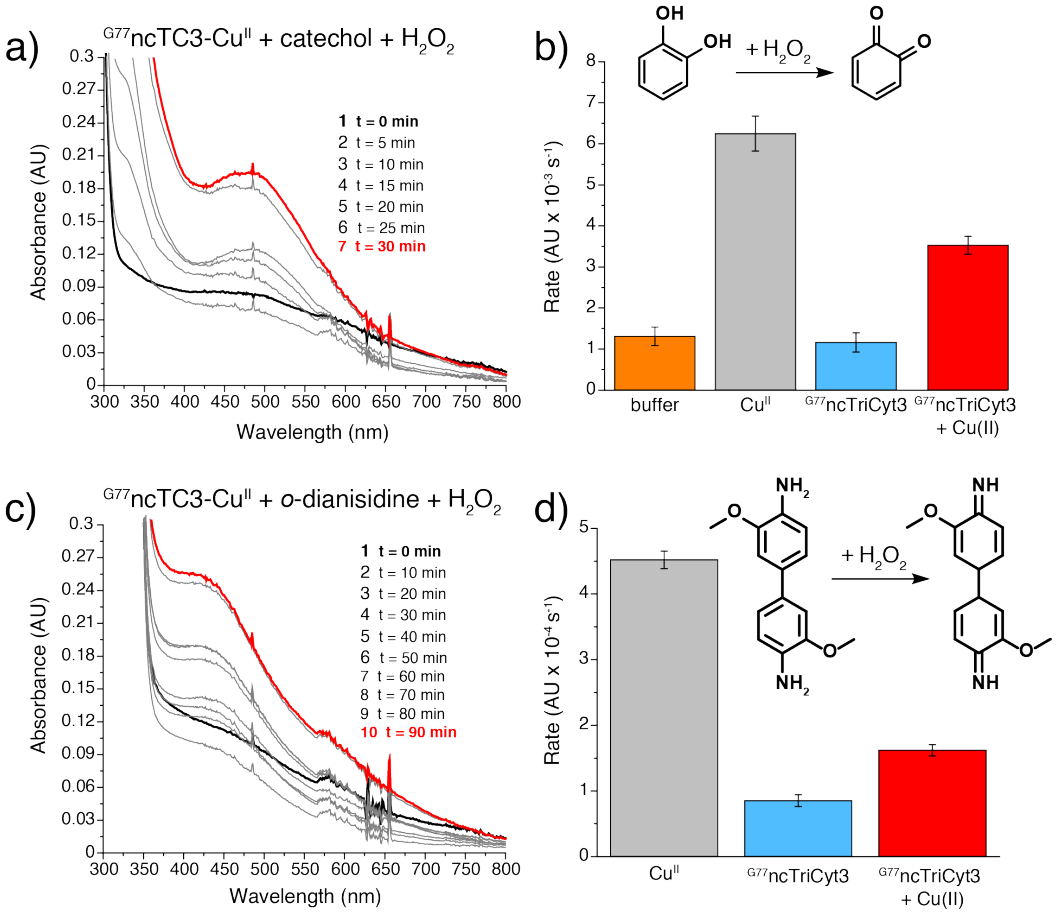

Figure 3.13 | Catechol and ODA oxidation activity of ^{G77}ncTriCyt3. (a) UV-vis spectroscopy following incubation of Cu^{II}-loaded ^{G77}ncTriCyt3 with 1 mM catechol and 10 mM H_2O_2. Spectra were recorded every 10 minutes over the course of 1 hour. (b) Summary of catechol oxidation rates under different reaction conditions. (c) UV-vis spectroscopy following incubation of Cu^{II}-loaded ^{G77}ncTriCyt3 with 1 mM ODA and 10 mM H_2O_2. Spectra were recorded every 10 minutes over the course of 90 minutes. (d) Summary of ODA oxidation rates under different reaction conditions. The oxidation rate in the absence of both protein and Cu^{II} was not measured.

dianisidine (ODA) as substrates, we observed a general correlation to the syringol oxidation data. whereby metal-loaded protein exhibited a higher peroxidase activity compared to apo protein but a lower activity compared to free Cu^{II} (**Figure 3.13 and Table 3.1**).

Our initial screens revealed that while Cu^{II}-bound ^{A77}ncTriCyt3 did not exhibit peroxidase activity for our most easily oxidizable substrate, ^{G77}ncTriCyt3 exhibited enhanced peroxidase activity for all three substrates of interest (**Figure 3.7a-c**) when pre-incubated with sub-stoichiometric equivalents of Cu^{II}. The initial peroxidase data obtained with the ^{G77}ncTriCyt3 scaffold motivated us to probe whether this emergent catalytic activity could be improved through directed evolution.

The reaction of interest for directed evolution experiments was ODA oxidation, as the difference in oxidation activity between apo and Cu^{II}-loaded ^{G77}ncTriCyt3 was lowest for the ODA substrate (**Figure 3.13d, Table 3.1**). Our library was generated through NYT codon randomization at residue positions 80 and 81 (**Figure 3.14a**), which results in 64 mutants with larger cavities compared to ^{G77}ncTriCyt3. These mutants also have a higher likelihood of engaging in CH-π or other non-covalent interactions with the phenyl rings of ODA than ^{G77}ncTriCyt3. Since the 80/81 mutations would eliminate the K80/E81 salt bridge, we installed a glutamate residue in position 78 to increase the probability of a compensatory, intermolecular salt bridge (**Figure 3.14a**).

Table 3.1 | Syringol, catechol, and ODA oxidation rates of ^{G77}ncTriCyt3[a]

Substrate	Rate (AU x s^{-1}) Buffer only	Rate (AU x s^{-1}) Protein only	Rate (AU x s^{-1}) Cu(II) only	Rate (AU x s^{-1}) Protein + Cu(II)
Syringol	1.3×10^{-3} (0.5)	3.0×10^{-3} (2)	2.0×10^{-2} (2)	9.1×10^{-3} (20)
Catechol	1.3×10^{-3} (2)	1.2×10^{-3} (2)	6.3×10^{-3} (4)	3.5×10^{-3} (2)
ODA	ND	8.5×10^{-5} (9)	4.5×10^{-4} (1)	1.6×10^{-4} (0.8)

[a] Values in parenthesis represent errors in the last digit reported, as calculated based on the standard deviations of replicate measurements.

Iterative rounds of site saturation mutagenesis were performed to ensure a random distribution of sequences (**Figure 3.14a**). All sequences included a media export tag, which allowed us to avoid labor-intensive cell lysis and thus maximize the protein yields of our high throughput screening protocol (**Figure 3.14b**). In this protocol, 192 colonies were each grown in 2 mL LB media to ensure an oversampling factor of 3 and then induced with 0.6 mM IPTG overnight. After a 16-hour induction period, the media was passed through centrifugal filters with a molecular weight cutoff (MWCO) of 100 kDa, then incubated overnight with 10 μM $CuSO_4$. After the addition of H_2O_2 and ODA, we observed for some samples a significant rate enhancement relative to free Cu^{II} ion, with improvements of up to 4-fold (**Figure 3.15a**). For these experiments, the oxidation rate was calculated by plotting the natural log of the concentration of product as a function of reaction time.

When sequencing the wells that yielded enhanced oxidation rates, only one sequence ($^{G77/E78/A80/L81}$ncTriCyt3) appeared twice despite the large oversampling factor (**Figure 3.15a**).

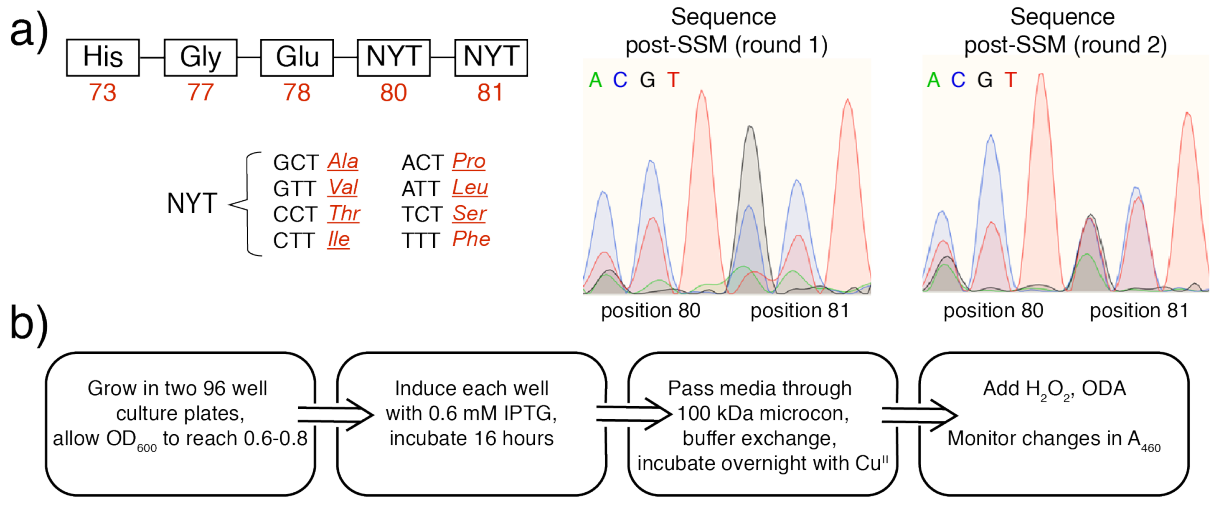

Figure 3.14 | Library generation and outline of high throughput screening protocol for improving ODA peroxidase activity of ^{G77}ncTriCyt3. (a) Sequence of TriCyt3 mutants and sequence chromatograms following two rounds of SSM on positions 80 and 81. Sequences from the first round of SSM served as templates for the second round. Oversampling factor = 3.

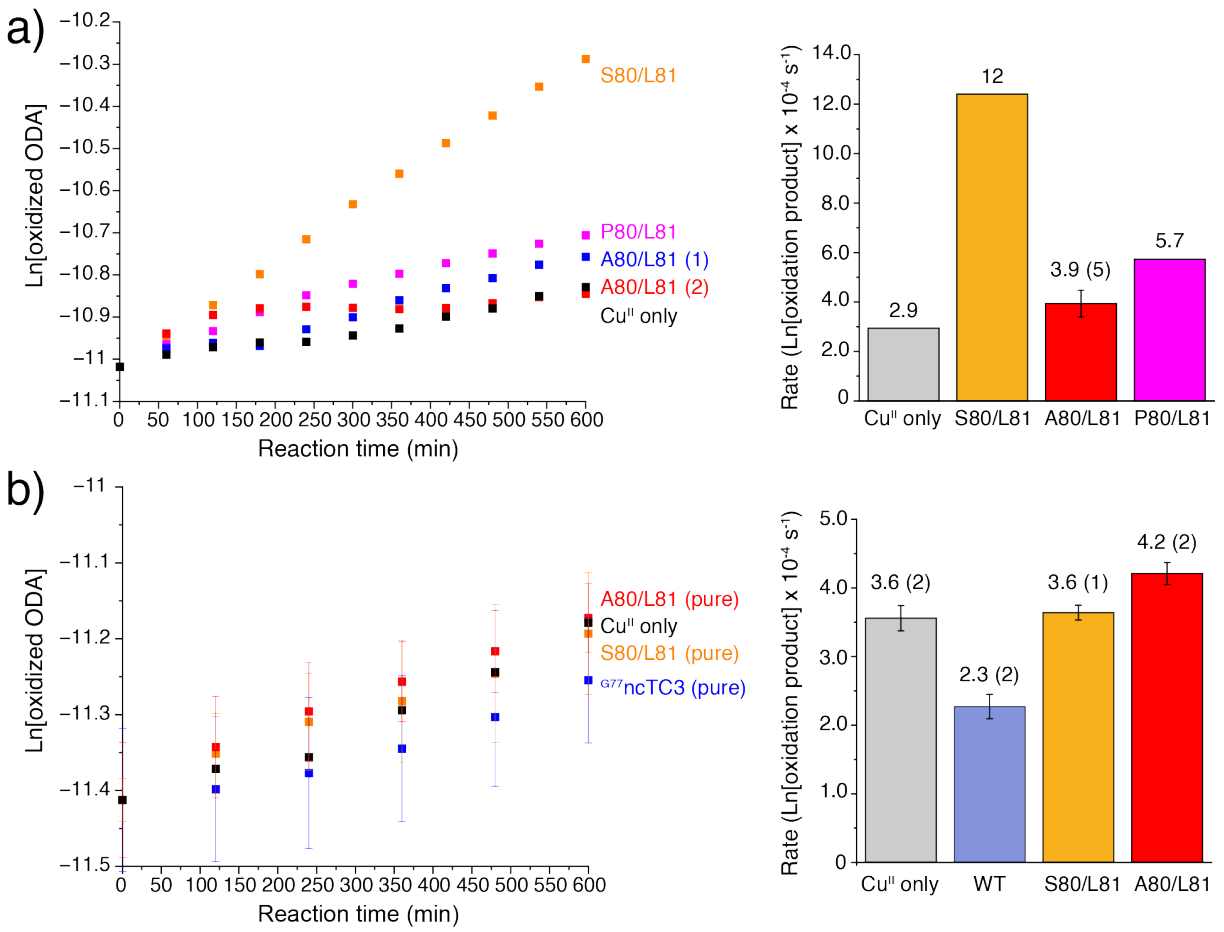

Figure 3.15 | ODA oxidation data for (a) Mutants directly characterized *in vivo* and (b) Mutants purified *in vitro*. For the *in vivo* experiments, the Cu^{II} only sample was prepared by adding $CuSO_4$ directly to LB media. ^{G77}ncTC3 is defined as WT. Notably, only one mutant characterized *in vivo* was sequenced more than once (GEAL). The concentration of oxidized ODA was determined based on ε_{460} = 11300 M^{-1} x cm^{-1}. Values in parenthesis represent errors in the last digit reported, as calculated based on the standard deviations of replicate measurements. Both *in vivo* and *in vitro* experiments were performed in the presence of 1 mM ODA and 10 mM H_2O_2.

Nevertheless, we set out to purify and characterize *in vitro* two unique mutants that were sequenced from the wells, $^{G77/E78/A80/L81}$ncTriCyt3 (GEAL) and $^{G77/E78/S80/L81}$ncTriCyt3 (GESL). When remeasuring ODA oxidation rates with *in vitro* purified protein, we observed significant deviation from the *in vivo* oxidation data (**Figure 3.15b**). Notably, the oxidation rate for the GESL mutant measured *in vivo* was over 3-fold higher than that determined *in vitro* (**Figure 3.15b**). For the GEAL mutant, the *in vitro* and *in vivo* rates were more closely correlated. The deviations between

the *in vivo* and *in vitro* rates may be due to the presence of endogenous peroxidase enzymes in the impure *in vivo* samples. In 40 mL scale media expressions of ^{G77}ncTriCyt3 and GEAL, the protein purity ranged from 42-52% (data not shown). Despite the inconsistencies between *in vivo* and *in vitro* oxidation rates, both GESL and GEAL mutants exhibited enhanced oxidation rates compared to ^{G77}ncTC3 (up to 1.8-fold higher) and an average oxidation rate either equal to or greater than that of free Cu^{II} ion. Encouraged by these results, we carried out SV-AUC measurements of the mutants in the absence and presence of 0.30 equiv. Cu^{II}/monomer to ensure that they assembled as trimers (**Figure 3.16**). For the GEAL mutant, the trimer yield was estimated to be 56% in the apo state and 72% upon the addition of Cu^{II}. By contrast, the trimer yields for the GESL mutant were 81% and >90% in the absence and presence of Cu^{II}, respectively. The SV-AUC data suggests that a significant fraction of the oxidation activity of Cu^{II}-loaded GEAL is attributable to free Cu^{II}, while the oxidation activity of GESL can be largely attributed to protein-bound Cu^{II}. Notably, the large differences in oligomerization yields are observed for mutants which differ by only one

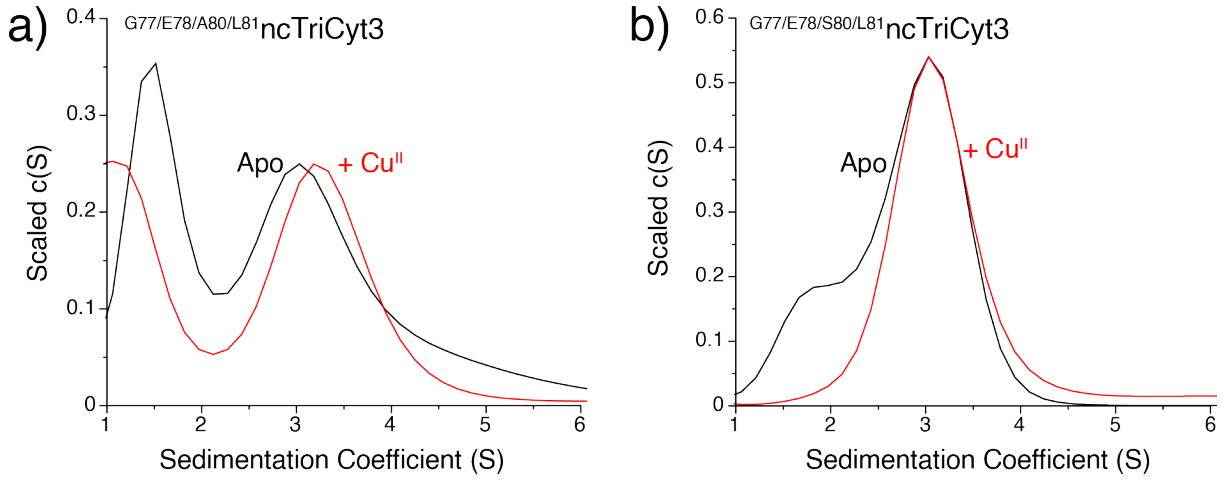

Figure 3.16 | SV-AUC distributions of (a) GEAL and (b) GESL mutants of TriCyt3. Measurements were performed with 60 μM protein monomer and 20 μM $CuSO_4$.

mutation, at residue position 80. We speculate that the A80 substitution, which results in a hydrophobic patch at the end of a helix, lowers protein fold stability and consequently the propensity to trimerize. We would need to perform circular dichroism (CD) spectroscopy measurements to test this hypothesis. Given its improved oxidation activity compared to ^{G77}ncTriCyt3 and propensity to trimerize both in the absence and presence of Cu^{II}, the GESL mutant of ncTriCyt3 represents a promising scaffold for future screening of our target catalytic functions (**Figure 3.7**).

Molecular zinc hydrides have been isolated using a variety of ligand frameworks[30-33] and are known to mediate hydride transfer to a broad range of substrates.[30] One ligand framework known to stabilize zinc hydrides is Tp (**Figure 3.6b**), which has been shown in its Zn^{II}-bound form to react with silanes to form a monomeric Zn^{II}-hydride species.[31] Ji et. al. recently recapitulated the catalytic activity of synthetic zinc hydride complexes in the protein scaffold human carbonic anhydrase II (hCAII).[23] The researchers demonstrated that in the presence of the hydride source phenylsilane, Zn^{II}-bound hCAII was able to catalyze the reduction of a broad range of ketones with yields and enantioselectivities as high as >99%.[23] This reaction was considered an attractive target for screening, as hCAII and ncTriCyt3 both feature mononuclear Zn^{II}:His_3 coordination sites and near-quantitative product formation was achieved without directed evolution of the active site of hCAII, suggesting less stringent secondary sphere coordination requirements. Notably, Ji et. al. observed that substitution of the active site residue Val121 to larger non-polar residues such as Phe and Trp abolished activity, demonstrating the impact of cavity size on zinc hydride formation and ketone reduction at the hCAII active site.[23] Given these results, we hypothesized that the requisite characteristics of the protein catalyst to achieve minimal activity could simply be its

ability to coordinate Zn^{II} in a tetrahedral, His_3 coordination environment and to accommodate the hydride donor/ketone substrate in the surrounding pocket.

To test this hypothesis, we aimed to screen the hydride transfer activity of Zn^{II}-bound ncTriCyt3 pocket mutants (**Figure 3.9**), with phenylsilane serving as the hydride donor and 4-acetylpyridine as the ketone substrate. Ketone reductions with phenylsilane were demonstrated by Ji et. al. to give the highest product yields, while acetyl pyridine is more water soluble and smaller than other potential substrates.[23] As the substrates and products do not have strong UV-vis absorbances, we chose to monitor product formation via ^1H NMR spectroscopy. We obtained both individual NMR spectra of reagents/products and an NMR spectrum of all components mixed at the stoichiometric ratios to be used for our catalysis screens (**Figure 3.17**). As these measurements were performed in 90% 50 mM MOPS (pH 8.0)/10% D_2O, a large signal corresponding to H_2O is present at a chemical shift of 4.7 parts per million (ppm), precluding full characterization and comparison of the NMR spectra. However, we identified a signal at ~8.38 ppm unique to the ketone reduction product (**Figure 3.17, inset**) that could serve as a spectroscopic handle for product formation during our initial catalysis screens. The first variant we tested was ^{A77}ncTriCyt3, which was predicted to have the smallest cavity size of all ncTriCyt3 pocket mutants (**Figure 3.** We first incubated 75 μM ^{A77}ncTriCyt3 monomer with 1.0-2.0 equiv. $ZnCl_2$ for 1 hour to ensure equilibration. We then added a large molar excess of hydride transfer reagents (2000 equiv. phenylsilane and 660 equiv. 4-acetyl pyridine) at room temperature and assayed for product

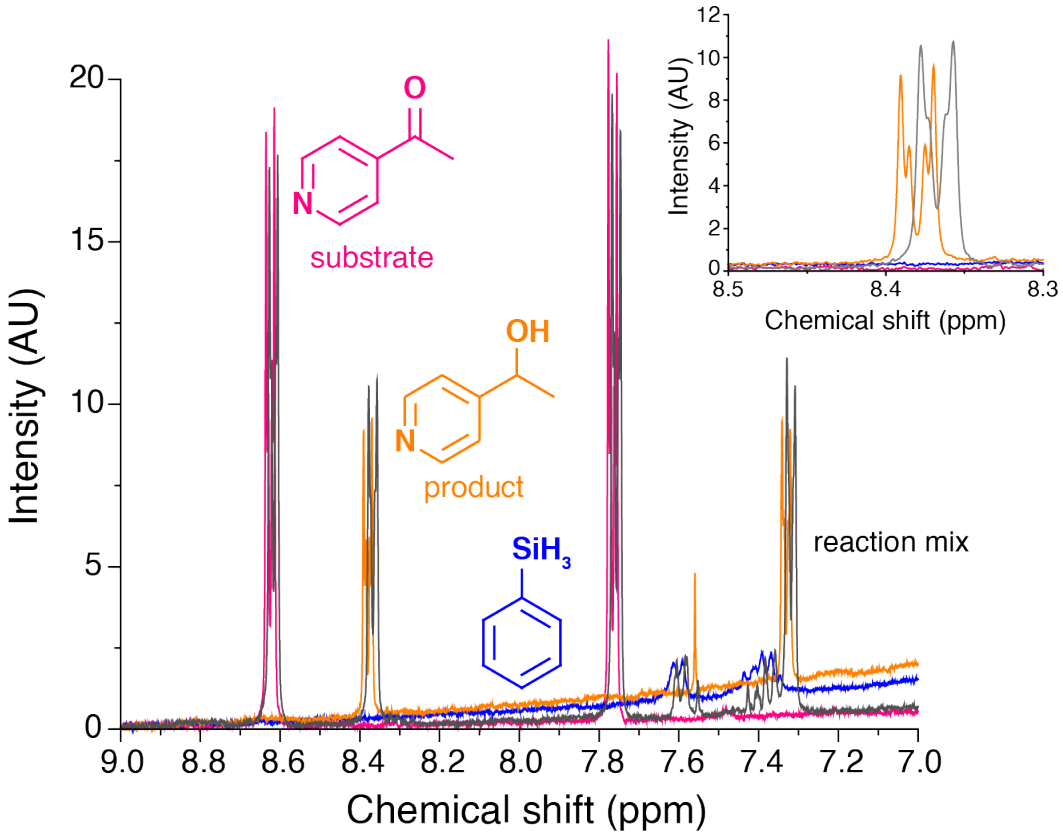

Figure 3.17 | Overlaid NMR spectra of 4-acetyl pyridine reduction reagents and products, including ketone substrate (magenta), product (orange), and phenylsilane (blue). The NMR spectrum of the component mixture (gray) features slight chemical shift differences compared to individual spectra of the components (inset). All samples were prepared in 50 mM MOPS (pH 8.0)/10% D_2O /10% D_2O. [Phenylsilane] = 36 mM, [substrate] = 12 mM, [product] = 12 mM.

formation via ^1H NMR (**Figure 3.18, red trace**). The signal at 8.38 ppm corresponding to the ketone reduction product was not present in the reaction mixture after an 18-hour incubation, suggesting that Zn^{II}-bound ^{A77}ncTriCyt3 was not active for the hydride reduction of 4-acetyl pyridine. However, we cannot rule out the possibility that the zinc-hydride species still forms at the coordination site. One experiment to establish whether this species forms would be to perform NMR analysis of the headspace gases of the reaction in the absence of ketone substrate to monitor the formation of H_2.[23]

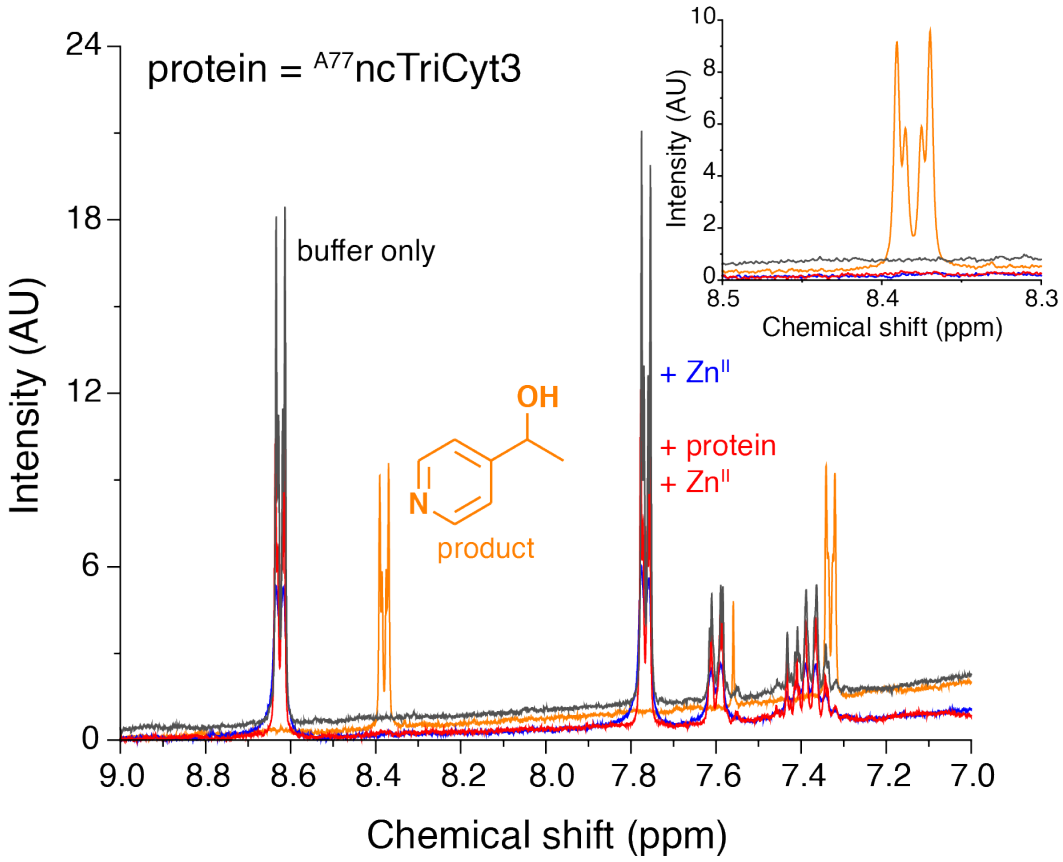

Figure 3.18 | Overlaid NMR spectra monitoring 4-acetyl pyridine reduction with no Zn^{II}/protein present (gray), with only Zn^{II} present (blue), and with Zn^{II} and protein present (red). An NMR spectrum of the product is shown in orange. All reactions were prepared in 50 mM MOPS (pH 8.0)/10% D_2O /10% D_2O. Phenylsilane and 4-acetylpyridine were present in reaction mixtures at 50 mM and 16.5 mM, respectively.

3.3.3 Incorporation of Fe_4S_4 clusters into TriCyt to obtain electronically coupled redox centers

Iron-sulfur (Fe-S) clusters are redox-active co-factors involved in a myriad of biological processes encompassing small molecule binding,[34, 35] electron transfer,[36-38] and catalysis.[39, 40] Commensurate with their diverse functional roles, protein-bound Fe-S clusters vary widely in their nuclearity, atomic composition, and redox potentials.[34] Fe-S clusters can be as compositionally simple as the [2Fe-2S] clusters of Rieske dioxygenases[34, 41] or as complex as the [8Fe-7S] and [Mo-7Fe-9S-C-homocitrate] clusters of nitrogenase,[39] while their redox potentials can span -600

to +450 mV.[34, 40] Given the functional importance of Fe-S proteins, there have been extensive efforts to both design such proteins *de novo*[42-45] and repurpose functional protein scaffolds into Fe-S cluster binding proteins.[46-48] Early *de novo* design efforts by the Dutton Lab involved the incorporation of [4Fe-4S] clusters into four-helix bundle peptides through the ligation of four protein-derived cysteines, rendering protein-cluster complexes reminiscent of ferredoxins.[42] Owing to the popularity of four-helix bundles as target folds and their ability to preorganize Cys_4 coordination environments, the vast majority of Fe-S proteins generated via *de novo* design recapitulate the coordination environment of ferredoxins.[43-45] While these *de novo* designed ferredoxin mimics have previously exhibited electron transfer activity,[44] each iron ion in ferredoxin-like [4Fe-4S] clusters is coordinatively saturated, precluding any possibility of engineering other functions such as small molecule binding and catalysis.

An alternative [4Fe-4S] cluster environment that could reasonably be stabilized in a designed protein is that found in aconitase, an enzyme which binds a single [4Fe-4S] cluster in a Cys_3 environment (**Figure 3.19**).[49, 50] In this coordination environment, the apical iron is coordinatively unsaturated and thus able to interact with citrate and catalyze its isomerization into isocitrate, a key step in the Krebs cycle (**Figure 3.19**).[49-52] When exposed to O_2, this apical iron "leaches" out of the cluster, resulting in a catalytically inactive [3Fe-4S] form (**Figure 3.19)**.

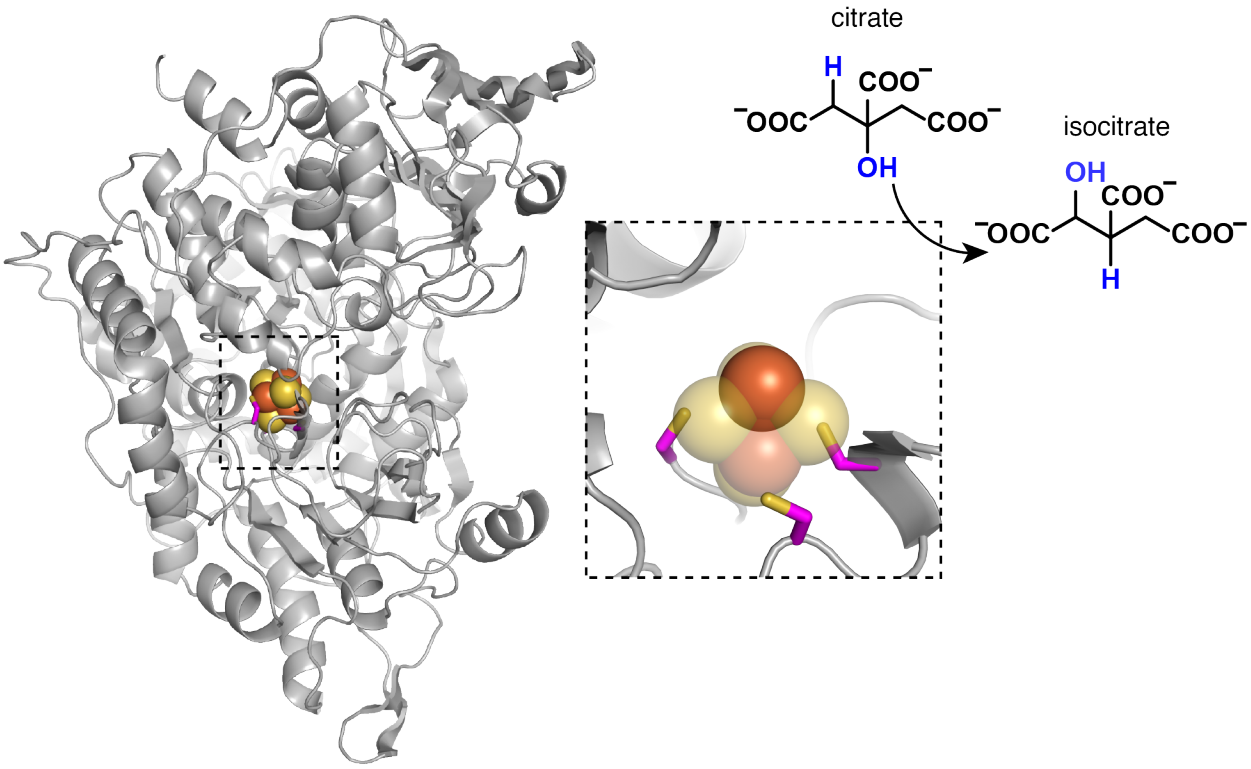

Figure 3.19 | Overview of aconitase structure (PDB ID: 1L5J) with close-up view of the [4Fe-4S] cluster and reaction scheme for the conversion of citrate to isocitrate. The cluster composition within aconitase can shift from [4Fe-4S] to [3Fe-4S]. In the [4Fe-4S] form, only three cysteine residues ligate the cluster, leaving one coordinatively unsaturated iron that functions as a Lewis acid to catalyze the isomerization of citrate to isocitrate. The catalytically active iron is shown as a solid sphere, while the rest of the cluster atoms are shown as transparent spheres.

Consequently, aconitase functions as both a Lewis acid catalyst and an O_2 sensor. Aussignargues and co-workers recently demonstrated that an aconitase-like site can be engineered into a bacterial microcompartment (BMC) shell protein through the installation of cysteine residues at a 3-fold interface.[46] The authors, while demonstrating successful incorporation of this cluster motif via solution methods and X-ray crystallography, did not indicate whether this cluster could bind exogenous ligands.[46] Notably, the three-fold symmetry axis at which the [4Fe-4S] cluster is placed in the BMC shell protein lacks an extended pocket comparable to that observed at the three-fold interfaces of the TriCyt scaffolds.[46] The extended pocket of our scaffolds (**Figure 3.9**) not only

provides the prospect of stabilizing ligand bound [Fe-S] clusters, but also to accommodate two distinct redox centers within the trimer: the exogenously added cluster and the endogenous M^{II}:His_3 site formed with His73 (**Figure 3.20**).

With the goal of placing an aconitase-like [4Fe-4S] cluster proximal to the M^{II}:His_3 site, we modeled cysteine mutations at residue position 77 for Fe^{II}:$(TriCyt2)_3$ (PDB ID: 6WZ0) and

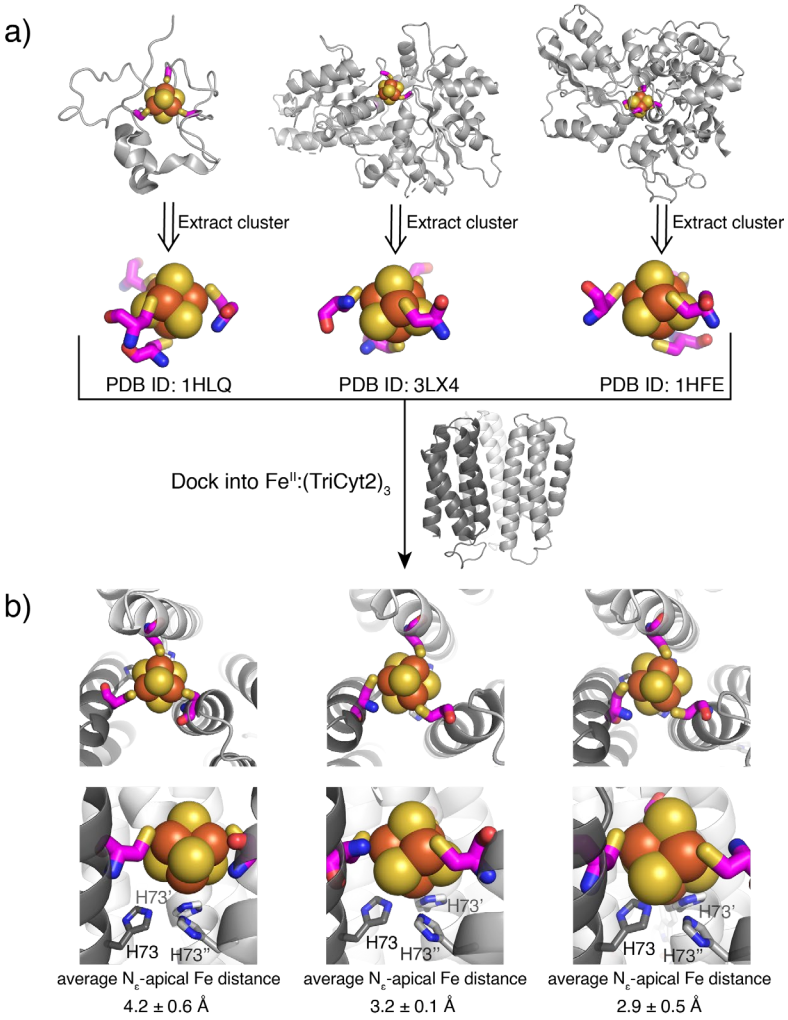

Figure 3.20 | Docking of [4Fe-4S] clusters from natural metalloproteins into Fe^{II}:$(TriCyt2)_3$. (a) Select Fe-S metalloproteins from which clusters were extracted and docked into Fe^{II}:$(TriCyt2)_3$. All docked clusters feature Cys_4 ligation. All bond angles and metal-ligand distances associated with the $Fe_4S_4Cys_4$ clusters were fixed during docking. (b) Topviews (top) and sideviews (bottom) of [4Fe-4S]-bound TriCyt2 models. The Fe^{II} ion of the Fe^{II}:His_6 site and His77 are omitted for clarity.

Ni^{II}:(TriCyt3)$_3$ (PDB ID: 6WZC). The average distance between the sulfur atoms of Cys77 was 6.2 Å for Fe^{II}:(TriCyt2)$_3$ and 7.5 Å for Ni^{II}:(TriCyt3)$_3$, with the intersulphur distance of the former more closely correlating to that of aconitase (6.4 ± 0.3 Å, PDB ID: 1L5J). We thus chose Fe^{II}:(TriCyt2)$_3$ as a model for [4Fe-4S] cluster docking calculations (**Figure 3.20a**). In these calculations, [4Fe-4S] clusters were extracted from protein crystal structures in the PDB and superimposed onto the three-fold interface of the trimer (**Figure 3.20a**). In addition to residue position 77, clusters were docked in residue positions 52, 54, 59, 62, 66, and 70. Ultimately, these positions were not selected for mutagenesis, as the docking simulations revealed that they were not as compatible for cluster binding as position 77 (data not shown). Cluster-docked structures varied significantly in the proximity between the apical iron of the cluster and the His73 coordination site (**Figure 3.20b**), an indication of the geometric variability of natural, protein-associated [4Fe-4S]. After confirming that residue position 77 was the optimal location to install cysteine residues to ligate the cluster, we mutated ncTriCyt2 into ^{C77}ncTriCyt2. Given that UV-vis features of both heme prosthetic groups and Fe-S clusters are present in the 300-600 nm range, it was imperative to use a heme-less version of the TriCyt2 scaffold to ensure that we could assess cluster incorporation via UV-vis and other spectroscopies.

After expressing and purifying our new mutant, we had to select a strategy for cluster incorporation into our trimeric scaffold (**Table 3.2**). Previous strategies for forming Fe-S proteins include *in-vitro* reconstitution,[44, 46] *in-vivo* biogenesis,[45] and incorporation of synthetic small-molecule [4Fe-4S] clusters (**Table 3.2**).[48, 53] Of these methods, *in-vitro* reconstitution is operationally the simplest. In this method, protein is mixed with an Fe^{3+} source (often $FeCl_3$), a sulfide source (Na_2S, Li_2S), and small-molecule thiol ligands such as dithiothreitol (DTT) and beta-mercaptoethanol (β-ME) under an anaerobic atmosphere.[54, 55] While it is hypothesized that

Table 3.2 | Cluster incorporation strategies

Method	Advantages	Disadvantages	References
In-vitro reconstitution	Operationally simple	Yield of *in-situ* generated cluster unknown	[43], [45]
In-vivo biogenesis	High incorporation fidelity/efficiency	Incorporation mechanism not well-established	[44]
Synthetic incorporation	High incorporation fidelity/efficiency	Synthesis requires dry, air-free conditions	[42], [47]

incorporation occurs via *in-situ* formation of a small-molecule DTT/β-ME-ligated cluster followed by substitution onto the Cys-containing protein scaffold,[54, 55] to the best of our knowledge there is no evidence that this small molecule cluster is generated in high yield *in situ*, if at all. Therefore, unless the protein scaffold already has some fraction of endogenously incorporated cluster to effectively "preorganize" reconstitution, the yield of [4Fe-4S] incorporation into the protein scaffold is limited by the yield of DTT/β-ME-ligated [4Fe-4S] cluster produced *in situ*. A second method, recently adopted by the Nanda Group for the design of Fe-S proteins,[45] is to exploit bacterial mechanisms for cluster incorporation into natural Fe-S proteins, namely the Isc and Suf pathways of *E.coli*.[56] This approach benefits from the highly efficient cellular machinery responsible for loading Fe-S proteins with their cognate co-factors, and thus has the potential to produce [4Fe-4S]-loaded proteins in higher yield compared to *in-vitro* reconstitution approaches. However, the mechanism by which the cellular machinery designates proteins for cluster installation is not well understood, potentially limiting the generalizability of this *in vivo* route to Fe-S proteins. A third option for cluster incorporation is to mix protein scaffolds with discrete, synthetic [4Fe-4S] clusters (**Figure 3.21**). This approach has been reported in only a few instances,[43, 48, 53] yet represents the most efficient path to incorporate clusters in high yield and with few side products. Moreover, the compositional diversity of synthetic Fe-S clusters[57, 58]

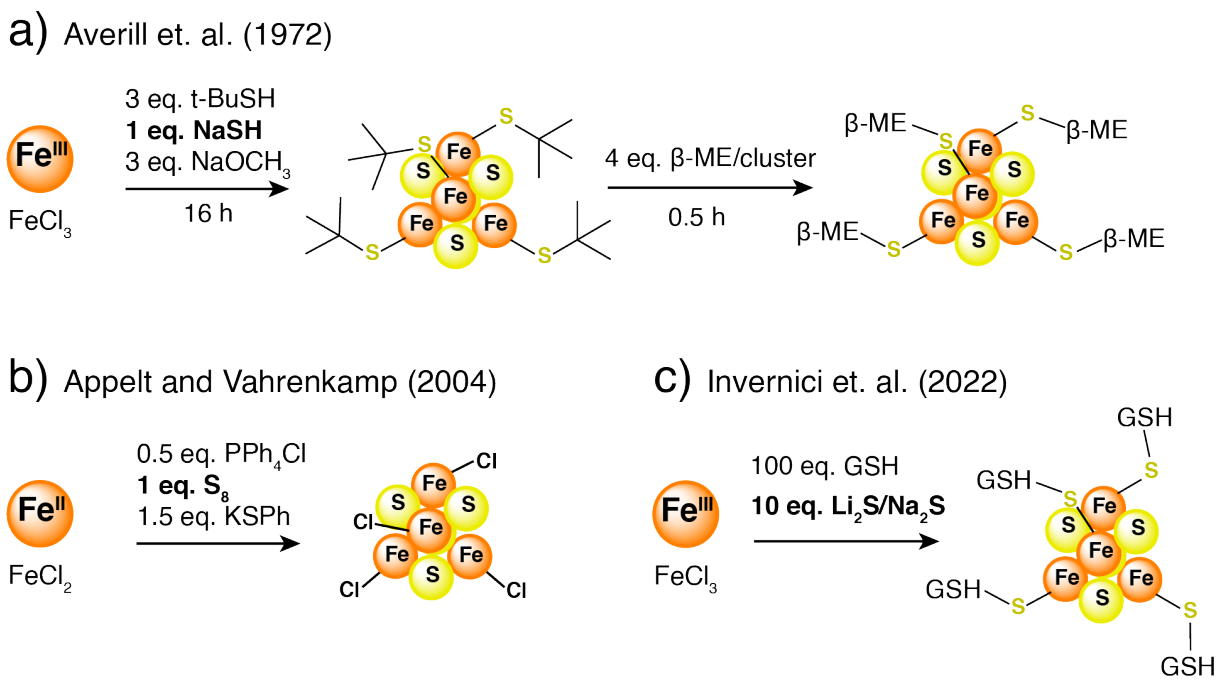

Figure 3.21 | Synthetic routes to [4Fe-4S] clusters. (a) Synthetic route developed by the Holm Group. It was later demonstrated that β-ME could substitute for the tert-butyl thiol substituents of this cluster. (b) Synthesis developed by the Vahrenkamp Group which results in a chloride-substituted cluster. Unlike the protocol developed by the Holm Group, which utilizes Fe^{III} and NaSH as the sulfide source, this synthesis utilizes Fe^{II} and elemental sulfur (S_8) as the sulfide source.

provides researchers the opportunity to form Fe-S proteins with greater structural and functional complexity than [4Fe-4S]-bound proteins.

Averill et. al. (**Figure 3.21a**) and Appelt and Vahrenkamp (**Figure 3.21b**) have reported two distinct procedures for [4Fe-4S] cluster synthesis in organic solvent, with the method of Averill et. al. rendering a thiol-substituted cluster and that of Appelt and Vahrenkamp resulting in a chloride substituted cluster. These clusters are insoluble in aqueous solution in the absence of thiol-based ligands such as β-ME, and even with excess β-ME one could infer based on previous studies of peptide-substituted Fe-S clusters that the β-ME-ligated cluster would degrade within hours.[59] However, when mixing the cluster with protein it is expected that the cysteine residues of the protein should readily substitute for the β-ME ligands and solubilize the cluster before a

significant fraction of it degrades. With regards to cluster incorporation, the main barrier presented by the above approaches is synthetic experience, as both syntheses must be carried out in stringently dry and air-free conditions. Recently, Invernici et. al. reported on a synthesis of glutathione-substituted [4Fe-4S] clusters (**Figure 3.21c**) that only requires mixing GSH, a sulfide source, and $FeCl_3$ in alkaline solution.[60] While the synthetic procedure is simple, 400 equiv. GSH/cluster are required to stabilize the cluster in aqueous solution. Given the large excess of GSH relative to our protein scaffolds that would be required during incorporation and the fact that GSH utilizes cysteine residues to ligate Fe-S clusters, we surmised that GSH would be a less labile cluster ligand than other small molecule thiols, leading to slower transfer to ^{C77}ncTriCyt2 (**Figure 3.21a-b**). Given the potentially slow incorporation kinetics posed by a GSH-ligated cluster, we opted to use the cluster synthesized by Appelt and Vahrenkamp as our [4Fe-4S] source.

Synthetic [4Fe-4S] cluster (hereafter referred to as $Fe_4S_4^{syn}$) was generously provided by the Suess Group at MIT. We considered two different incorporation protocols. In the first, a 60 mM stock of $Fe_4S_4^{syn}$ in DMF would be added dropwise to 3 mM ^{C77}ncTriCyt2 monomer to give 1-3 mM cluster, or 1-3 equiv. cluster/trimer. In the second, 300 μM ^{C77}ncTriCyt2 would be initially mixed with 1 equiv. $CuCl_2$, and then, following a one-hour incubation at room temperature, would be mixed with 60 mM $Fe_4S_4^{syn}$ such that 1-3 equiv. cluster/trimer is present in solution. The purpose of the second incorporation protocol would be to assess whether the protein scaffold could simultaneously form the M^{II}:His_3 site while coordinating $Fe_4S_4^{syn}$ at the Cys_3 site (**Figure 3.20**). Both reactions would be carried out in a buffer containing 50 mM TRIS and 25 mM β-ME (pH 8.5), under anaerobic conditions, at room temperature, and for up to 20 hours. As a qualitative assessment of whether cluster could be incorporated into ^{C77}ncTriCyt2, we compared color changes in a reaction based on the first incorporation protocol with those of a solution with only

$Fe_4S_4^{syn}$, TRIS, and β-ME present. Over the course of six hours, the color of the (-) protein solution changed from yellow to clear while a significant quantity of black precipitate formed (**Figure 3.22a**). While some black precipitate was also observed in the (+) protein solution, the solution remained yellowish brown, suggesting that ^{C77}ncTriCyt2 bound some fraction of the cluster molecules present and prevented their degradation. After 20 hours, the reaction was passed over a PD-10 column equilibrated in 50 mM TRIS, 25 mM β-ME (pH 8.5) to remove unreacted cluster. As a negative control, we added free $Fe_4S_4^{syn}$ to the column and carried out the gel filtration step under the same conditions performed for the {protein + cluster} reaction. After eluting in 0.7 mL buffer, UV-vis spectroscopic features suggestive of Fe-S species were observed for the {protein + cluster} reaction but not for $Fe_4S_4^{syn}$ (**Figure 3.22b**), validating gel filtration as an effective method for separating unbound and protein-bound $Fe_4S_4^{syn}$. The UV-vis spectrum of ^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$ featured a broad absorption band with a λ_{max} of 426 nm (**Figure 3.22b-c**), which is red-

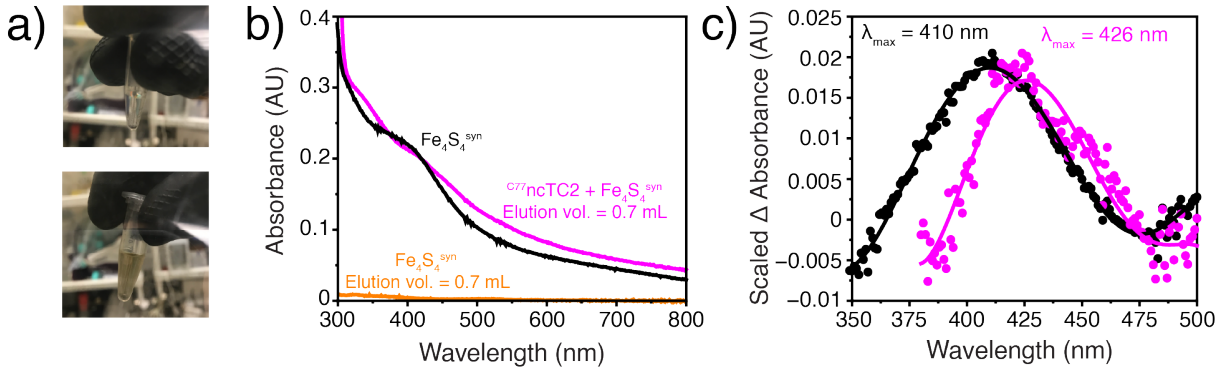

Figure 3.22 | Initial observations and UV-vis characterization of {^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$}. (a) $Fe_4S_4^{syn}$ in the absence (top) and presence (bottom) of ^{C77}ncTriCyt2 following a six-hour incubation in 50 mM TRIS, 25 mM β-ME. In both reactions, [$Fe_4S_4^{syn}$] = 3 mM. (b) UV-vis spectra of protein + $Fe_4S_4^{syn}$ (magenta trace) and $Fe_4S_4^{syn}$ (orange trace) following elution from a PD-10 desalting column. At an elution volume of 0.7 mL, no visible spectroscopic features of free $Fe_4S_4^{syn}$ are observed, indicating that free cluster is retained on the column while protein-cluster adducts elute. (c) UV-vis spectra of $Fe_4S_4^{syn}$ (black trace) and protein + $Fe_4S_4^{syn}$ (magenta trace) following a linear baseline correction, indicating distinct λ_{max} values for the Cys⁻Fe LMCT transition in the 400-450 nm range. The red shift in λ_{max} has been previously observed when mixing protein scaffolds with synthetic Fe_4S_4 clusters. [47].

shifted relative to the UV-vis spectrum of $Fe_4S_4^{syn}$ (λ_{max} = 410 nm). The red shift in λ_{max} ($\Delta\lambda_{max}$ = +16 nm) upon incubation of the cluster with protein suggests that, assuming $Fe_4S_4^{syn}$ is bound to ^{C77}ncTriCyt2, the coordination environment around $Fe_4S_4^{syn}$ is distinct from that of the unbound cluster.[48]

The UV-vis spectrum of {protein + cluster} revealed putative LMCT bands at 350 and 426 nm suggestive of Cys77-iron bonding. Assuming that all protein-bound iron was associated with $Fe_4S_4^{syn}$, we aimed to quantitate the yield of cluster incorporation by quantitating iron content with a 2,2'-bipyridine (bipy) chelation assay (**Figure 3.23**). Bipy molecules bind Fe^{II} in an octahedral Fe^{II}-$(bipy)_3$ complex with a metal-to-ligand charge transfer (MLCT) band centered at 522 nm (ε_{522} = 8650 M^{-1} x cm^{-1}).[61, 62] In the assay, {protein + cluster} is mixed with a "bipy cocktail" (**Figure 3.23a**) composed of 7 M guanidine HCl (GdnHCl), 10 mM DT, and 2 mM bipy.

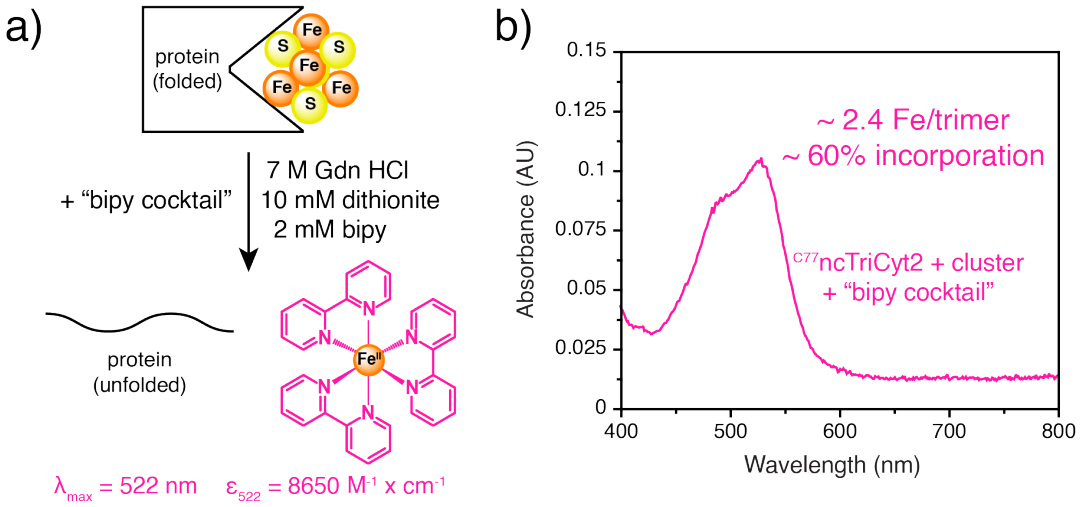

Figure 3.23 | Bipy chelation assay to quantitate [Fe] in {protein + cluster} reaction. (a) Schematic of the bipy chelation assay. All additions were made in an anaerobic Coy chamber. (b) UV-vis spectrum of {protein + cluster} following incubation with the "bipy cocktail" for at least 15 minutes. Correction factors were applied to the A_{522} values to account for the intrinsic absorbances of the bipy cocktail and {protein + cluster}.

The iron content was estimated based on the [Fe^{II}-(bipy)$_3$] determined by A_{522}. The purpose of including Gdn HCl and DT was to unfold the protein and ensure all iron present is reduced to Fe^{II}, respectively. In parallel to UV-vis measurements on {protein + cluster} mixed with the bipy cocktail, we measured the A_{280} of {protein + cluster} to determine [protein], with the assumption that absorbance features of the cluster contribute minimally to the total absorbance at 280 nm.[44] Based on the ratio A_{522}/A_{280}, we estimated 2.4 Fe atoms/protein trimer, or, assuming that all Fe present is cluster-associated, ~60% $Fe_4S_4^{syn}$ incorporation efficiency (**Figure 3.23b**). Notably, this incorporation efficiency is significantly lower than that previously reported for $Fe_4S_4^{syn}$ incorporation into a designed protein , which was nearly quantitative.[48] In this report, the authors unfolded the protein prior to the reaction with $Fe_4S_4^{syn}$. Such an approach, though likely to lower potential kinetic barriers to cysteinyl coordination of the cluster, would require extensive screening to identify conditions for refolding TriCyt variants.

Having estimated the iron content of {protein + cluster}, we next aimed to characterize the electronic states of the iron atoms via electron paramagnetic resonance (EPR) spectroscopy. In the absence of exogenous reductants or oxidants, $Fe_4S_4^{syn}$ is expected to be EPR silent due to the antiferromagnetic coupling of two [Fe^{2+}, Fe^{3+}] centers with S = $\frac{9}{2}$.[63] To our surprise, {^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$} in the absence of reductants/oxidants was EPR-active, with the EPR data featuring an isotropic signal at g ~ 2.01 (**Figure 3.24a, top**). This feature has previously indicated a Fe_3S_4 cluster in the oxidized 1+ state with S = $\frac{1}{2}$.[64, 65] Previous studies of protein-bound Fe_3S_4 clusters have shown that the $[Fe_3S_4]^{1+}$species can be reduced to the EPR-silent (S = 2) $[Fe_3S_4]^0$ with DT.[65, 66] Upon the addition of 10 mM DT to {protein + cluster}, the g ~ 2.01 signal remained, though the overall signal from g = 1.98-2.03 was reduced ~ 1.8-fold (**Figure 3.24a, middle**). If the g ~ 2.01

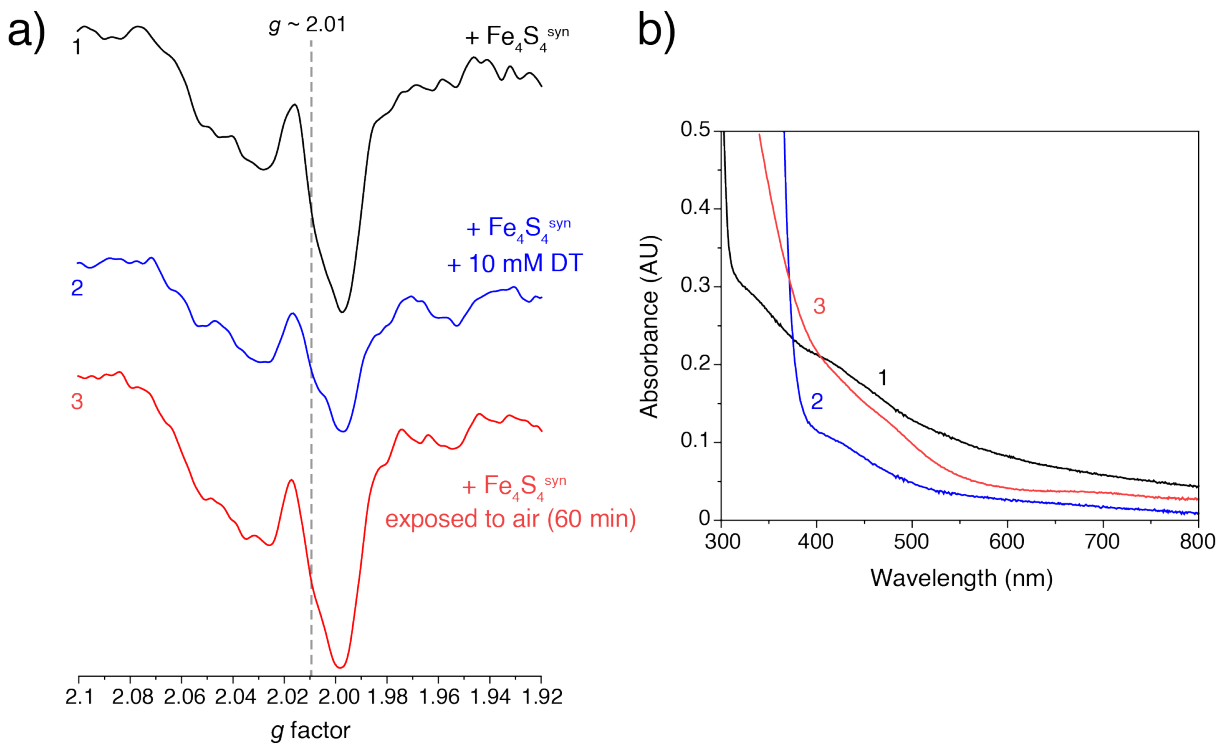

Figure 3.24 | EPR and UV-vis characterization of {^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$} in the presence of redox stimuli. (a) X-band EPR of {protein + cluster} in the absence of redox stimuli (top), in the presence of 10 mM DT (middle), and after prolonged exposure to air (bottom). (b) Corresponding UV-vis spectra of EPR samples. The $g \sim 2.01$ signal is present in all samples despite UV-vis spectra suggesting a change in the electronic states of protein-bound Fe species. EPR parameters: temperature = 40K, microwave frequency = 9.36 GHz, modulation amplitude = 9.77 G, microwave power = 0.2 mW. All EPR samples were transferred to sealed tubes in an anaerobic Coy chamber and then flash frozen in liquid N_2 prior to measurements. All samples were prepared with 10% (v/v) glycerol as a glassing agent.

feature indeed represents a $[Fe_3S_4]^{1+}$ species, then it would have an unusually low redox potential such that the addition of excess dithionite does not fully yield the EPR-silent $[Fe_3S_4]^0$ species.

Notably, the UV-vis spectrum of {protein + cluster} in the presence of DT shows a ~ 2-fold decrease in ε_{426}, suggesting a change in the electronic states of the protein-bound Fe species (**Figure 3.24b, blue trace**). The same discrepancy between the UV-vis and EPR spectroscopic data is observed upon extended exposure of {protein + cluster} to air. After over 60 minutes of exposure, the UV-vis spectrum reveals that the characteristic shoulder centered at 426 nm and attributed to a Cys-Fe LMCT band is no longer present (**Figure 3.24b, red trace**). However, the

EPR spectrum still features the $g \sim 2.01$ signal, with little difference in the signal intensity relative to the spectrum of {protein + cluster} in the absence of stimuli. Given these discrepancies, the observation that dithionite addition does not generate an EPR silent species, and the lack of Mössbauer spectroscopy data, we could not say with confidence based on our initial solution data that the iron species bound to the protein over the course of the {protein + $Fe_4S_4^{syn}$) reaction is $[Fe_3S_4]^{1+}$.

In parallel to solution characterization via EPR and UV-vis spectroscopies, we aimed to characterize the composition and coordination environment of potential protein-bound clusters through X-ray crystallography. We observed a mixture of clear and yellow-brown crystals in conditions with either Poly(ethylene glycol) (PEG) 1500 or PEG 400 (**Table 3.3**). Yellow-brown crystals were brought outside the tent, incubated for < 30 seconds in cryoprotectant (**Table 3.3**), then flash frozen in liquid N_2. We observed that while this approach did temporarily expose

Table 3.3 | Crystallization conditions for {^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$}

Condition	Appearance	Cryoprotectant	Space Group/Unit Cell /Resolution
25% PEG400, 200 mM $CaCl_2$, 100 mM Bis-Tris (pH 6.5)	Yellow-brown Plates/needles/cubes	25% glycerol, 12.5% PEG400, 100 mM $CaCl_2$, 50 mM Bis-Tris (pH 6.5)	NA (twinned)
25% PEG1500, 200 mM $MgCl_2$, 100 mM Bis-Tris (pH 5.5)	Yellow-brown Plates	25% glycerol, 12.5% PEG1500, 100 mM $MgCl_2$, 50 mM Bis-Tris (pH 5.5)	NA (twinned)
25% PEG1500, 200 mM NaCl, 100 mM HEPES (pH 7.5)	Yellow-brown Diamond-shaped	25% glycerol, 12.5% PEG1500, 100 mM NaCl, 50 mM HEPES (pH 7.5)	$P\,3_2$ a = 43.24, b = 43.24, c = 265.2, α = 90, β = 90, γ = 120 2.5 Å
25% PEG1500, 200 mM NaCl, 100 mM Tris pH 8.5	Yellow-brown Triangular prisms	25% glycerol, 12.5% PEG1500, 100 mM NaCl, 50 mM Tris (pH 6.5)	NA (twinned)

crystals to air, the crystals maintained their yellow-brown color, suggesting qualitatively that oxygen damage did not occur. Notably, exposing the crystals to air for at least 30 minutes led to a loss of color. Prior to data collection, wavelength scans around the K-edge of Fe were performed. A large increase in X-ray absorption around the Fe K-edge (**Figure 3.25a**) was observed, suggesting that Fe was bound to the protein. While twinning was observed for most crystals (**Table 3.3**), we were able to determine the structure for one by solving it in the *P* 3_2 space group (**Table 3.3**) at a resolution of 2.5 Å, with two trimers (termed trimer1 and trimer2, **Figure 3.25b-c**) comprising the asymmetric unit (ASU). To identify Fe binding sites within the ASU, we collected diffraction data above (E = 7200 eV) and below (E = 7000 eV) the Fe K-edge (**Figure 3.25b-c**). At E = 7200 eV, discrete anomalous density that can be modeled with two Fe atoms, one proximal to the Cys77 layer and another to the His73 layer, is observed in both trimers of the ASU (**Figure 3.25b-c**). However, density corresponding to an Fe atom at the His73 layer is much weaker in trimer2 than in trimer1 (**Figure 3.25c**), suggesting lower occupancy. Indeed, the occupancy of Fe

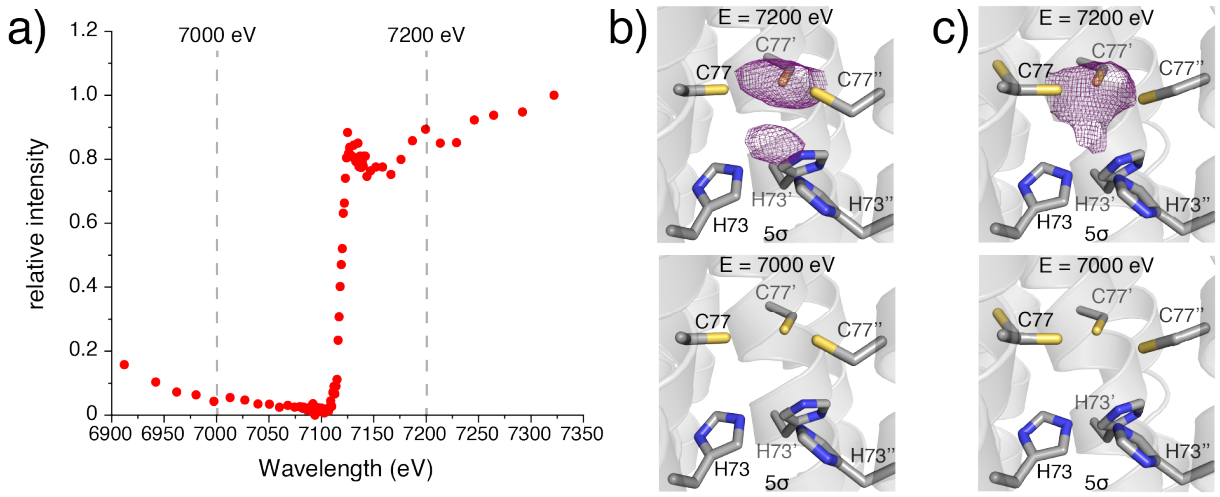

Figure 3.25 | Characterization of Fe binding in {^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$}. (a) Wavelength scan from 6900-7350 eV, illustrating a large increase in X-ray absorption around the K-edge of Fe. Based on this scan, diffraction data was collected at 7200 and 7000 eV. (b-c) Anomalous difference maps at 7200 (above) and 7000 eV (below) for trimer1 and trimer2 of the ASU. The large difference in anomalous density at wavelengths just above and below the Fe K-edge points to the presence of Fe at the three-fold interface.

at the His73 site of trimer2 was ultimately refined to 0.51, while the latter three Fe atoms in the ASU were refined to full occupancy (**Figure 3.26**).

The lack of discrete anomalous density assignable to four individual Fe atoms at the Cys77 layers, coupled with the uncharacteristically short S_γ distances (4.9 ± 1.0 Å, compared to 6.3 ± 0.3 Å for aconitase) between Cys77 residues, led us to conclude that Fe-S clusters were not bound to the protein. The crystallographic data instead suggests the presence of two unusual di-Fe sites at the three-fold interfaces of each trimer of the ASU, each with distinct coordination numbers and geometries (**Figure 3.26**). Given the highly diffuse nature of the anomalous difference maps and the quality of the dataset, we are not fully confident that our placement of the Fe atoms and calculation of Fe-ligand distances is accurate, especially for trimer2. At the trimer1 interface, Fe1 is coordinated by two Cys77 residues and an aqua ligand in a distorted trigonal planar geometry, while Fe2 is coordinated by the N_ε atoms of three His73 residues, again in a distorted trigonal planar geometry (**Figure 3.26a**). While in other crystal structures of TriCyt coordinatively unsaturated Fe:His_3 complexes include an aqua ligand and are thus tetrahedral, we speculate that in this case Fe1 sterically occludes coordination of an aqua ligand to Fe2. The coordination environment around Fe2 is similar at the interface of trimer2 (**Figure 3.26b**). However, the estimated Fe1-Fe2 distance (2.5 Å) at the trimer2 interface is 0.7 Å shorter than at the trimer1 interface and suggestive of metal-metal bonding. Notably, none of the Cys77 sulfur atoms are within coordinating distance of Fe1 (**Figure 3.26b**). Given that Fe1 is at full occupancy, it is likely that our placement of the Fe1 atom is incorrect. However, to limit bias we modeled all Fe atoms at the center of the anomalous densities, and so an alternative position would not reflect the best placement based on the initial anomalous difference maps. A higher resolution crystal structure is

expected to provide more clarity about the coordination environments of the Fe atoms in ^{C77}ncTriCyt2.

While the di-Fe site present in trimer1 of the ASU was an unexpected finding, we were intrigued by its coordination environment, which has no known precedence in Nature. We aimed to establish whether this di-Fe site could be isolated in solution and the scope of metal ions that can form a similar coordination site. Although the inter-S_γ distances in the crystal structure (4.9

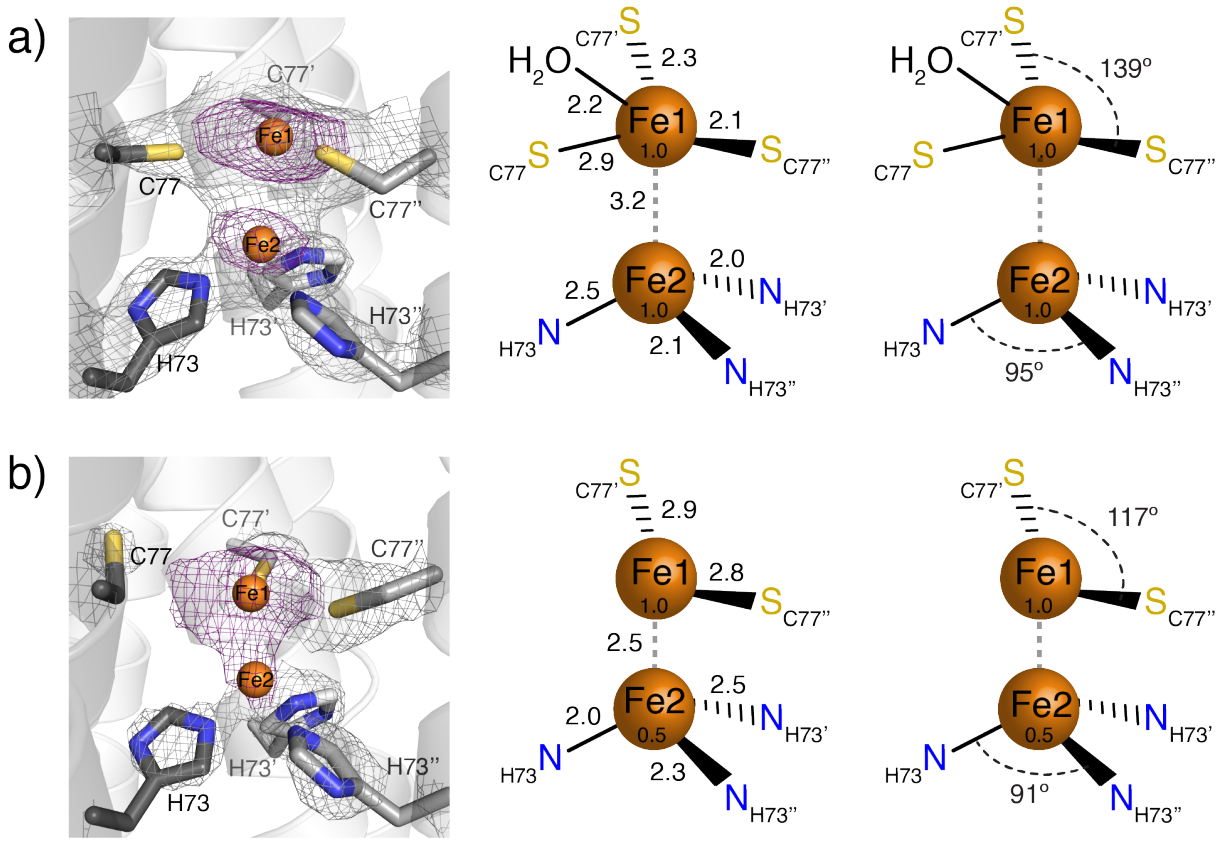

Figure 3.26 | Electron density maps around His73/Cys77 residues of {^{C77}ncTC2 + $Fe_4S_4^{syn}$} crystal structure and schemes illustrating salient geometric features of the di-Fe coordination sites. (a) Superimposition of anomalous difference (magenta) and 2mFo-Dfc (gray) maps at trimer1 interface (left), bond distances for the interfacial Fe-ligand complexes (middle), select bond angles for the Fe-ligand complexes (right). (b) Superimposition of anomalous difference (magenta) and 2mFo-Dfc (gray) maps at trimer2 interface (left), bond distances for the interfacial Fe-ligand complexes (middle), select bond angles for the Fe-ligand complexes (right). The evidence for a di-Fe site is stronger at the trimer1 interface than at the trimer2 interface. Anomalous difference maps were contoured at 5σ and 2mFo-Dfc maps at 2σ.

Å) were ~1.3-fold smaller than those predicted by the PYMOL model, the metal coordination metrics at trimer2 (**Figure 3.26b**) suggest that a scaffold in which the cysteine residues may approach closer to each other would be better suited for isolating dinuclear complexes for a wide range of metal ions. One scaffold of interest is the chemically-inducible trimer (CIT), based on the TriCyt3 scaffold (**Figure 3.27a**, CIT = $^{I67/A73/M76/N80/H81/L83}$ncTriCyt3). While originally designed to form a ligand-directed assembly, it was found instead to trimerize in the absence of stimuli. Modeling Cys77 mutations into CIT yielded an average inter-S_γ distance of 5.4 Å (**Figure 3.27b**), 0.8 Å smaller than the average distance observed when modeling Cys77 into ncTriCyt2.

We hypothesized based on the PYMOL model of ^{C77}CIT that the cysteine layer in this scaffold could more closely approach the 3-fold axis than in ^{C77}ncTriCyt2 and thus be in position to coordinate metal ions encompassing a larger range of ionic radii.

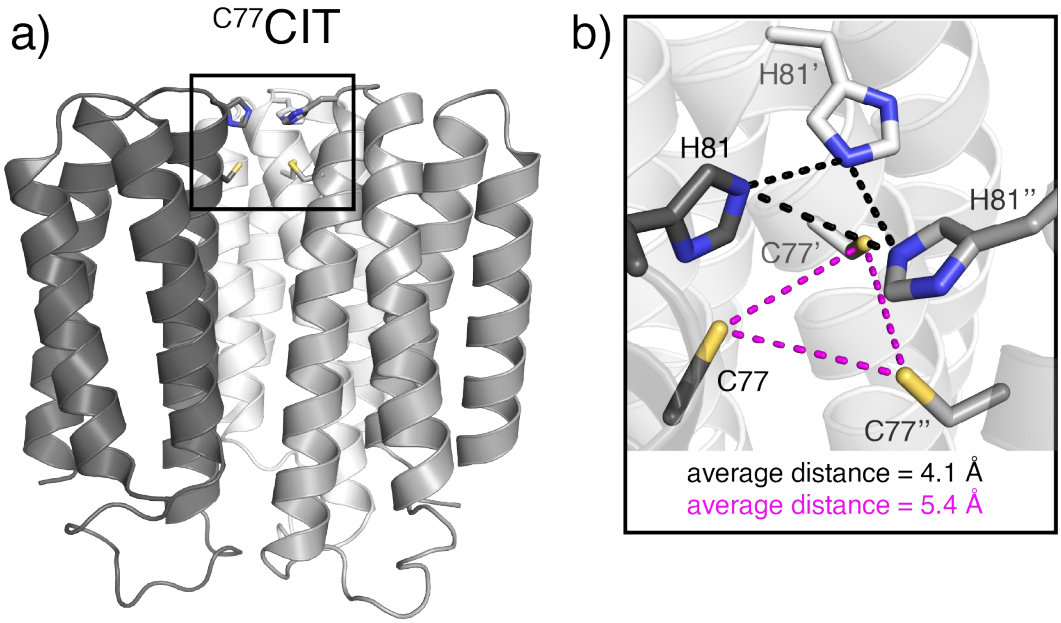

Figure 3.27 | Crystal structure of CIT and modeling of Cys77 mutation. (a) Overview of CIT with Cys77 mutations installed. (b) Closeup of putative coordination site of ^{C77}CIT, with inter-S_γ of Cys77 and inter-N_ε distances of His81 indicated in magenta and black, respectively. Unlike in ^{C77}ncTriCyt2, the His_3 layer is positioned closer to the surface than the Cys_3 layer.

We initially screened for ^{C77}CIT binding to Fe^{II}-Zn^{II} in solution using ESI-MS (**Figure 3.28**). Herein, we report mass/charge (m/z) peaks corresponding to trimeric species at z = 15 and z = 13 (**Figure 3.28**). At z = 15 and z = 13, m/z peaks above background corresponding to metal-trimer complexes with 1:1 and 3:1 metal:trimer stoichiometries are observed for ^{C77}CIT + Ni^{II} and ^{C77}CIT + Cu^{II} (**Figure 3.28**). Notably, at z = 13 only a triply metalated trimeric species is observed for ^{C77}CIT + Ni^{II} **(Figure 3.28, right)**.

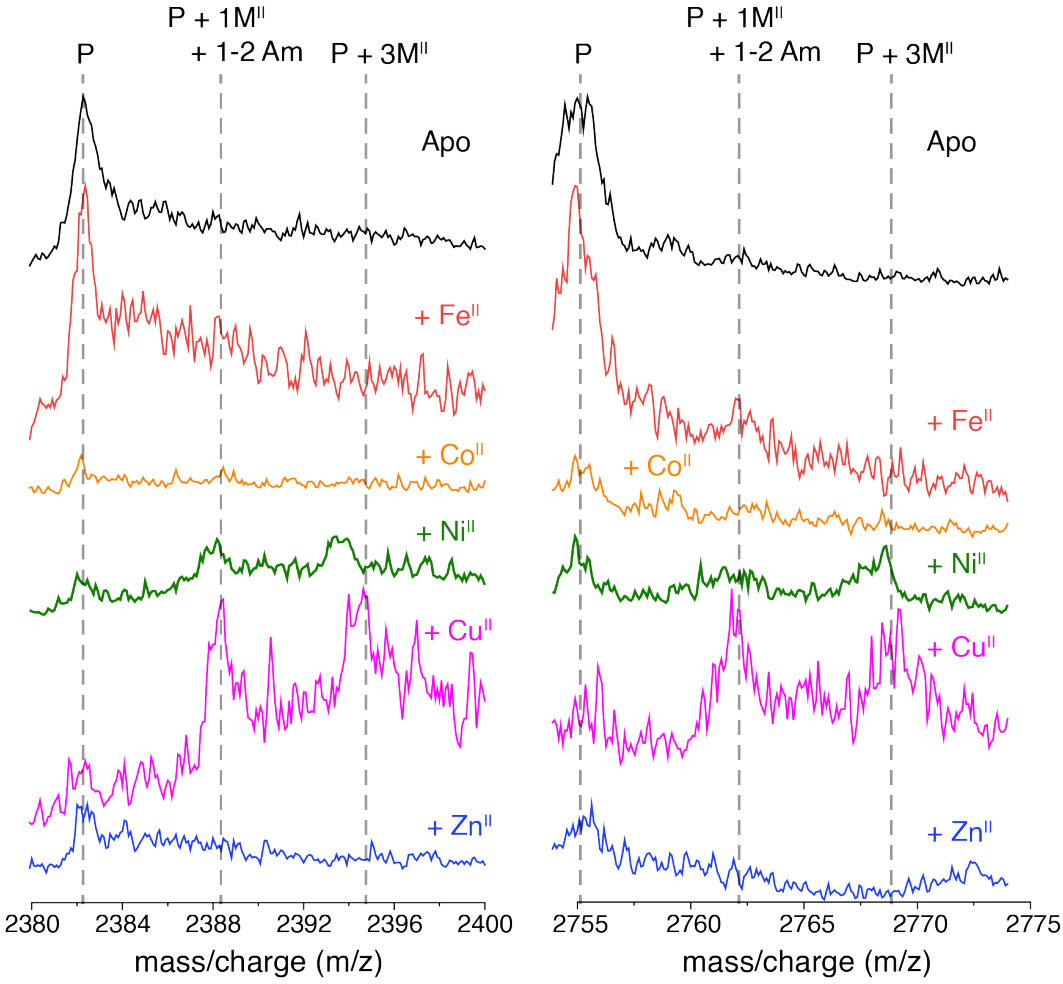

Figure 3.28 | ESI-MS of ^{C77}CIT + {Fe^{II}-Zn^{II}}. The m/z peaks shown correspond to trimeric species with z = 15 (left) and z = 13 (right). All samples were measured in 20 mM ammonium bicarbonate buffer, which was prepared using mass spectrometry-grade water. Am = ammonium ion.

We were surprised to observe 3:1 metal:trimer stoichiometries, as all metal additions were made at 2.0 equiv. metal/trimer. Considering this result, we surmise that the protein/metal mixtures were not allowed enough time to equilibrate, giving rise to both singly and triply metalated trimeric species. Another surprising observation regarding the preliminary ESI-MS data is the absence of metal-protein complexes for ^{C77}CIT + Zn^{II} at z =15 and z = 13, as Zn^{II} is positioned above Ni^{II} on the IW Series and is thus expected to form more stable complexes with ^{C77}CIT. We would need to carry out additional characterizations, including competitive binding titrations and X-ray crystallography, to confirm the potential anti-IW behavior.

While no ^{C77}CIT-Fe^{II} complexes were observed by ESI-MS, the addition of 0.6 equiv. $FeCl_2$/monomer at a monomer concentration of 1 mM led to the formation of a magenta-colored species with an absorption band centered at 496 nm (**Figure 3.29a, blue trace**). This result was surprising given that such absorption features are often associated with LMCTs to Fe^{III}, yet the addition of $FeCl_2$ to ^{C77}CIT was performed under anaerobic conditions and in the absence of exogenous oxidants. The addition of 0.6 mM $FeCl_2$ to buffer in the absence of protein did not yield the same visible absorption band, confirming that the UV-vis feature at 496 nm was not the result of iron oxidation by buffer components (**Figure 3.29a, black trace**). The addition of DT to {^{C77}CIT + 0.6 equiv. Fe^{II}/monomer} led to the disappearance of this absorption band (**Figure 3.29a, orange trace**), while extended exposure (0.5 h) to air resulted in a > 2-fold increase in the absorbance at 496 nm (**Figure 3.29a, red trace**). Taken together, the UV-vis spectra suggested the formation of Fe^{3+}-containing species, yet X-band EPR measurements of {^{C77}CIT + 0.6 equiv. Fe^{II}/monomer} revealed an EPR-silent species (**Figure 3.29b**).

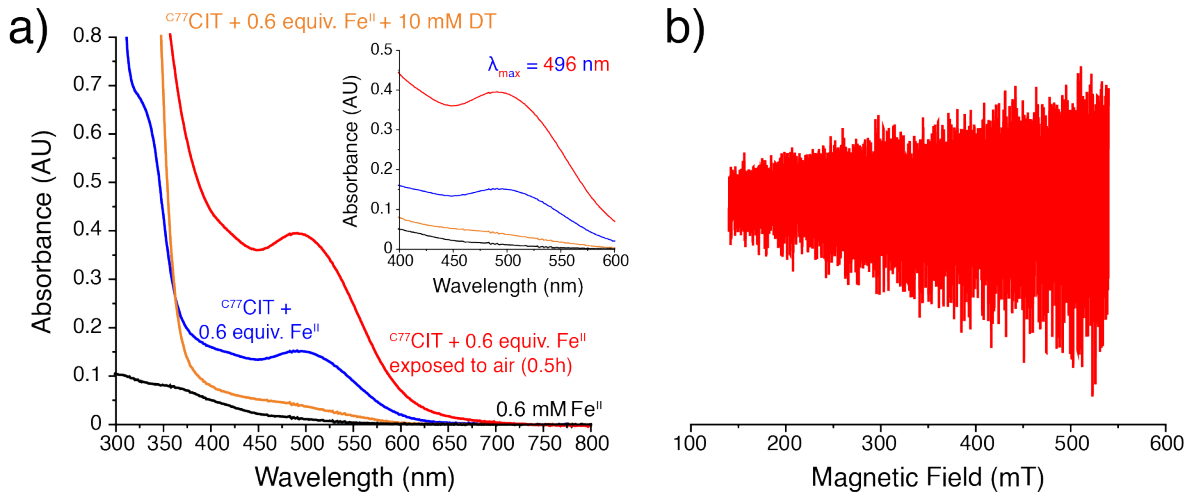

Figure 3.29 | UV-vis and EPR spectra of ^{C77}CIT mixed with 0.6 equiv. Fe^{II}/monomer. (a) UV-vis spectra of {^{C77}CIT + Fe^{II}} in the absence (blue trace) and presence (red and orange traces) of redox stimuli. A UV-vis spectrum of $FeCl_2$ in the absence of protein (black trace) did not reveal the same spectral features as those observed in the protein-containing samples, confirming that the broad absorption band observed at 496 nm (blue and red traces) was not attributable solely to Fe^{II} oxidation by buffer components. (b) X-band EPR spectrum of {^{C77}CIT + Fe^{II}} indicating the presence of an EPR-silent species. EPR parameters: temperature = 10K, microwave frequency = 9.36 GHz, modulation amplitude = 5.00 G, microwave power = 2.0 mW.

Crystals of ^{C77}CIT loaded with 0.6 equiv./monomer of $FeCl_2$ were prepared under anaerobic conditions and had a magenta hue, as observed in solution (data not shown). Crystal structures of {^{C77}CIT + Fe^{II}} revealed a trimeric ASU (**Figure 3.30a**). Two observations immediately surprised us. First, the His81 residues were not observed to coordinate a Fe ion, and second, instead of engaging in the T-stacking interactions observed in all other TriCyt variants, Trp70 residues were engaged in cation-π interactions with Arg34 (**Figure 3.30b**). Instead of being coordinated at a dinuclear site composed of Cys77 and His81, we observed that Fe was coordinated by Cys77 residues in a mononuclear, trigonal planar, Cys_3 coordination environment (**Figure 3.30b-c and 3.31a**). The identity of the metal ion was confirmed by collecting reflection data at the K-edge of Fe (**Figure 3.30b**). According to the database MetalPDB,[67] no metalloproteins have

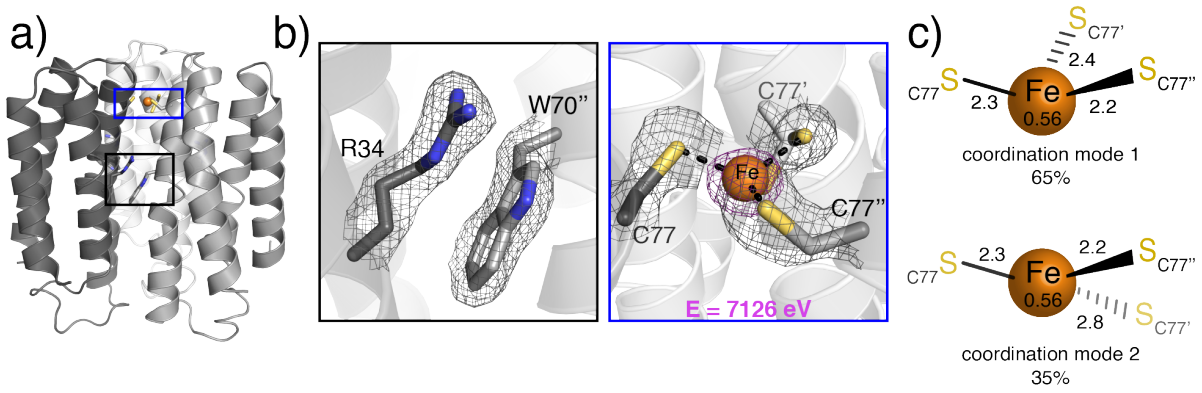

Figure 3.30 | Crystal structure of {^{C77}CIT + Fe^{II}}. (a) Overview of trimer. (b) Electron density maps highlighting a R34/W70 cation-π interaction (average inter-ring distance = 3.5 ± 0.2 Å, left) and a mononuclear, Fe:Cys$_3$ coordination site (right) with two discrete coordination modes. 2mFo-Dfc maps (gray) and the anomalous difference map (magenta) are contoured at 1σ and 9σ, respectively. (c) Schemes of Fe coordination modes observed in the crystal structure. The occupancy of the Fe atom is 0.56, thus the total occupancy of trigonal planar-coordinated Fe in the ASU is 0.36.

been reported which feature a trigonal planar, Fe:Cys$_3$ coordination site. The closest synthetic analogue to this site is a trigonal planar [FeSR$_3$]$^-$ complex (R = C$_6$H$_2$-2,4,6-tertbutyl$_3$) whose electronic structure was extensively characterized by the Holm and Power Groups using Mossbauer spectroscopy.[68, 69] Thus far, no studies have been performed probing the reactivity of this complex, but three-coordinate Fe complexes have previously been shown to possess a wide range of functions encompassing C-F bond cleavage,[70] oxygen activation,[71] NO binding,[72] and hydrazine reduction.[73] As indicated by the pyramidalization angle (θ_1-90°, **Figure 3.31b**), the Fe:Cys$_3$ site observed in our structure has significant distortions away from a trigonal planar and toward a trigonal pyramidal geometry. At our Fe:Cys$_3$ complex, θ_1-90° = 10°, indicative of a coordination geometry intermediate between trigonal planar (θ_1-90° = 0°) and trigonal pyramidal/tetrahedral (θ_1-90° = 19.5°). Such a distortion could enable the coordination complex to bind small molecule ligands through π back-bonding, such as CO and CN (**Figure 3.31c**). It has previously been observed by Chambers *et. al.*

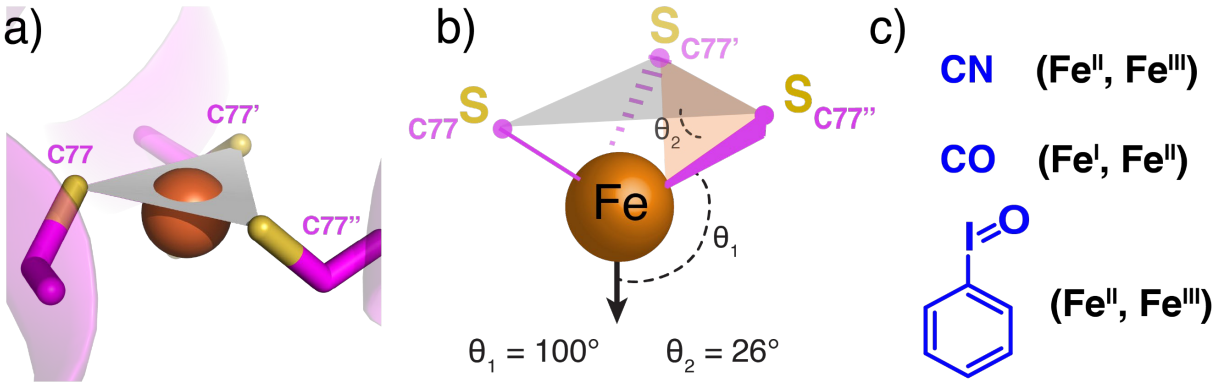

Figure 3.31 | Geometry of coordination mode 1 observed in {^{C77}CIT + Fe^{II}} crystal structure and candidate ligands to bind the complex. (a) Modeled Fe ion superimposed onto the Cys_3 plane. As the coordination plane does not divide the sphere into two equal parts, we can surmise that there is deviation from the ideal trigonal planar coordination geometry. (b) Scheme of the coordination site with selected angles (θ_1 and θ_2). In an ideal trigonal planar geometry, $\theta_1 = 90°$, whereas in a tetrahedral geometry, $\theta_1 = 109.5°$. (c) Small molecules that could coordinate to or react with the Fe:Cys_3 site. Iodosylbenzene (bottom) is an oxo-transfer reagent previously used to obtain Fe-oxo adducts. Possible resting redox states of the iron center required for binding/reacting with each molecule are denoted in parentheses.

that Fe^{II} complexes featuring coordination by alkoxide ligands in a trigonal planar geometry react with oxo-transfer reagents to render high valent, Fe^{IV}-oxo species.[71] Like alkoxide, cysteine is a weak field ligand, and so it is possible that a Fe:Cys_3 complexes could mediate the same reactivities observed in the Fe-alkoxide complex. Unlike alkoxide, however, cysteine is itself prone to oxidation, potentially leading to oxo-transfer to the ligands rather than or in addition to the metal center.[74, 75] In future studies, we will assess the reactivity of the Fe:Cys_3 site toward various ligands (**Figure 3.31c**) using UV-vis, Mössbauer, X-band EPR, and NMR spectroscopies. These solution characterizations will parallel efforts to crystallize ligand-bound Fe:Cys_3 complexes.

3.4 Conclusions

The primary and secondary sphere characteristics of a metalloprotein play a critical role in determining its function. Achieving multiple, distinct metal-based functions within a designed metalloprotein thus requires primary and secondary sphere diversification. In pursuit of selective

Ln^{III} coordination, redox-dependent/hydrolytic catalysis, and Fe-S cluster binding, we assessed the functional potential of TriCyt2 and TriCyt3 mutants featuring a wide range of primary and secondary sphere environments.

Our vision for obtaining selective Ln^{III} binding proteins centered on high-throughput screening protocols in which a library of TriCyt2 mutants would be screened for lanthanide-directed trimerization via either SEC-HPLC or fluorescence spectroscopy. Control experiments revealed that our periplasmic extraction method, which we used to isolate our mutants without column purifications, introduced artifacts to our readouts. Possibly due to the high metal content and heterogeneity of the periplasm,[76] periplasmic extracts of TriCyt2 were revealed by SEC-HPLC to assemble as trimers in the *absence* of exogenously added metal ions, while fluorescence spectroscopy measurements on the extracts had low reproducibility. Future protocols would include media expression of protein constructs to ensure samples are in a less heterogenous milieu and the utilization of ESI-MS to identify protein-Ln^{III} complexes, which would enable us to screen for Ln^{III} binding in both TriCyt2 and TriCyt3-based libraries.

In pursuit of a catalytic variant of TriCyt3, we first mutated His77 into Gly/Ala and sought to establish whether the resulting mutants had metal-dependent activity. ^{G77}ncTriCyt3 exhibited Cu^{II}-dependent peroxidase activity for three distinct substrates, including ODA, and thus became a starting point for directed evolution. We aimed to improve our emergent ODA peroxidase activity by generating and screening mutants with larger cavities than the starting point. Following library generation and high-throughput screening of catalytic activity via UV-vis spectroscopy, we identified a variant (GESL) which exhibited a ~1.6-fold improvement in peroxidase activity compared to ^{G77}ncTriCyt3 while remaining trimeric in the presence of Cu^{II}. While the increase in catalytic rate is notable, we envision that GESL could serve as a novel starting point for obtaining

even more active peroxidases through additional modifications of the secondary sphere. Future studies could include catalytic screening of a GESL-based library in which residue position 77 or positions along the 80s loop are randomized.

The coordination of two or more compositionally distinct yet electronically coupled metal centers within protein assemblies is essential to complex transformations encompassing proton and nitrogen reduction.[39, 77] In pursuit of electronically coupled metal centers, we mutated His77 of TriCyt2 into Cys to generate a metalloprotein construct (^{C77}ncTriCyt2) which could simultaneously bind a synthetic 4Fe-4S cluster ($Fe_4S_4^{syn}$) and single metal ions at $(Cys77)_3$ and $(His73)_3$ coordination sites, respectively. While initial solution characterization suggested the incorporation of an unusually low potential, protein-Fe_3S_4 complex, crystal structures revealed that ^{C77}ncTriCyt2 coordinated single Fe ions at Cys77 and His73 residues, resulting in a dinuclear coordination site with a maximum Fe-Fe distance of 3.2 Å. This serendipitous discovery has led to two new pursuits, the first to stabilize heterometallic, dinuclear coordination sites in ^{C77}ncTriCyt2 by harnessing differences in metal binding affinity for His/Cys ligands, and the second to assess the reactivity of mononuclear, $Fe:Cys_3$ complexes within our trimeric scaffold. These investigations could lead to the discovery of new-to-nature metalloprotein functions while improving our understanding of reactivity at low-coordinate iron complexes.

3.5 Materials and Methods

3.5.1 Mutagenesis, Expression, and Purification

Mutagenesis and Sequencing. PCR-based site-directed mutagenesis of plasmids encoding TriCyt2 or TriCyt3 (pET20b-[TriCyt2] or pET20b-[TriCyt3]) was performed as previously described.[78] To generate heme-less variants for cytoplasmic expression, the periplasmic leader sequences of pET20b-[TriCyt2] and pET20b-[TriCyt3] were deleted via PCR-based site-directed mutagenesis. To generate heme-less variants for expression in LB media, the periplasmic leader sequences of pET20b-[TriCyt2] and pET20b-[TriCyt3] were replaced via PCR-based site-directed mutagenesis with the following media export sequence: *MRKSLLAILAVSSLVFSSASFA*.[79] All PCR products were transformed into competent XL-1 Blue *E. coli* cells and allowed to grow on LB/agar plates containing ampicillin (100 μg/mL). When generating mutant libraries, the number of colonies grown following transformation and for DNA sequencing exceeded the number of expected unique sequences by a factor of ~3. Colonies were grown in LB media supplemented with ampicillin (100 μg/mL) and shaken at 150 rpm for 16 hours at 37°C prior to miniprep and sequencing.

Protein expression. Purified plasmids of heme-containing mutants were transformed into competent BL21 (DE3) *E. coli* cells containing the *ccm* (cytochrome c maturation) cassette plasmid, pEC86,[80] and allowed to grow for 20 hours on LB/agar plates containing ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL). Purified plasmids of heme-less mutants were transformed into competent BL21 (DE3) *E. coli* cells without the *ccm* cassette plasmid and allowed to grow for 20 hours on LB/agar plates containing ampicillin (100 μg/mL). Starter cultures were grown overnight for 16 hours at 37°C in LB media supplemented with the same antibiotic concentrations, diluted either 50- or 100-fold into fresh, antibiotic supplemented LB media, and

then grown at 37°C until the OD_{600} reached 0.6-1. For heme-containing mutants, cultures were then diluted 10,000-fold into LB media supplemented with ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) and shaken at 100 RPM for 20-24 hours at 37°C. For heme-less mutants, expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.6 mM) and cultures were shaken at 200 rpm for 6 hours at 37°C post-induction. For mutants expressed in the periplasm or cytoplasm, cells were pelleted via centrifugation (5,000 x g, 4°C, 5 min) and the media discarded. For mutants exported to the media, media was decanted off the cell pellet following centrifugation. Mutants exported to the media did not undergo column purifications prior to analysis.

Protein Purification. Cell pellets were resuspended in a 10 mM NaP_i buffer solution (pH 8.0) and vigorously stirred until all pellets were resuspended. The resulting mixture was sonicated for 15 min in cycles of 30 seconds on and 60 seconds off (Qsonica). After the pH of the lysate was readjusted to 8.0, it was clarified by centrifugation (10,000 x g, 4°C, 20 min). The cleared lysate was diluted 10-fold into a 10 mM NaP_i buffer solution (pH 8.0) and applied to a Q Sepharose Fast Flow (Biorad) resin preequilibrated with the same buffer solution. Protein was eluted in 125 mL fractions with a step-gradient of 0-1 M NaCl (total volume = 1.25-2.50 L). Fractions which contained target protein in significant quantities as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were pooled, concentrated, and exchanged into 10 mM sodium acetate buffer solution (pH 4.5). The protein was then loaded onto a 5 mL High-S cartridge column (Biorad) preequilibrated with the same buffer solution and eluted using a step-gradient of 0-400 mM NaCl. Protein identity and purity were determined by ESI-MS and 15% SDS-PAGE gels, respectively. When purifying mutants with Cys residues, all buffer solutions were supplemented with 5 mM DTT.

For the high-throughput screening of lanthanide-binding metalloproteins, heme-containing TriCyt2 mutants were expressed in the periplasm and extracted via chloroform shock.[81] Cell pellets were resuspended in chloroform and allowed to incubate for 15 minutes. Afterwards, the suspension was diluted with 10 equivalents by volume of 10 mM HEPES (pH 7.5) and centrifuged (10,000 x g, 4°C, 15 min) to separate the aqueous and organic layers. The red aqueous layer was decanted off the organic layer and analyzed without further purification.

For the high-throughput screening of ODA oxidation by TriCyt3 mutants, individual colonies were grown in ~2 mL LB media in deep well culture plates (Genesee Scientific) supplemented with ampicillin (100 μg/mL). After the OD_{600} of three unique wells reached 0.6-0.8 OD, each well was induced with 0.6 mM IPTG and allowed to grow for 16 hours. Following the 16-hour growth period, cells were pelleted via centrifugation (5,000 x g, 4°C, 10 min) and the media transferred to 96-well filter plates with a 100 kDa cutoff (Pall). The media was then passed through the filter plates via centrifugation (5,000 x g, 4°C, 30 mins) and transferred to 96-well filter plates with a 3 kDa cutoff (Pall). Samples were then extensively buffer exchanged into a 50 mM HEPES buffer solution supplemented with 150 mM NaCl (pH 7.5) and transferred to UV-star 96 well microplates (Grenier Bio-One) for analysis. Pellets obtained from the initial centrifugation step were preserved for miniprep and sequencing.

3.5.2 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

Sedimentation velocity (SV) measurements were made in solutions of either 20 mM MOPS (pH 7.5) or 20 mM HEPES (pH 7.5) at 25°C on a Beckman XL-A instrument equipped with a AN-60 Ti rotor and at 41,000 RPM. Heme-containing samples were monitored at 415 nm (corresponding to the Soret band of the heme) for up to 12 h. To ensure $A_{415} \geq 1.0$ AU, SV-AUC measurements were performed at a minimum concentration of 7 μM protein monomer. Heme-less

samples were monitored at 280 nm, and to ensure $A_{280} \geq 1.0$ AU, measurements were performed at a minimum concentration of 60 μM protein monomer. When analyzing metal-directed oligomerization, protein samples were mixed with metal chloride salts at 1 equiv. metal ion/trimer. Following metal addition, samples were equilibrated for at least one hour at 25°C prior to SV measurements. Scans were processed and molecular weight distributions calculated using SEDFIT software.[82] Fitting parameters such as the buffer density (0.9988 g/mL), buffer viscosity (0.01007 poise), and partial specific volume (0.7313 mL/g) were calculated by SEDNTERP.[82] SV profiles are shown at a confidence level of 95%. Oligomerization yields were estimated based on Riemann integrations of the peaks of the SV profiles, where the bounds of each discrete peak were defined by the FWHM.

3.5.3 Spectroscopic characterization methods

UV-vis spectra were acquired on Agilent 8453 and Cary 60 spectrometers. For UV-vis measurements of potentially air-sensitive samples, all samples were prepared under anaerobic conditions and transferred to quartz cuvettes sealed with parafilm and septa. X-band continuous wave (CW) EPR experiments were carried out on a Bruker EMXplus EPR spectrometer (Bruker Biospin) equipped with a high-sensitivity resonator (ER4119HS) operating in perpendicular mode at 9.4 GHz. Spectra were collected at 10-40 K using a Waveguide cryostat (ColdEdge Technologies). Samples of $\{^{C77}\text{ncTriCyt2} + Fe_4S_4^{syn}\}$ were prepared at ~1.2 mM protein monomer and nominally ~0.96 mM Fe^{II}/Fe^{III}. Samples of $\{^{C77}\text{CIT} + Fe^{II}\}$ were prepared at ~0.83 mM protein monomer and 0.5 mM $FeCl_2$. All samples were prepared under anaerobic conditions in a 25-50 mM HEPES buffer solution supplemented with 10% (v/v) glycerol (pH 7.5) as a glassing agent. Spectra were plotted using EasySpin.[83]

Tryptophan fluorescence intensity scans and high-throughput UV-vis measurements were performed on an Infinite M Nano 200Pro plate reader (Tecan). To measure tryptophan fluorescence intensity, fluorescence excitation was carried out at 280 nm and emission monitored from 300-450 nm (step size = 1 nm). Samples were prepared at ~30 μM protein monomer and 10 μM metal chloride salt and allowed to equilibrate for at least 30 minutes prior to being transferred to UV-star 96 well microplates (Grenier Bio-One). Before measuring Cu^{II}-dependent ODA oxidation activities, media containing TriCyt3 mutants was supplemented with 10 μM $CuSO_4$ and allowed to equilibrate overnight in UV-star 96 well microplates (Grenier Bio-One). The media was then further supplemented with 10 mM H_2O_2 and 1 mM ODA immediately prior to analysis. To monitor ODA oxidation, A_{460} was measured every 60 s across a total reaction time of 1800 s. Natural log plots were generated for the first ten time points. To estimate protein concentrations in the sample wells, absorbance scans were performed with a wavelength range of 250-800 nm (step size = 2 nm).

3.5.4 Size exclusion chromatography-HPLC (SEC-HPLC)

Prior to SEC-HPLC analysis, periplasmic extracts of TriCyt2 were passed through a spin filter with a 100 kDa cutoff to remove potential protein aggregate. The flowthrough was then buffer exchanged extensively into filtered 20 mM HEPES (pH 7.5) and the protein concentration estimated based on the absorbance at 415 nm. Samples were prepared at ~30 μM protein monomer and 10 μM metal chloride salt in a solution of 20 mM HEPES (pH 7.5) and allowed to equilibrate for at least one hour. Samples were passed through a Zenix SEC-300 column (Sepax) at 25°C, operating at a flow rate of 1 mL/min and in 20 mM HEPES (pH 7.5). Sample elution was monitored at 415 nm and over the course of 12-15 minutes.

3.5.5 X-ray structure determination

Crystals of TriCyt2 and TriCyt3 mutants were obtained by sitting-drop vapor diffusion at 25°C. Samples of {^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$} were prepared at ~0.55 mM protein monomer in 50 mM TRIS supplemented with 25 mM β-ME (pH 8.5). All other samples were prepared at ~3 mM protein monomer in 25 mM HEPES (pH 7.5). Cys-containing mutants were mixed with HEPES buffer solution supplemented with 25 mM β-ME before being extensively buffer exchanged into HEPES buffer solution without β-ME. All samples with Cys-containing mutants or initially mixed with $Fe_4S_4^{syn}$ were crystallized under anaerobic conditions. Samples were mixed at a ratio of 1 µL: 1 µL with mother liquor. Crystals were transferred into either mother liquor supplemented with glyercol (25% v/v) or perfluoro polyether (Hampton) for cryoprotection prior to freezing. Diffraction data for crystals of {^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$} and {^{C77}CIT + Fe^{II}} were collected at 100 K on ALS Beamline 8.3.1. Diffraction data were processed using XDS and scaled using XSCALE.[84] Molecular replacement was carried out using Phaser with either monomeric ncTriCyt2 ({^{C77}ncTriCyt2 + $Fe_4S_4^{syn}$}) or trimeric CIT ({^{C77}CIT + Fe^{II}}) as the search model.[85] Refinement was performed using phenix.refine[86] while model building and placement of metal ions/water was performed using COOT.[87]

3.5.6 Native electrospray ionization (ESI) mass spectrometry

ESI-MS experiments were performed using an LTQ XL linear ion trap mass spectrometer equipped with a heated-electrospray ionization (HESI) source (Thermo). All solutions for ESI-MS were prepared using LC-MS grade water (JT Baker). Samples were prepared at ~50 µM protein monomer and 33 µM metal chloride salt in 20 mM NH_4HCO_3 (pH 8.0). Samples were analyzed in high mass mode for a m/z range of 1500-3000 and with the following instrument parameters: source voltage = 2.52 kV, source temperature = 101.36°C, sheath gas flow rate = 11.94 AU,

auxiliary gas flow rate = 11.97 AU, capillary temperature = 225°C. To prevent cross-contamination, peek tubing and steel capillary were sequentially washed using 0.5 mL 0.1% formic acid and 0.5 mL 20 mM NH_4HCO_3 between measurements. Distributions were analyzed using Qual Browser (Thermo).

3.6 Acknowledgements

We thank C. Yu and T. Choi for their assistance with EPR and native ESI-MS spectrometry measurements, respectively. This work was funded by the NSF (CHE1607145), by the NIH (a CBI traineeship to A.K. through T32GM112584 and R01GM138884 to F.A.T), by NASA (80NSSC18M0093; ENIGMA: Evolution of Nanomachines in Geospheres and Microbial Ancestors (NASA Astrobiology Institute Cycle 8)), and by the National Institutes of Health (NIH; R01-GM138884). Portions of this research were carried out at the Advanced Light Source at the Lawrence Berkeley National Laboratory, which are supported by the DOE, Office of Science, Office of Basic Energy Sciences under contracts DE-AC02-76SF00515 and DE-AC02-05CH11231, respectively. EPR data was acquired on an instrument obtained through a Major Research Instrumentation fund supported by the National Science Foundation (NSF; CHE-2019066)

3.7 References

1. Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F. A., Metal-Templated Design of Chemically Switchable Protein Assemblies with High-Affinity Coordination Sites. *Angew. Chem. Int. Ed. Engl.* **2020,** *59* (49), 21940-21944.
2. Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A., Metals in protein-protein interfaces. *Annu. Rev. Biophys.* **2014,** *43*, 409-31.
3. Kiefer, L. L.; Paterno, S. A.; Fierke, C. A., Hydrogen bond network in the metal binding site of carbonic anhydrase enhances zinc affinity and catalytic efficiency. *J. Am. Chem. Soc.* **2002,** *117* (26), 6831-6837.
4. Fernandes, A.; Afonso, J. C.; Dutra, A. J. B., Separation of nickel(II), cobalt(II) and lanthanides from spent Ni-MH batteries by hydrochloric acid leaching, solvent extraction and precipitation. *Hydrometallurgy* **2013,** *133*, 37-43.
5. Teo, R. D.; Termini, J.; Gray, H. B., Lanthanides: Applications in Cancer Diagnosis and Therapy. *J. Med. Chem.* **2016,** *59* (13), 6012-24.
6. Woodruff, D. N.; Winpenny, R. E.; Layfield, R. A., Lanthanide single-molecule magnets. *Chem. Rev.* **2013,** *113* (7), 5110-48.
7. Dushyantha, N.; Batapola, N.; Ilankoon, I. M. S. K.; Rohitha, S.; Premasiri, R.; Abeysinghe, B.; Ratnayake, N.; Dissanayake, K., The story of rare earth elements (REEs): Occurrences, global distribution, genesis, geology, mineralogy and global production. *Ore Geol. Rev.* **2020,** *122*.
8. Cotruvo, J. A., Jr.; Featherston, E. R.; Mattocks, J. A.; Ho, J. V.; Laremore, T. N., Lanmodulin: A Highly Selective Lanthanide-Binding Protein from a Lanthanide-Utilizing Bacterium. *J. Am. Chem. Soc.* **2018,** *140* (44), 15056-15061.
9. Cook, E. C.; Featherston, E. R.; Showalter, S. A.; Cotruvo, J. A., Jr., Structural Basis for Rare Earth Element Recognition by Methylobacterium extorquens Lanmodulin. *Biochemistry* **2019,** *58* (2), 120-125.
10. Deblonde, G. J.; Mattocks, J. A.; Park, D. M.; Reed, D. W.; Cotruvo, J. A., Jr.; Jiao, Y., Selective and Efficient Biomacromolecular Extraction of Rare-Earth Elements using Lanmodulin. *Inorg. Chem.* **2020,** *59* (17), 11855-11867.
11. Bünzli, J.-C. G., Review: Lanthanide coordination chemistry: from old concepts to coordination polymers. *J. Coord. Chem.* **2014,** *67* (23-24), 3706-3733.
12. Bunzli, J. C.; Piguet, C., Taking advantage of luminescent lanthanide ions. *Chem. Soc. Rev.* **2005,** *34* (12), 1048-77.
13. Slope, L. N.; Daubney, O. J.; Campbell, H.; White, S. A.; Peacock, A. F. A., Location-Dependent Lanthanide Selectivity Engineered into Structurally Characterized Designed Coiled Coils. *Angew. Chem.* **2021,** *133* (46), 24678-24682.
14. Vivian, J. T.; Callis, P. R., Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophys. J.* **2001,** *80* (5), 2093-2109.
15. Taraska, J. W.; Puljung, M. C.; Olivier, N. B.; Flynn, G. E.; Zagotta, W. N., Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nat. Methods* **2009,** *6* (7), 532-7.
16. Taraska, J. W.; Puljung, M. C.; Zagotta, W. N., Short-distance probes for protein backbone structure based on energy transfer between bimane and transition metal ions. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106* (38), 16227-32.

17. Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R., Stability and folding kinetics of structurally characterized cytochrome c-b562. *Biochemistry* **2006,** *45* (35), 10504-11.
18. Hoffnagle, A. M.; Eng, V. H.; Markel, U.; Tezcan, F. A., Computationally Guided Redesign of a Heme-free Cytochrome with Native-like Structure and Stability. *Biochemistry* **2022,** *61* (19), 2063-2072.
19. Joseph, C. A.; Maroney, M. J., Cysteine dioxygenase: structure and mechanism. *Chem. Commun.* **2007,** (32), 3338-49.
20. Sheet, D.; Paine, T. K., Aerobic alcohol oxidation and oxygen atom transfer reactions catalyzed by a nonheme iron(ii)-alpha-keto acid complex. *Chem. Sci.* **2016,** *7* (8), 5322-5331.
21. Ponduru, T. T.; Sun, Z.; Cundari, T. R.; Rasika Dias, H. V., Nitrene Insertion into Aromatic and Benzylic C−H Bonds Catalyzed by Copper Complexes of Fluorinated Bis- and Tris(pyrazolyl)borates. *ChemCatChem* **2019,** *11* (19), 4966-4973.
22. Brand, U.; Rombach, M.; Vahrenkamp, H., Methylation of zinc bound thiolates; a model for cobalamine independent methionine synthase. *Chem. Commun.* **1998,** (24), 2717-2718.
23. Ji, P.; Park, J.; Gu, Y.; Clark, D. S.; Hartwig, J. F., Abiotic reduction of ketones with silanes catalysed by carbonic anhydrase through an enzymatic zinc hydride. *Nat. Chem.* **2021,** *13* (4), 312-318.
24. Okrasa, K.; Kazlauskas, R. J., Manganese-substituted carbonic anhydrase as a new peroxidase. *Chem. Eur. J.* **2006,** *12* (6), 1587-96.
25. Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; Basanta, B.; Bender, B. J.; Blacklock, K.; Bonet, J.; Boyken, S. E.; Bradley, P.; Bystroff, C.; Conway, P.; Cooper, S.; Correia, B. E.; Coventry, B.; Das, R.; De Jong, R. M.; DiMaio, F.; Dsilva, L.; Dunbrack, R.; Ford, A. S.; Frenz, B.; Fu, D. Y.; Geniesse, C.; Goldschmidt, L.; Gowthaman, R.; Gray, J. J.; Gront, D.; Guffy, S.; Horowitz, S.; Huang, P. S.; Huber, T.; Jacobs, T. M.; Jeliazkov, J. R.; Johnson, D. K.; Kappel, K.; Karanicolas, J.; Khakzad, H.; Khar, K. R.; Khare, S. D.; Khatib, F.; Khramushin, A.; King, I. C.; Kleffner, R.; Koepnick, B.; Kortemme, T.; Kuenze, G.; Kuhlman, B.; Kuroda, D.; Labonte, J. W.; Lai, J. K.; Lapidoth, G.; Leaver-Fay, A.; Lindert, S.; Linsky, T.; London, N.; Lubin, J. H.; Lyskov, S.; Maguire, J.; Malmstrom, L.; Marcos, E.; Marcu, O.; Marze, N. A.; Meiler, J.; Moretti, R.; Mulligan, V. K.; Nerli, S.; Norn, C.; O'Conchuir, S.; Ollikainen, N.; Ovchinnikov, S.; Pacella, M. S.; Pan, X.; Park, H.; Pavlovicz, R. E.; Pethe, M.; Pierce, B. G.; Pilla, K. B.; Raveh, B.; Renfrew, P. D.; Burman, S. S. R.; Rubenstein, A.; Sauer, M. F.; Scheck, A.; Schief, W.; Schueler-Furman, O.; Sedan, Y.; Sevy, A. M.; Sgourakis, N. G.; Shi, L.; Siegel, J. B.; Silva, D. A.; Smith, S.; Song, Y.; Stein, A.; Szegedy, M.; Teets, F. D.; Thyme, S. B.; Wang, R. Y.; Watkins, A.; Zimmerman, L.; Bonneau, R., Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **2020,** *17* (7), 665-680.
26. Van Stappen, C.; Deng, Y.; Liu, Y.; Heidari, H.; Wang, J. X.; Zhou, Y.; Ledray, A. P.; Lu, Y., Designing Artificial Metalloenzymes by Tuning of the Environment beyond the Primary Coordination Sphere. *Chem. Rev.* **2022**.
27. Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J., CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **2012,** *8* (10), e1002708.

28. Pistone, L.; Ottolina, G.; De, S.; Romero, A. A.; Martins, L. O.; Luque, R., Encapsulated Laccases for the Room-Temperature Oxidation of Aromatics: Towards Synthetic Low-Molecular-Weight Lignins. *ChemSusChem* **2016,** *9* (7), 756-62.
29. Giummarella, N.; Pylypchuk, I. V.; Sevastyanova, O.; Lawoko, M., New Structures in Eucalyptus Kraft Lignin with Complex Mechanistic Implications. *ACS Sustain. Chem. Eng.* **2020**.
30. Wiegand, A.-K.; Rit, A.; Okuda, J., Molecular zinc hydrides. *Coord. Chem. Rev.* **2016,** *314*, 71-82.
31. Klaui, W.; Schilde, U.; Schmidt, M., Fluoro[eta(3)-hydrotris(3-R-5-methylpyrazol-1-yl)borato]zinc(II): The First TpZnF Complexes, Convenient Precursors to Zinc Hydride Complexes(,). *Inorg. Chem.* **1997,** *36* (8), 1598-1601.
32. Ballmann, G.; Grams, S.; Elsen, H.; Harder, S., Dipyrromethene and β-Diketiminate Zinc Hydride Complexes: Resemblances and Differences. *Organometallics* **2019,** *38* (14), 2824-2833.
33. Shlian, D. G.; Amemiya, E.; Parkin, G., Synthesis of bis(2-pyridylthio)methyl zinc hydride and catalytic hydrosilylation and hydroboration of CO2. *Chem. Commun.* **2022,** *58* (26), 4188-4191.
34. Fontecave, M., Iron-sulfur clusters: ever-expanding roles. *Nat. Chem. Biol.* **2006,** *2* (4), 171-4.
35. Bai, F.; Morcos, F.; Sohn, Y. S.; Darash-Yahana, M.; Rezende, C. O.; Lipper, C. H.; Paddock, M. L.; Song, L.; Luo, Y.; Holt, S. H.; Tamir, S.; Theodorakis, E. A.; Jennings, P. A.; Onuchic, J. N.; Mittler, R.; Nechushtai, R., The Fe-S cluster-containing NEET proteins mitoNEET and NAF-1 as chemotherapeutic targets in breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (12), 3698-703.
36. Kiley, P. J.; Beinert, H., The role of Fe–S proteins in sensing and regulation in bacteria. *Curr. Opin. Microbiol.* **2003,** *6* (2), 181-185.
37. Pakharukova, N.; Malmi, H.; Tuittila, M.; Dahlberg, T.; Ghosal, D.; Chang, Y. W.; Myint, S. L.; Paavilainen, S.; Knight, S. D.; Lamminmaki, U.; Uhlin, B. E.; Andersson, M.; Jensen, G.; Zavialov, A. V., Archaic chaperone-usher pili self-secrete into superelastic zigzag springs. *Nature* **2022**.
38. Schurmann, P.; Buchanan, B. B., The ferredoxin/thioredoxin system of oxygenic photosynthesis. *Antioxid. Redox Signal.* **2008,** *10* (7), 1235-74.
39. Rutledge, H. L.; Tezcan, F. A., Electron Transfer in Nitrogenase. *Chem. Rev.* **2020,** *120* (12), 5158-5193.
40. Liu, J.; Chakraborty, S.; Hosseinzadeh, P.; Yu, Y.; Tian, S.; Petrik, I.; Bhagi, A.; Lu, Y., Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev.* **2014,** *114* (8), 4366-469.
41. Barry, S. M.; Challis, G. L., Mechanism and Catalytic Diversity of Rieske Non-Heme Iron-Dependent Oxygenases. *ACS Catal.* **2013,** *3* (10).
42. Gibney, B. R.; Mulholland, S. E.; Rabanal, F.; Dutton, P. L., Ferredoxin and ferredoxin-heme maquettes. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93* (26), 15041-6.
43. Laplaza, C. E.; Holm, R. H., Helix-loop-helix peptides as scaffolds for the construction of bridged metal assemblies in proteins: the spectroscopic A-cluster structure in carbon monoxide dehydrogenase. *J. Am. Chem. Soc.* **2001,** *123* (42), 10255-64.
44. Roy, A.; Sommer, D. J.; Schmitz, R. A.; Brown, C. L.; Gust, D.; Astashkin, A.; Ghirlanda, G., A de novo designed 2[4Fe-4S] ferredoxin mimic mediates electron transfer. *J. Am. Chem. Soc.* **2014,** *136* (49), 17343-9.

45. Jagilinki, B. P.; Ilic, S.; Trncik, C.; Tyryshkin, A. M.; Pike, D. H.; Lubitz, W.; Bill, E.; Einsle, O.; Birrell, J. A.; Akabayov, B.; Noy, D.; Nanda, V., In Vivo Biogenesis of a De Novo Designed Iron-Sulfur Protein. *ACS Synth. Biol.* **2020,** *9* (12), 3400-3407.
46. Aussignargues, C.; Pandelia, M. E.; Sutter, M.; Plegaria, J. S.; Zarzycki, J.; Turmo, A.; Huang, J.; Ducat, D. C.; Hegg, E. L.; Gibney, B. R.; Kerfeld, C. A., Structure and Function of a Bacterial Microcompartment Shell Protein Engineered to Bind a [4Fe-4S] Cluster. *J. Am. Chem. Soc.* **2016,** *138* (16), 5262-70.
47. Mirts, E. N.; Petrik, I. D.; Hosseinzadeh, P.; Nilges, M. J.; Lu, Y., A designed heme-[4Fe-4S] metalloenzyme catalyzes sulfite reduction like the native enzyme. *Science* **2018,** *361* (6407), 1098-1101.
48. Coldren, C. D.; Hellinga, H. W.; Caradonna, J. P., The rational design and construction of a cuboidal iron-sulfur protein. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94* (13), 6635-40.
49. Beinert, H.; Kennedy, M. C.; Stout, C. D., Aconitase as Ironminus signSulfur Protein, Enzyme, and Iron-Regulatory Protein. *Chem. Rev.* **1996,** *96* (7), 2335-2374.
50. Villafranca, J. J.; Mildvan, A. S., The Mechanism of Aconitase Action. *J. Biol. Chem.* **1971,** *246* (3), 772-779.
51. Tortora, V.; Quijano, C.; Freeman, B.; Radi, R.; Castro, L., Mitochondrial aconitase reaction with nitric oxide, S-nitrosoglutathione, and peroxynitrite: mechanisms and relative contributions to aconitase inactivation. *Free Radic. Biol. Med.* **2007,** *42* (7), 1075-88.
52. Lauble, H.; Kennedy, M. C.; Beinert, H.; Stout, C. D., Crystal structures of aconitase with trans-aconitate and nitrocitrate bound. *J. Mol. Biol.* **1994,** *237* (4), 437-51.
53. Tanifuji, K.; Lee, C. C.; Sickerman, N. S.; Tatsumi, K.; Ohki, Y.; Hu, Y.; Ribbe, M. W., Tracing the 'ninth sulfur' of the nitrogenase cofactor via a semi-synthetic approach. *Nat. Chem.* **2018,** *10* (5), 568-572.
54. Freibert, S. A.; Weiler, B. D.; Bill, E.; Pierik, A. J.; Muhlenhoff, U.; Lill, R., Biochemical Reconstitution and Spectroscopic Analysis of Iron-Sulfur Proteins. *Meth. Enzymol.* **2018,** *599*, 197-226.
55. Dizicheh, Z. B.; Halloran, N.; Asma, W.; Ghirlanda, G., De Novo Design of Iron-Sulfur Proteins. *Meth. Enzymol.* **2017,** *595*, 33-53.
56. Roche, B.; Aussel, L.; Ezraty, B.; Mandin, P.; Py, B.; Barras, F., Reprint of: Iron/sulfur proteins biogenesis in prokaryotes: formation, regulation and diversity. *Biochim. Biophys. Acta* **2013,** *1827* (8-9), 923-37.
57. Holm, R. H.; Lo, W., Structural Conversions of Synthetic and Protein-Bound Iron-Sulfur Clusters. *Chem. Rev.* **2016,** *116* (22), 13685-13713.
58. Sickerman, N. S.; Tanifuji, K.; Hu, Y.; Ribbe, M. W., Synthetic Analogues of Nitrogenase Metallocofactors: Challenges and Developments. *Chemistry* **2017,** *23* (51), 12425-12432.
59. Scintilla, S.; Bonfio, C.; Belmonte, L.; Forlin, M.; Rossetto, D.; Li, J.; Cowan, J. A.; Galliani, A.; Arnesano, F.; Assfalg, M.; Mansy, S. S., Duplications of an iron-sulphur tripeptide leads to the formation of a protoferredoxin. *Chem. Commun.* **2016,** *52* (92), 13456-13459.
60. Invernici, M.; Selvolini, G.; Silva, J. M.; Marrazza, G.; Ciofi-Baffoni, S.; Piccioli, M., Interconversion between [2Fe-2S] and [4Fe-4S] cluster glutathione complexes. *Chem. Commun.* **2022,** *58* (21), 3533-3536.
61. de Graaf, C.; Sousa, C., Study of the light-induced spin crossover process of the [Fe(II)(bpy)3]2+ complex. *Chemistry* **2010,** *16* (15), 4550-6.

62. Rutledge, H. L.; Rittle, J.; Williamson, L. M.; Xu, W. A.; Gagnon, D. M.; Tezcan, F. A., Redox-Dependent Metastability of the Nitrogenase P-Cluster. *J. Am. Chem. Soc.* **2019,** *141* (25), 10091-10098.
63. Holm, R. H.; Berg, J. M., Structures and reactions of iron-sulfur protein clusters and their synthetic analogs. In *Iron-Sulfur Proteins*, Spiro, T. G., Ed. Wiley-Interscience: New York, 1982; Vol. 4, pp 1-66.
64. Zhou, J.; Hu, Z.; Münck, E.; Holm, R. H., The Cuboidal Fe3S4 Cluster: Synthesis, Stability, and Geometric and Electronic Structures in a Non-Protein Environment. *J. Am. Chem. Soc.* **1996,** *118* (8), 1966-1980.
65. Hoppe, A.; Pandelia, M. E.; Gartner, W.; Lubitz, W., [Fe(4)S(4)]- and [Fe(3)S(4)]-cluster formation in synthetic peptides. *Biochim. Biophys. Acta* **2011,** *1807* (11), 1414-22.
66. Thomson, A. J.; Breton, J.; N. Butt, J.; A. Armstrong, F.; Hatchikian, E. C., Iron—sulphur clusters with labile metal ions. *J. Inorg. Biochem.* **1992,** *47* (1), 197-207.
67. Andreini, C.; Cavallaro, G.; Lorenzini, S.; Rosato, A., MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res.* **2013,** *41* (Database issue), D312-9.
68. MacDonnell, F. M.; Ruhlandt-Senge, K.; Ellison, J. J.; Holm, R. H.; Power, P. P., Sterically Encumbered Iron(II) Thiolate Complexes: Synthesis and Structure of Trigonal Planar [Fe(SR)3]- (R = 2,4,6-t-Bu3C6H2) and Moessbauer Spectra of Two- and Three-Coordinate Complexes. *Inorg. Chem.* **1995,** *34* (7), 1815-1822.
69. Sanakis, Y.; Power, P. P.; Stubna, A.; Munck, E., Mossbauer study of the three-coordinate planar Fe(II) thiolate complex [Fe(SR)(3)](-) (R = C(6)H(2)-2,4,6-tBu(3)): model for the trigonal iron sites of the MoFe(7)S(9):homocitrate cofactor of nitrogenase. *Inorg. Chem.* **2002,** *41* (10), 2690-6.
70. Vela, J.; Smith, J. M.; Yu, Y.; Ketterer, N. A.; Flaschenriem, C. J.; Lachicotte, R. J.; Holland, P. L., Synthesis and reactivity of low-coordinate iron(II) fluoride complexes and their use in the catalytic hydrodefluorination of fluorocarbons. *J. Am. Chem. Soc.* **2005,** *127* (21), 7857-70.
71. Chambers, M. B.; Groysman, S.; Villagran, D.; Nocera, D. G., Iron in a trigonal tris(alkoxide) ligand environment. *Inorg. Chem.* **2013,** *52* (6), 3159-69.
72. Lu, T. T.; Chiou, S. J.; Chen, C. Y.; Liaw, W. F., Mononitrosyl tris(thiolate) iron complex [Fe(NO)(SPh)3]- and dinitrosyl iron complex [(EtS)2Fe(NO)2]-: formation pathway of dinitrosyl iron complexes (DNICs) from nitrosylation of biomimetic rubredoxin [Fe(SR)4]2-/1- (R = Ph, Et). *Inorg. Chem.* **2006,** *45* (21), 8799-806.
73. DeRosha, D. E.; Chilkuri, V. G.; Van Stappen, C.; Bill, E.; Mercado, B. Q.; DeBeer, S.; Neese, F.; Holland, P. L., Planar three-coordinate iron sulfide in a synthetic [4Fe-3S] cluster with biomimetic reactivity. *Nat. Chem.* **2019,** *11* (11), 1019-1025.
74. McDonald, A. R.; Bukowski, M. R.; Farquhar, E. R.; Jackson, T. A.; Koehntop, K. D.; Seo, M. S.; De Hont, R. F.; Stubna, A.; Halfen, J. A.; Munck, E.; Nam, W.; Que, L., Jr., Sulfur versus iron oxidation in an iron-thiolate model complex. *J. Am. Chem. Soc.* **2010,** *132* (48), 17118-29.
75. Villar-Acevedo, G.; Lugo-Mas, P.; Blakely, M. N.; Rees, J. A.; Ganas, A. S.; Hanada, E. M.; Kaminsky, W.; Kovacs, J. A., Metal-Assisted Oxo Atom Addition to an Fe(III) Thiolate. *J. Am. Chem. Soc.* **2017,** *139* (1), 119-129.
76. Helbig, K.; Bleuel, C.; Krauss, G. J.; Nies, D. H., Glutathione and transition-metal homeostasis in Escherichia coli. *J. Bacteriol.* **2008,** *190* (15), 5431-8.

77. Mulder, D. W.; Shepard, E. M.; Meuser, J. E.; Joshi, N.; King, P. W.; Posewitz, M. C.; Broderick, J. B.; Peters, J. W., Insights into [FeFe]-hydrogenase structure, mechanism, and maturation. *Structure* **2011,** *19* (8), 1038-52.
78. Liu, H.; Naismith, J. H., An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **2008,** *8*, 91.
79. Rittle, J.; Field, M. J.; Green, M. T.; Tezcan, F. A., An efficient, step-economical strategy for the design of functional metalloproteins. *Nat. Chem.* **2019,** *11* (5), 434-441.
80. Arslan, E.; Schulz, H.; Zufferey, R.; Kunzler, P.; Thony-Meyer, L., Overproduction of the Bradyrhizobium japonicum c-type cytochrome subunits of the cbb3 oxidase in Escherichia coli. *Biochem. Biophys. Res. Commun.* **1998,** *251* (3), 744-7.
81. Ames, G. F.; Prody, C.; Kustu, S., Simple, rapid, and quantitative release of periplasmic proteins by chloroform. *J. Bacteriol.* **1984,** *160* (3), 1181-3.
82. Lebowitz, J.; Lewis, M. S.; Schuck, P., Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* **2002,** *11* (9), 2067-79.
83. Stoll, S.; Schweiger, A., EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **2006,** *178* (1), 42-55.
84. Kabsch, W., Xds. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 125-32.
85. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994,** *50* (Pt 5), 760-3.
86. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta. Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-21.
87. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004,** *60* (Pt 12 Pt 1), 2126-32.

Chapter 4: Redox- and metal-directed structural diversity in designed protein assemblies

4.1 Abstract

Herein we describe a designed protein building block whose self-assembly behavior is dually gated by the redox state of disulfide bonds and the identity of exogenous metal ions. This protein construct is shown–through extensive structural and biophysical characterization–to access five distinct oligomeric states, exemplifying how the complex interplay between hydrophobic, metal-ligand, and reversible covalent interactions could be harnessed to obtain multiple, responsive protein architectures from a single building block.

4.2 Introduction

The propensity of a single protein sequence to form multiple conformations or assembly states has been crucial for the generation of structural and functional diversity during evolution.[1-3] For instance, protein folds such as the Rossman,[4] four-helix bundle,[5] and βαββ motifs[6] have been repeatedly used as modular building blocks for larger architectures or quaternary assemblies with a wide variety of functions. Similarly, obtaining multiple structural outcomes from a single protein sequence also is a prerequisite for building switchable systems that transduce external stimuli into functionally relevant changes to their tertiary folds or quaternary assembly states.[7-12] These processes are underpinned by reversible interactions between proteins and external stimuli, which include ligand binding,[13, 14] metal coordination,[15, 16] phosphorylation,[17, 18] and cysteine oxidation/reduction.[19, 20] While most designed protein switches process a single type of stimulus to toggle between two structural states,[3, 13-18, 21-23] the ability to design proteins that respond to more than one type of stimulus or to obtain more than two structurally distinct states from a single protein sequence/structure has been limited (**Figure 4.1a**).[24] This is primarily due to the fact that most protein design strategies involve the implementation of extensive noncovalent interactions

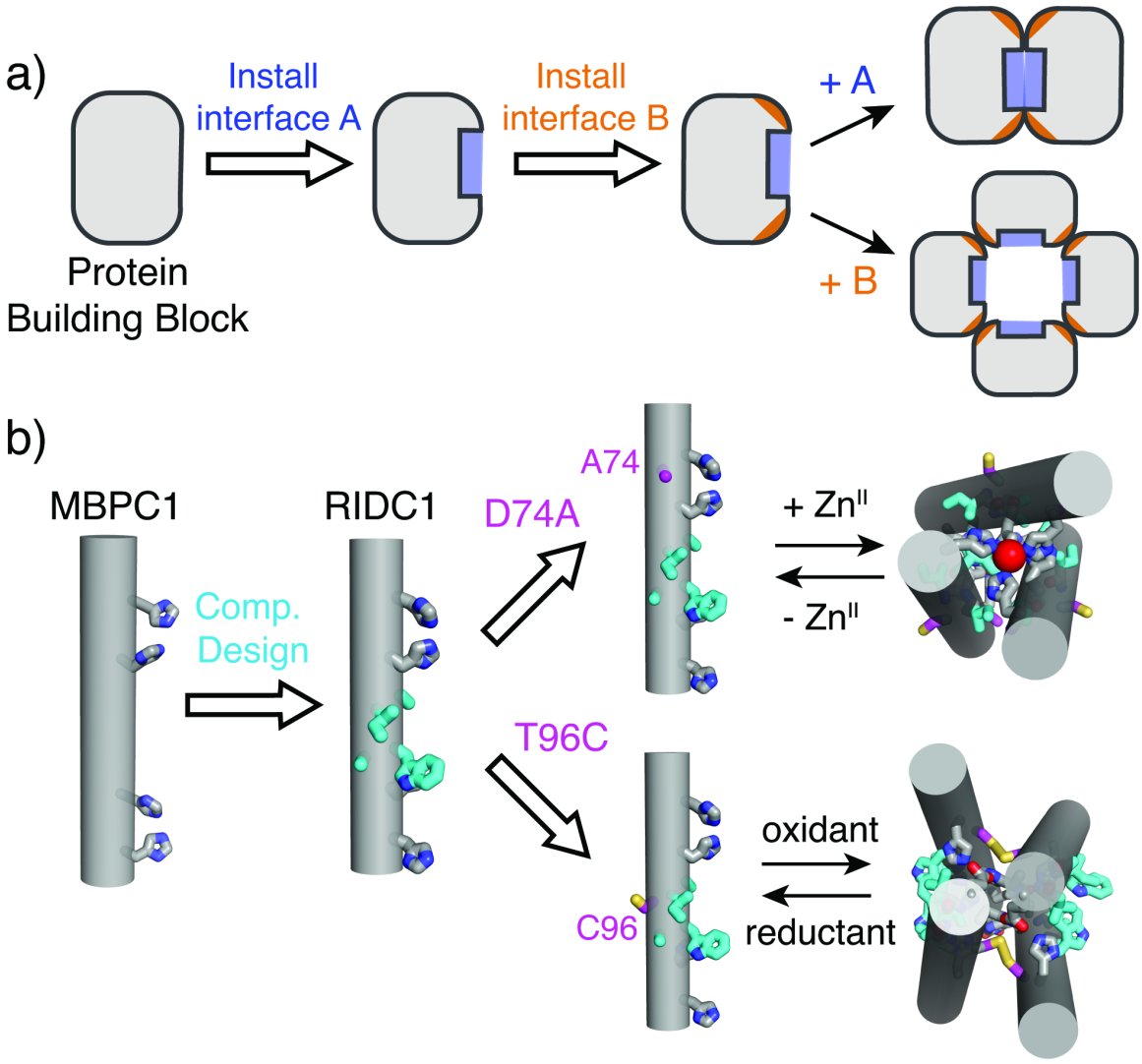

Figure 4.1 | Engineering multi-stimuli responsiveness in protein assemblies. (a) General workflow to design a multi-stimuli-responsive protein construct. A and B represent different stimuli. (b) Cartoon schemes of previously designed cytochrome *cb*$_{562}$ variants. Hydrophobic surface mutations are highlighted in cyan.

(in particular, hydrophobic packing) to obtain single, stable structures that correspond to deep free energy minima.[25-27] This strategy lowers the potential for structural diversification and the ability of the resulting protein architecture to be stimuli-responsive and reconfigurable.

Due to their simultaneous strength and reversibility, metal-ligand and disulfide bonding interactions represent promising conduits for the design of dynamic, multistate protein switches.[20,

[28-31] Our group has previously exploited metal coordination, disulfide bonding, and hydrophobic packing to construct cytochrome (cyt) cb_{562}-based assemblies with diverse conformations, oligomeric states and metal coordination environments.[15, 32-37] MBPC1, an early designed variant of cyt cb_{562}, was shown to form different assemblies with distinct oligomeric states and structures based on the coordination preferences of exogenously added metal ions.[15, 32] One of these assemblies, the Zn-directed tetramer Zn_4:$MBPC1_4$, served as a structural template for the computational design of a hydrophobic interface (highlighted in cyan in **Figure 4.1b, Figure 4.2a**) on the surface of MBPC1.[34] Owing to the designed interactions between the hydrophobic surface residues, the resulting variant, RIDC1, formed a considerably more stable Zn-directed tetramer (Zn_4:$RIDC1_4$) with a nearly identical structure to that of Zn_4:$MBPC1_4$ (**Figure 4.2b**).[34] Importantly, Zn_4:$RIDC1_4$ served as a starting point for designing assemblies with functions that

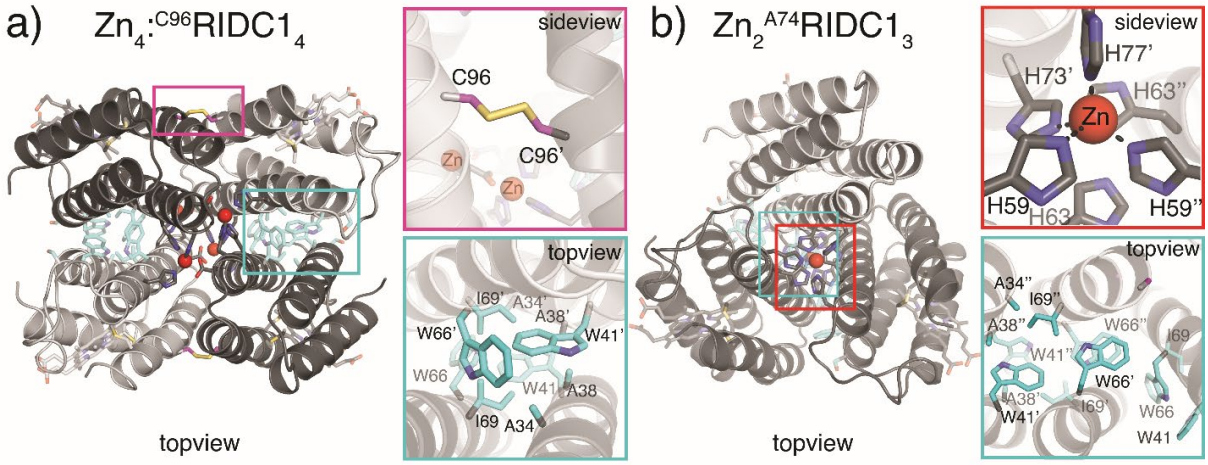

Figure 4.2 | Crystal structures of previously designed RIDC1 variants. (a) Crystal structure of Zn_4:$^{C96}RIDC1^{ox}_4$ (PDB ID: 3IQ6). One of the two C96-C96 disulfide bonds is highlighted in magenta while the designed hydrophobic residues are highlighted in cyan. (b) Crystal structure of Zn_2:$^{A74}RIDC1_3$ (PDB ID: 3M15). One of the two Zn:His_4 coordination sites is highlighted in red while the designed hydrophobic residues are highlighted in cyan. ^{C96}RIDC1 enforces the tetrameric architecture through covalent preorganization of the *i1* interface while ^{A74}RIDC1 features a decoupling of Zn binding to tetramerization. This decoupling takes place despite the presence of hydrophobic residues installed to stabilize a tetramer, which suggests that Zn coordination by D74 plays a valuable role in directing the tetramerization of RIDC1.

ranged from selective metal binding and metal-based allostery to *in vivo* enzymatic activity.[35, 38, 39]

In the course of our previous studies, we observed that single mutations of the RIDC1 construct alter the assembly outcomes.[40, 41] One RIDC1 variant, ^{C96}RIDC1, formed a redox-dependent but metal-independent tetramer (Zn_4:RIDC1$_4$) stabilized by both hydrophobic and Cys96-Cys96 disulfide bonding interactions (**Figure 4.1b**). A second RIDC1 variant, ^{A74}RIDC1 (wherein the metal binding residue Asp74 was mutated to Ala), assembled into a Zn-dependent trimer (Zn_2:^{A74}RIDC1$_3$) stabilized by hydrophobic packing interactions and tetrahedral, Zn:His$_4$ coordination sites that had not been observed in other RIDC1 variants (**Figure 4.1b**, **Figure 4.2b**).[41] Here, with the aim of developing a protein construct that can respond to redox and metal-based stimuli to access multiple structural states, we combined the A74 and C96 mutations to

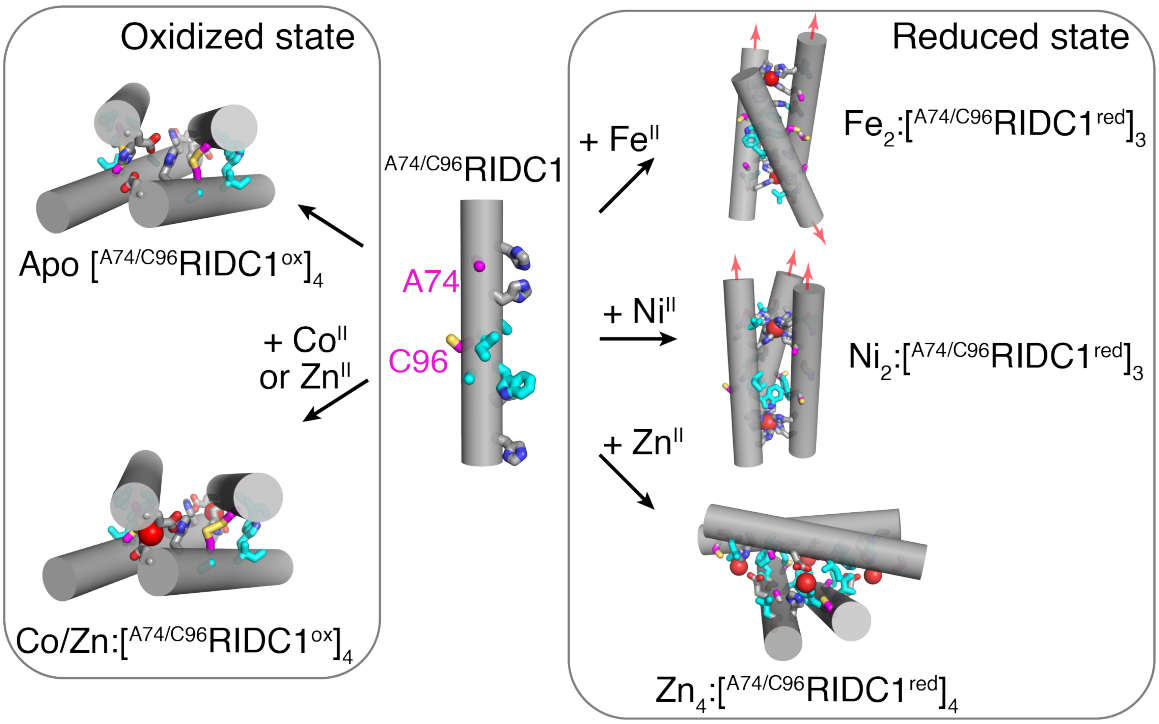

Figure 4.3 | Structural states of $^{A74/C96}$RIDC1 obtained through the addition of redox and/or metal-based stimuli. Hydrophobic mutations are highlighted in cyan. Each cylinder represents a protein monomer.

generate $^{A74/C96}$RIDC1 (**Figure 4.3**). We found that the interplay between hydrophobic, metal-ligand, and covalent interactions enabled this variant to form five discrete structural states in a redox- and metal-responsive fashion (**Figure 4.3**).

4.3 Results and Discussion

4.3.1 Assembly properties of $^{A74/C96}$RIDC1^{ox}

We surmised that in the oxidized state of $^{A74/C96}$RIDC1 ($^{A74/C96}$RIDC1^{ox}), the Cys96-Cys96 disulfide bonds would enforce tetramerization, as observed in the case of ^{C96}RIDC1.[40] Indeed, both in the absence and presence of metal ions (Co^{II}, Ni^{II}, Cu^{II}, Zn^{II}), $^{A74/C96}$RIDC1^{ox} formed a tetrameric species in solution in near quantitative yields as determined by sedimentation velocity- analytical ultracentrifugation (SV-AUC) measurements (**Figure 4.4**). We were also able to confirm tetramerization in solution using native electrospray ionization mass spectrometry (ESI-MS, **Figure 4.5**). The crystal structures of Co^{II}-, and Zn^{II}-bound [$^{A74/C96}$RIDC1^{ox}]$_4$ are nearly identical to one another, with a root-mean-square deviation (RMSD) of 1.24 Å between all α-C's (**Figure 4.6-4.8)**. On average, the buried surface area (BSA) of the metal-bound tetramers is about 40% smaller (1018 Å^2) than that of the apo structure (1388 Å^2) (**Table 4.1**). This indicates that the tetrameric assembly undergoes a structural change upon metal binding, with an average RMSD of 2.55 Å between apo and metal-bound structures (**Figure 4.6**). Both Co and Zn-bound [$^{A74/C96}$RIDC1^{ox}]$_4$ tetramers contain two C_2-symmetry-related coordination sites, with E81 and H77 residues from two different monomers serving as ligands. (**Figure 4.7-4.9**). The [$^{A74/C96}$RIDC1^{ox}]$_4$ structures illustrate that simultaneously exploiting the flexibility of and structural constraints imposed by disulfide bonds and hydrophobic packing interactions can engender a flexible protein assembly with well-defined metal coordination sites (**Figure 4.3, 4.7-4.9**).

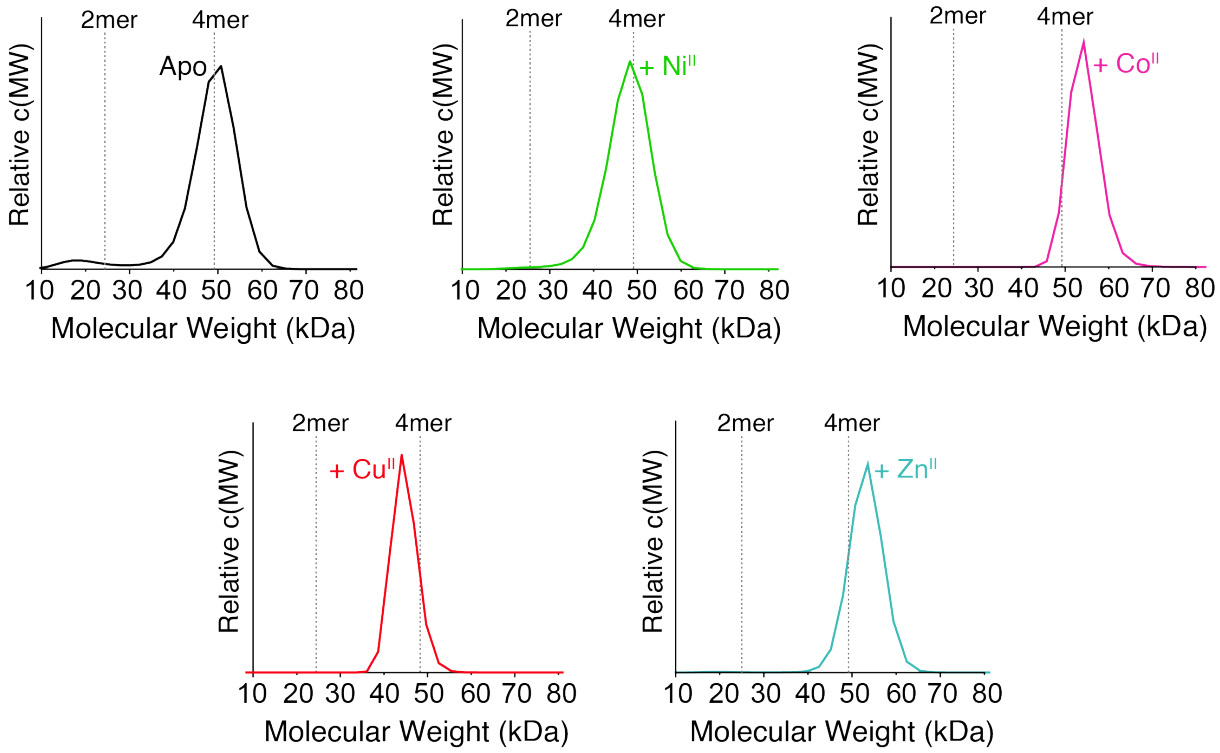

Figure 4.4 | SV-AUC distributions of metal-supplemented $^{A74/C96}$RIDC1^{ox}. Analysis was performed with 200 μM protein monomer and 1 equivalent metal salt.

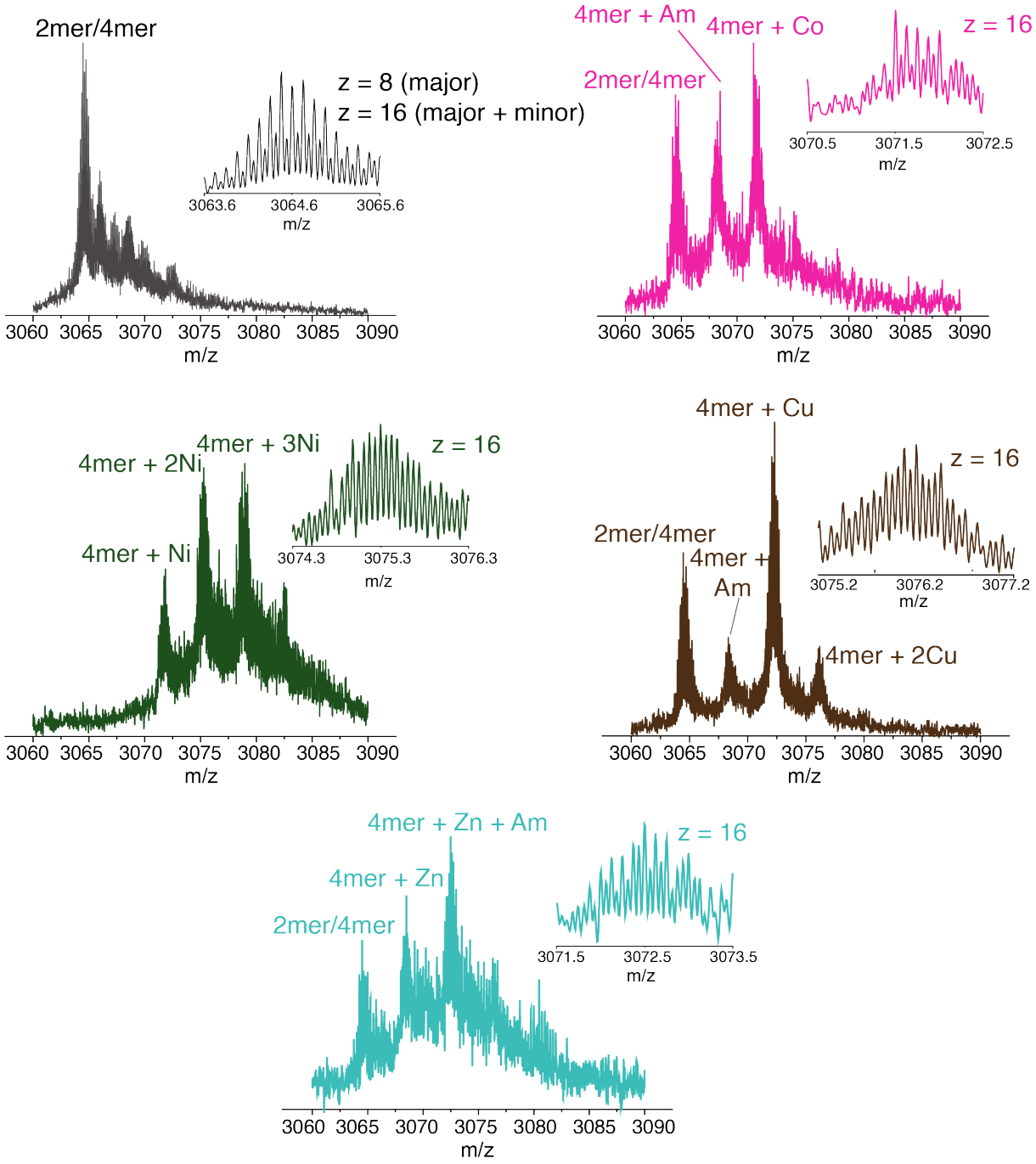

Figure 4.5 | Native ESI-MS distributions of metal-supplemented $^{A74/C96}$RIDC1^{ox}. Analysis was performed with 50 μM protein monomer and 1 equivalent metal salt. Distributions with a charge state of +8 correspond to a dimer, while those with a charge state of +16 correspond to a tetramer.

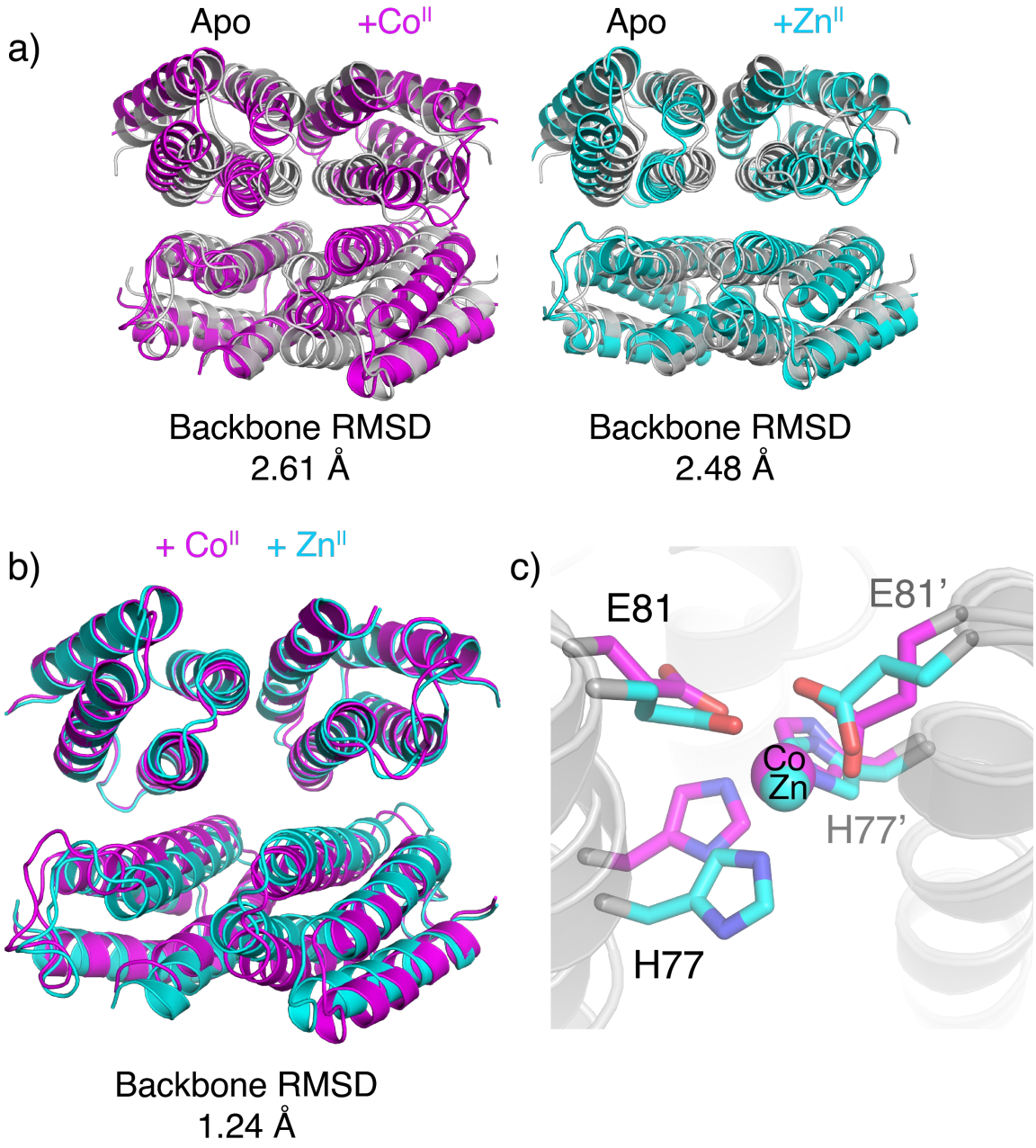

Figure 4.6 | Crystal structures of metal-bound $^{A74/C96}$RIDC1^{ox}. (a) Overlays of Apo, Co^{II}-, and Zn^{II}-bound $^{A74/C96}$RIDC1^{ox} crystal structures (b) Overlay of Co^{II}-, and Zn^{II}-bound $^{A74/C96}$RIDC1^{ox} crystal structures. (c) Overlay of C_2 coordination sites of the metal-loaded assemblies, illustrating close correlation in the positions of coordinating residues.

Table 4.1 | Buried surface area (BSA) calculations of RIDC1 assemblies[a]

Variant	Reductant	M^{II}	BSA/monomer ($Å^2$)
$^{A74/C96}$RIDC1	-	-	1665
$^{A74/C96}$RIDC1	-	Co	968
$^{A74/C96}$RIDC1	-	Zn	1068
$^{A74/C96}$RIDC1	THPP	Fe	1038
$^{A74/C96}$RIDC1	THPP	Ni	768
$^{A74/C96}$RIDC1	THPP	Cu	1743
$^{A74/C96}$RIDC1	THPP	Zn	1650
^{C96}RIDC1	-	-	1063
^{C96}RIDC1	-	Zn	1388
^{A74}RIDC1	-	Zn	1007

[a] BSA calculations based on crystal structures

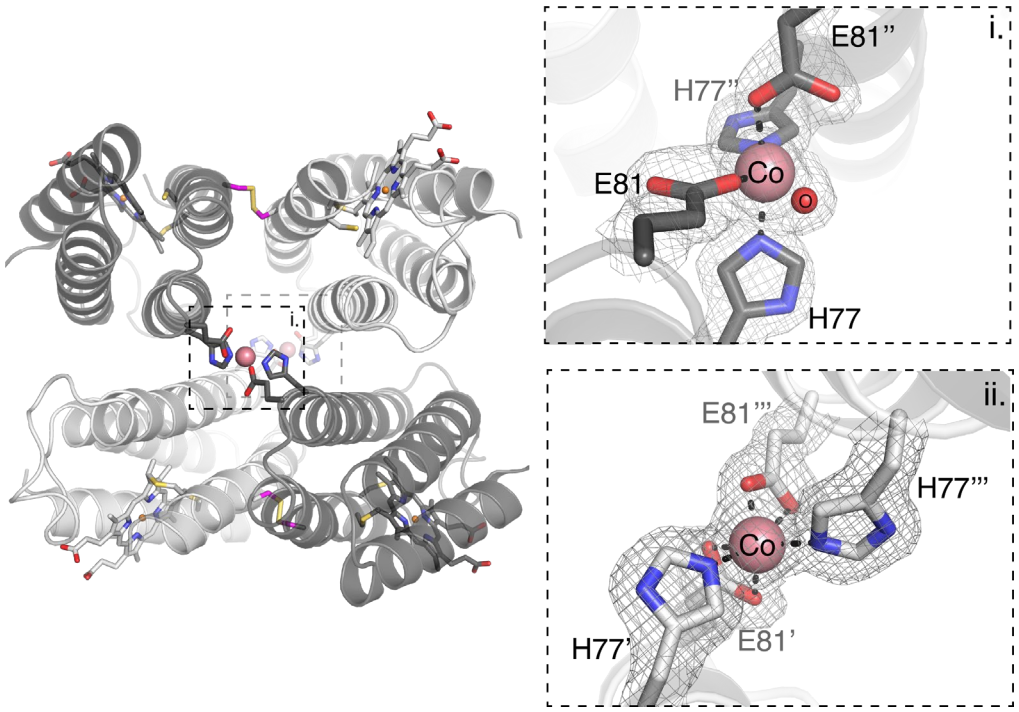

Figure 4.7 | Overview of Co_2:[$^{A74/C96}$RIDC1^{ox}]$_4$ crystal structure, including 2Fo-2Fc maps of metal coordination sites. All electron density maps were generated in CCP4i.

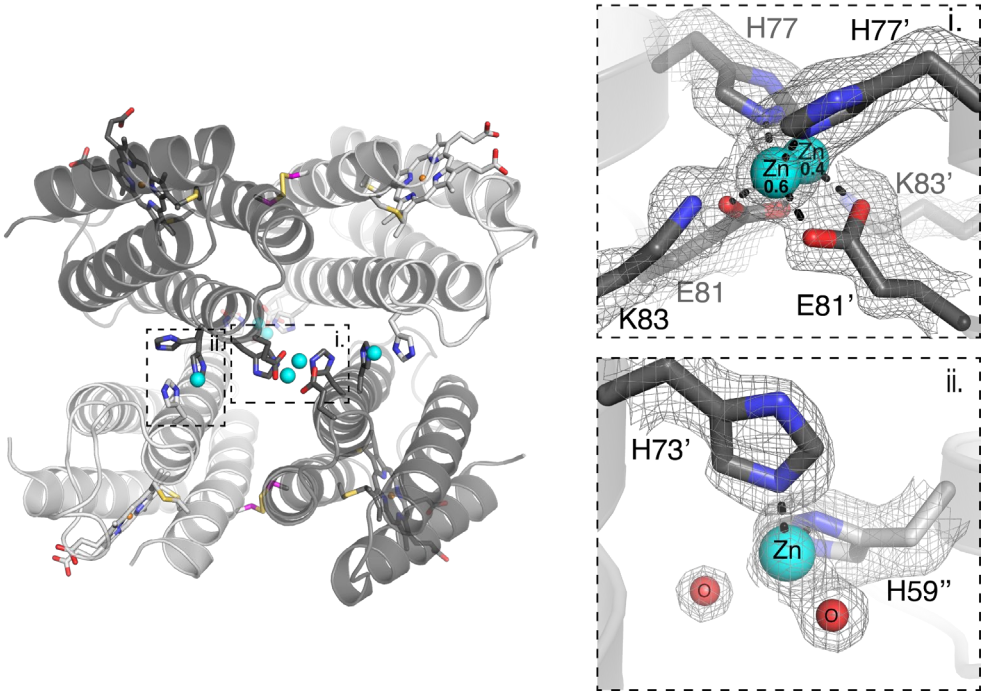

Figure 4.8 | Overview of Zn_4:[$^{A74/C96}$RIDC1^{ox}]$_4$ crystal structure, including 2Fo-2Fc maps of metal coordination sites. All electron density maps were generated in CCP4i. At the C_2 coordination site, Zn occupies two discrete positions with occupancies of 0.6 and 0.4.

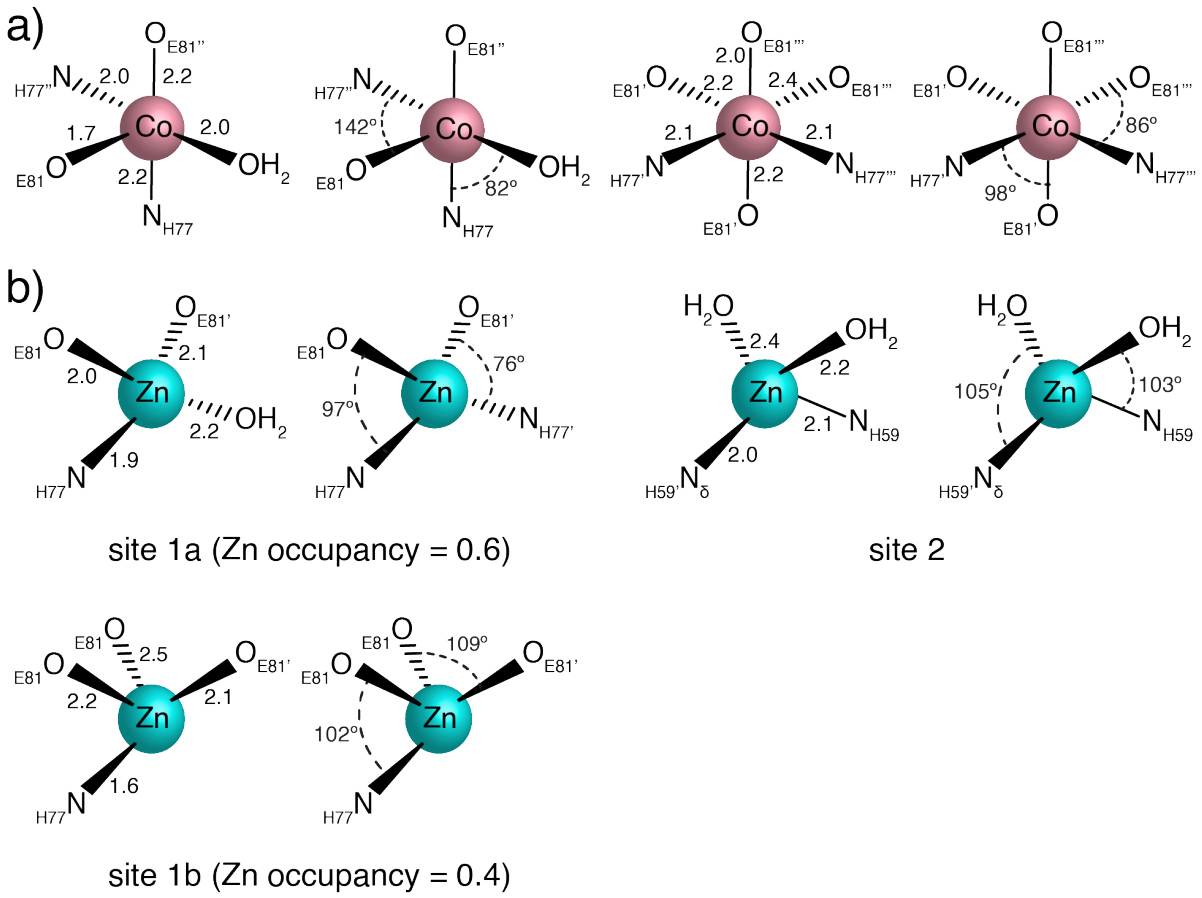

Figure 4.9 | Geometries of coordination sites of Co- and Zn-bound $^{A74/C96}$RIDC1^{ox}, including bond distances and angles. (a) Coordination geometries of Co-bound $^{A74/C96}$RIDC1^{ox}, (b) Coordination geometries of Zn-bound $^{A74/C96}$RIDC1^{ox}.

4.3.2 Assembly properties of $^{A74/C96}$RIDC1^{red}

We next turned to the reduced form of our construct, $^{A74/C96}$RIDC1^{red}, with the hypothesis that the lack of disulfide-mediated interfacial constraints could allow it to access different oligomeric states upon metal coordination. $^{A74/C96}$RIDC1^{red} was obtained by adding 5-fold excess of the reductant tris(3-hydroxylpropyl) phosphine (THPP) to $^{A74/C96}$RIDC1^{ox}. At a protein concentration of ≥200 μM and up to 5-fold excess of metal ions, Fe^{II}, Ni^{II} and Cu^{II} addition to $^{A74/C96}$RIDC1^{red} led primarily to trimeric species in solution, whereas Zn^{II} and Co^{II} addition yielded tetrameric or higher-order assemblies (**Figure 4.10, Table 4.2**). All metal-directed $^{A74/C96}$RIDC1^{red}

oligomers could be completely disassembled by the addition of a mixture of ethylenediaminetetraacetic acid (EDTA) and dipicolinic acid (DPA) (**Figure 4.10**).

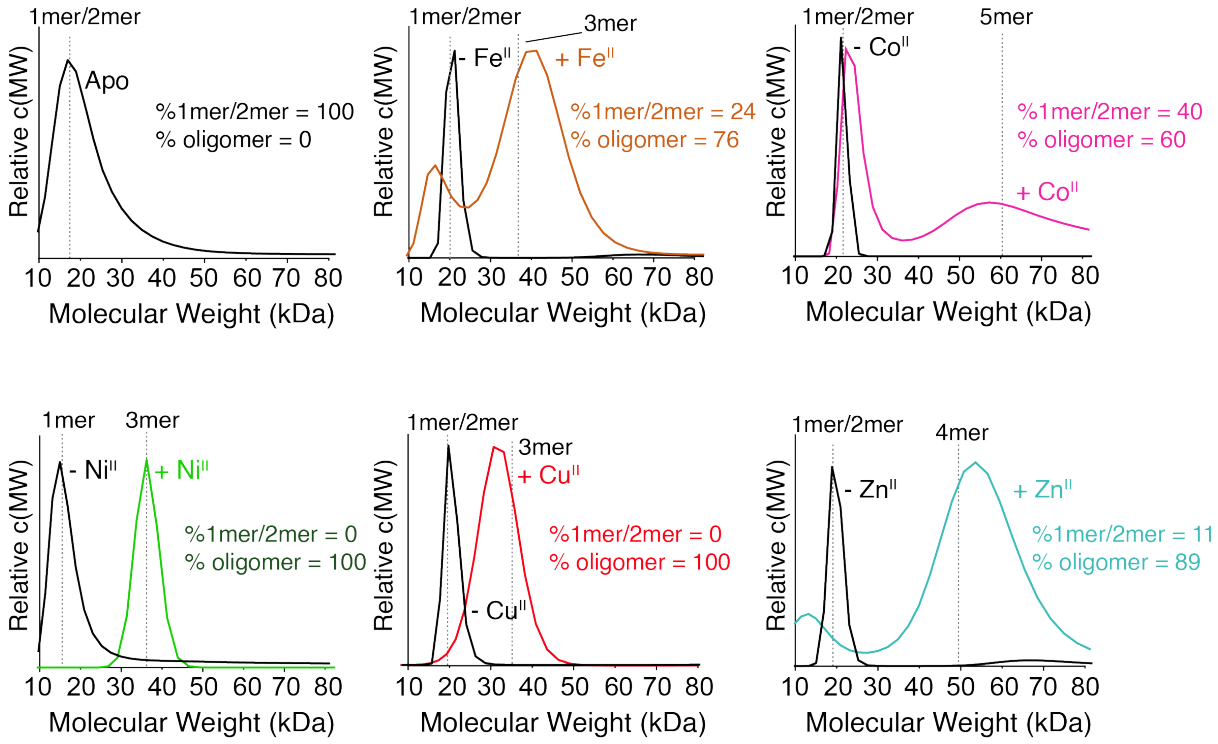

Figure 4.10 | SV-AUC distributions of $^{A74/C96}$RIDC1^{red} upon the addition/removal of metal ions. Analysis was performed with 200 μM protein monomer and 1 equivalent (Co^{II}, Ni^{II}, Cu^{II}, Zn^{II}) or 5 equivalents (Fe^{II}) metal salt. Metal ions were removed from the protein via the addition of 10 mM EDTA/DPA (black traces)**.**

Table 4.2 | SV-AUC parameters for apo, metal-loaded $^{A74/C96}$RIDC1[a]

Reductant[b]	Metal[c]	Frictional Ratio (f/f_0)	MW_{max} (kDa)[d]
-	-	1.31	51
-	$CoCl_2$	1.20	54
-	$NiCl_2$	1.20	48
-	$CuCl_2$	1.21	44
-	$ZnCl_2$	1.12	54
THPP	-	1.20	17
THPP	$FeSO_4$[e]	1.37	41
THPP	$CoCl_2$	1.16	22
THPP	$NiCl_2$	1.21	33
THPP	$CuCl_2$	1.21	31
THPP	$ZnCl_2$	1.19	54

[a] [protein monomer] = 0.2 mM monomer
[b] [THPP] = 1 mM
[c] Unless otherwise indicated, [metal] = 0.2 mM
[d] Theoretical MW = 12258 (monomer), 24516 (dimer), 36774 (trimer), 49032 (tetramer), 61305 (pentamer)
[e] [$FeSO_4$] = 1 mM

The crystal structures of Fe^{II}-, Ni^{II}-, Cu^{II}-, and Zn^{II}-directed assemblies of $^{A74/C96}$RIDC1^{red} were determined at resolutions of 1.6 Å to 2.7 Å (**Table 4.3**). These structures revealed a correspondence between the oligomerization states observed in solution and crystals for the Fe^{II} (n=3), Ni^{II} (n=3) and Zn^{II} (n=4) complexes (**Figure 4.11-4.13**). By contrast, there was a deviation in the case of the Cu^{II}-directed $^{A74/C96}$RIDC1^{red} assembly (n=4 in crystals vs. n=3 in solution) (**Figure 4.14**). A closer look at the latter structure showed that all four Cu centers in the tetrameric assembly adopted a tetrahedral coordination geometry, strongly suggesting that they were in the +1 oxidation state and thus reduced by the excess THPP present in the crystallization solution (**Figure 4.14-4.15, Table 4.4**). In light of the complex redox equilibrium that exists between Cu ions, Cys-disulfide bonds, and THPP, we decided to focus our further analyses on Fe^{II}, Ni^{II} and Zn^{II} complexes of $^{A74/C96}$RIDC1^{red}.

Table 4.3 | X-ray refinement statistics for $^{A74/C96}$RIDC1 crystal structures (numbers in parentheses correspond to values in the highest resolution shell).

Variant	Apo $[^{A74/C96}RIDC1^{ox}]_4$	Co_2: $[^{A74/C96}RIDC1^{ox}]_4$	Zn_4: $[^{A74/C96}RIDC1^{ox}]_4$	Fe_2: $[^{A74/C96}RIDC1^{red}]_3$
PDB ID	7RWV	7SU2	7RWW	7RWY
Space group	*I*4	$P2_1$	$P2_1$	$P2_1$
Cell dimensions (Å)	92.58, 92.58, 91.26	47.88 90.02 52.04	47.79, 86.91, 49.18	52.55, 82.68, 74.14
Cell angles (°)	90.00, 90.00, 90.00	90.00 95.91 90.00	90.00, 109.16, 90.00	90.00, 95.22, 90.00
Resolution (Å)	46.29-2.20	37.00-2.00	35.96-1.70	44.65-2.20
No. unique reflections	19413	29302	40172	30358
R_{merge}	0.150 (0.544)	0.017 (0.059)	0.115 (0.752)	0.096 (0.374)
Multiplicity	2.5 (2.6)	2.0 (2.0)	14.8 (10.9)	3.9 (3.8)
CC½	0.984 (0.363)	1 (0.989)	0.996 (0.617)	0.992 (0.863)
$< I / \sigma(I) >$	3.7 (1.9)	26.7 (7.4)	11.7 (3.2)	8.0 (2.8)
Completeness (%)	98.0 (98.7)	98.4 (96.0)	99.9 (99.5)	99.1 (99.7)
Refinement				
R_{work}/R_{free}	0.1946/0.2520	0.1904/0.1915	0.1704/0.2020	0.2023/0.2592
B-factors (Å^2)	34.36	34.32	29.84	39.71
Protein	34.76	34.37	29.15	39.64
Ligand/ion	28.36	29.71	29.63	39.86
Solvent	32.76	37.65	39.39	41.08
R.m.s deviations				
Bond lengths (Å)	0.016	0.009	0.012	0.009
Bond angles (°)	1.871	1.150	1.123	1.159
Clashscore	16.87	7.63	10.23	19.73
Ramachandran plot (%)				
Favored	99.76	97.84	99.76	97.44
Outliers	0.00	0.00	0.00	0.16
Rotamer outliers (%)	0.00	2.60	0.29	1.99

Table 4.3 (cont.) | X-ray refinement statistics for $^{A74/C96}$RIDC1 crystal structures (numbers in parentheses correspond to values in the highest resolution shell).

Variant	Ni_2: $[^{A74/C96}RIDC1^{red}]_3$	Cu_4: $[^{A74/C96}RIDC1^{red}]_4$	Zn_4: $[^{A74/C96}RIDC1^{red}]_4$
PDB ID	7RWU	7TEP	7RWX
Space group	$P4_132$	$P2_1$	R 3 2
Cell dimensions (Å)	94.26, 94.26, 94.26	47.08 80.57 49.44	111.65, 111.65, 148.33
Cell angles (°)	90.00, 90.00, 90.00	90.00 102.33, 90.00	90.00, 90.00, 120.00
Resolution (Å)	42.15-1.80	45.99-2.70	37.01-1.60
No. unique reflections	25101	9589	27879
R_{merge}	0.069 (0.191)	0.0571 (0.146)	0.055
Multiplicity	68.0 (68.5)	3.3 (3.3)	4.1 (3.7)
CC ½	1.000 (0.998)	0.998 (0.953)	0.995 (0.898)
$< I / \sigma(I) >$	58.1 (28.2)	16.66 (6.50)	8.5 (2.7)
Completeness (%)	100.0 (100.0)	95.9 (89.6)	96.2 (91.3)
Refinement			
R_{work}/R_{free}	0.1641/0.1892	0.2536/0.3385	0.2369/0.2897
B-factors (Å^2)	23.60	28.62	50.21
Protein	21.49	28.66	50.60
Ligand/ion	16.75	28.86	44.24
Solvent	36.29	26.39	49.03
R.m.s deviations			
Bond lengths (Å)	0.010	0.019	0.008
Bond angles (°)	0.994	1.960	1.006
Clashscore	15.07	51.51	10.12
Ramachandran plot (%)			
Favored	99.04	77.88	98.08
Outliers	0.00	6.75	0.00
Rotamer outliers (%)	0.00	10.5	4.22

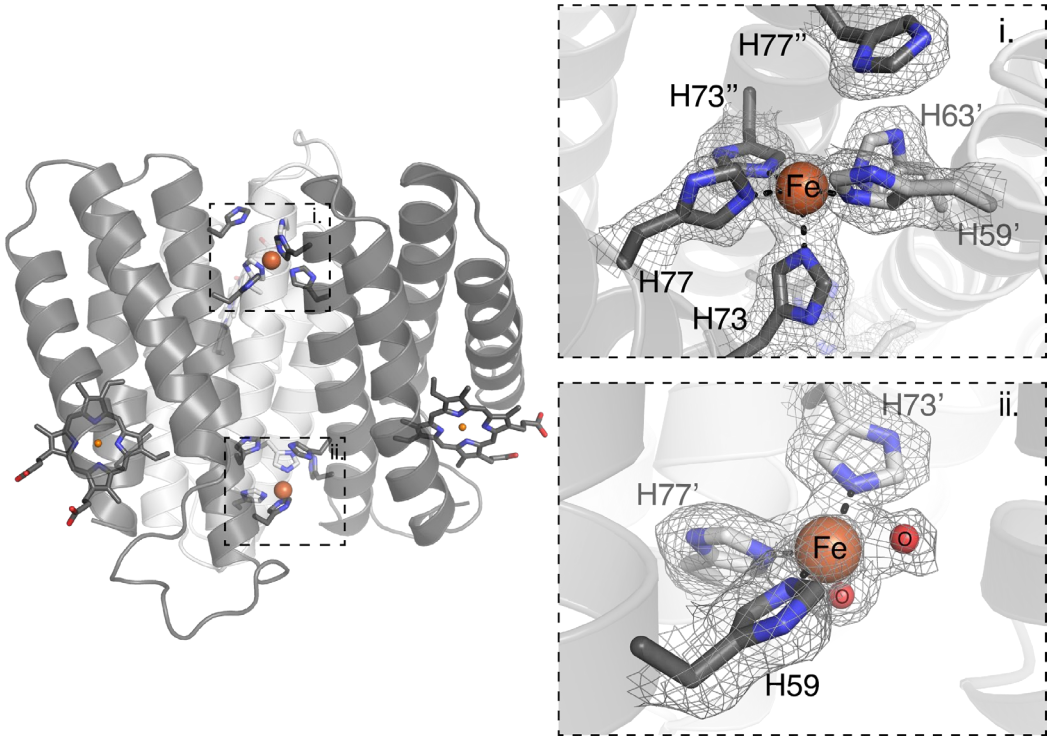

Figure 4.11 | Overview of Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$ crystal structure, including 2Fo-2Fc maps of metal coordination sites. All electron density maps were generated in CCP4i.

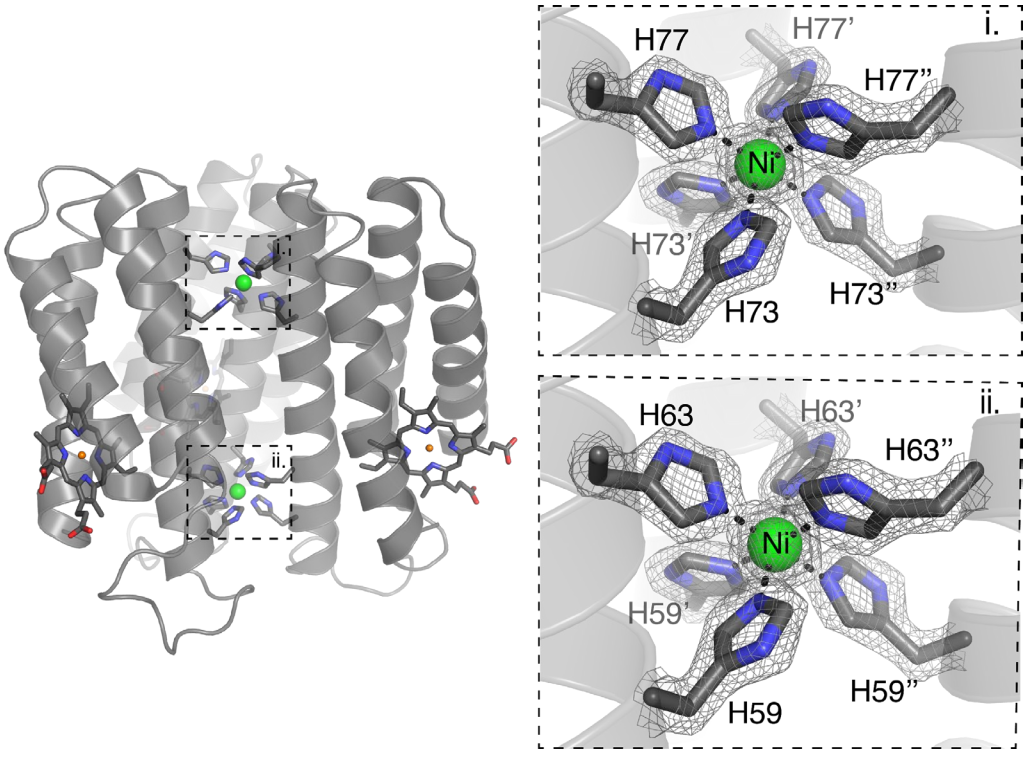

Figure 4.12 | Overview of Ni_2:[$^{A74/C96}$RIDC1^{red}]$_3$ crystal structure, including 2Fo-2Fc maps of metal coordination sites. All electron density maps were generated in CCP4i.

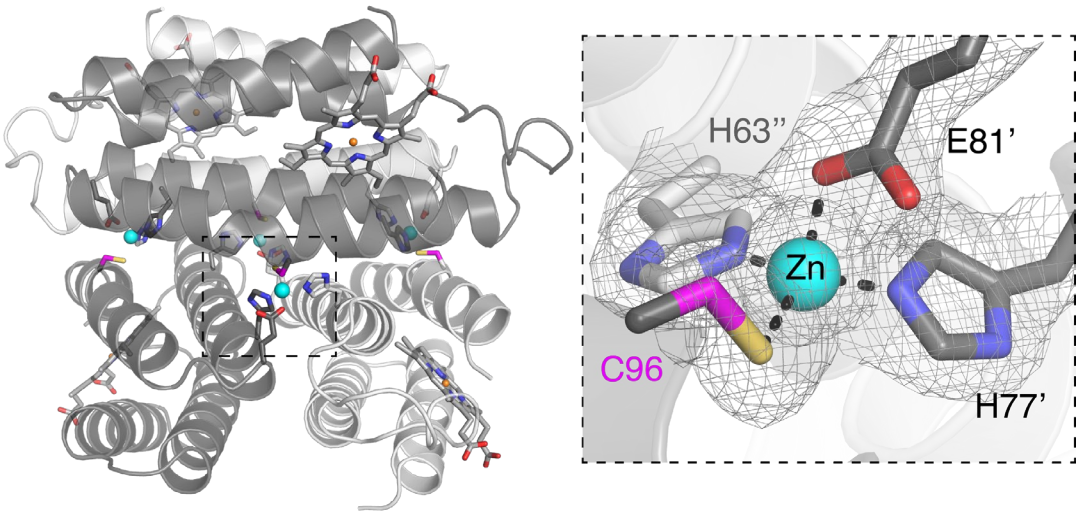

Figure 4.13 | Overview of Zn_4:[$^{A74/C96}$RIDC1^{red}]$_4$ crystal structure, including 2Fo-2Fc maps of metal coordination sites. All electron density maps were generated in CCP4i.

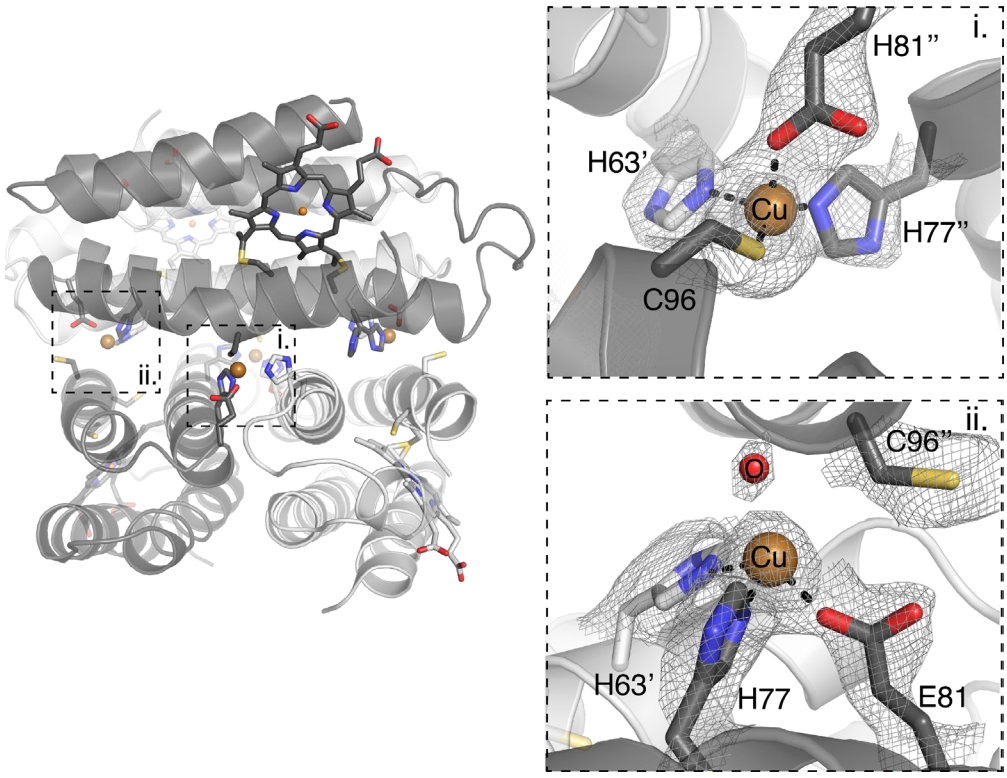

Figure 4.14 | Overview of the Cu_4:[$^{A74/C96}$RIDC1^{red}]$_4$ crystal structure, including 2Fo-2Fc maps of metal coordination sites. The tetrahedral coordination of the Cu ions suggests *in situ* reduction of copper ions by THPP to Cu^{I}, which is more frequently observed in tetrahedral coordination geometries.

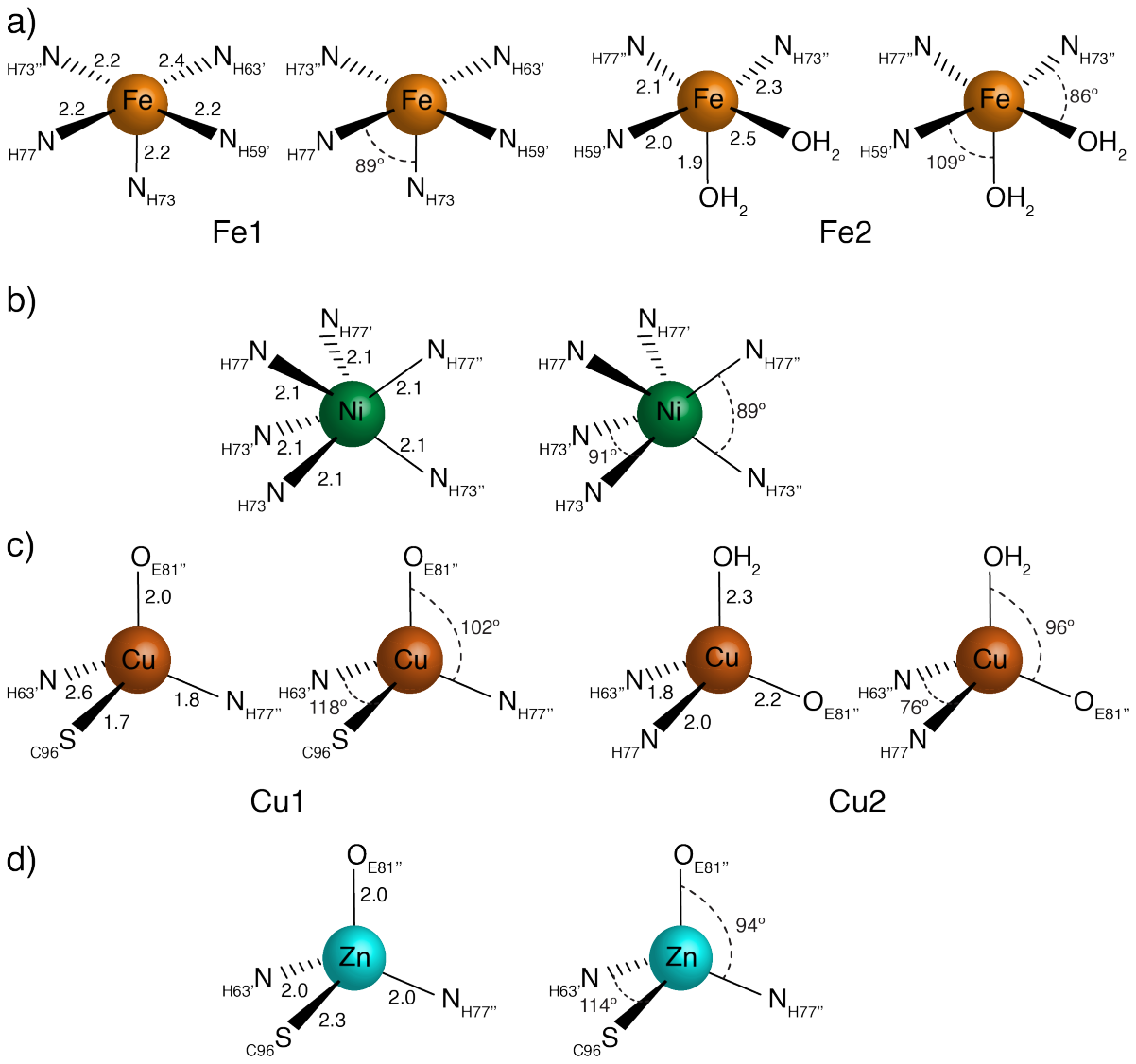

Figure 4.15 | Geometries of coordination sites of metal-bound $^{A74/C96}$RIDC1^{red}, including bond distances and angles. (a) Coordination geometries of Fe-bound $^{A74/C96}$RIDC1^{red}, (b) Coordination geometries of Ni-bound $^{A74/C96}$RIDC1^{red}, (c) Coordination geometries of Cu-bound $^{A74/C96}$RIDC1^{red}, (d) Coordination geometries of Zn-bound $^{A74/C96}$RIDC1^{red}.

Table 4.4 | Crystallization conditions for apo, metal-loaded $^{A74/C96}$RIDC1

Reductant[a]	Metal	[Protein][b]	[Metal]	Mother liquor
-	-	2.6 mM	-	30% PEG400, 0.1 M HEPES pH 7.5, 0.2 M $MgCl_2$
-	$CoCl_2$	3.0 mM	3.0 mM	30% PEG400, 0.1 M HEPES pH 7.5, 0.2 M NaCl
-	$ZnCl_2$	2.6 mM	2.6 mM	45% MPD, 0.1 M HEPES pH 7.5, 0.2 M $MgCl_2$
THPP[a]	$FeSO_4$	2.8 mM	9.0 mM	40% PPG, 0.1 M Bis-Tris 6.5, no salt
THPP[a]	$NiCl_2$	2.6 mM	4.8 mM	45% MPD, 0.1 M Tris pH 8.5, 0.2 M $MgCl_2$
THPP[a]	$CuCl_2$	3.0 mM	3.0 mM	25% PEG1500, 0.1 M Bis-Tris pH 6.5, 0.2 M AmAc
THPP[a]	$ZnCl_2$	2.6 mM	5.0 mM	30% PEG400, 0.1 M HEPES pH 7.5, 0.2 M $CaCl_2$

[a] [THPP] = 10 mM

[b] Indicates monomer concentration

Interestingly, the Fe^{II}- and Ni^{II}-directed assemblies, while both trimeric, adopt different structural conformations and metal coordination environments (**Figure 4.3, 4.11-4.12, 4.15**). The Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$ complex features two protein monomers with their C-termini projecting downward and one monomer with its C-terminus projecting upward, resulting in an antiparallel, "up-up-down" arrangement similar to that observed for Zn_2:^{A74}RIDC1$_3$ (**Figure 4.3, 4.11, 4.16**).[41] The two Fe^{II} centers, termed Fe1 and Fe2, are distinct from one another, with Fe1 in a square pyramidal geometry formed by five His residues and Fe2 in a similar geometry but with three His and two H_2O ligands (**Figure 4.11, 4.15a**). By contrast, the Ni_2:[$^{A74/C96}$RIDC1^{red}]$_3$ assembly has C_3 symmetry with an all-parallel, "up-up-up" arrangement of protein monomers and two octahedral, hexa-His-coordinated Ni^{II} centers (**Figure 4.3, 4.12, and 4.15b**). The structure of Ni_2:[$^{A74/C96}$RIDC1^{red}]$_3$ is nearly identical to that of the previously characterized Ni_2:MBPC1$_3$ assembly (RMSD = 0.66 Å, **Figure 4.16**). While buried surface area and Rosetta interface calculations predict that the "up-up-up" trimer is less stable compared to the "up-up-down" configuration based purely on interfacial hydrophobic interactions (**Table 4.1, 4.5**), we propose

based on DFT calculations of the metal coordination sites that the stability of the two octahedral Ni^{II}:His_6 coordination motifs in the "up-up-up" trimer compared to that provided by hydrophobic packing in the "up-up-down" trimer is sufficiently high such that Ni^{II}-directed assembly is biased toward the "up-up-up" trimer (**Figure 4.17, Table 4.6**). Notably, the "up-up-up" trimer appears to preorganize an ideal octahedral coordination environment more effectively than the "up-up-down" trimer, as enforcing a His_6 coordination environment in Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$ led to a more distorted octahedral coordination environment (**Figure 4.18**).

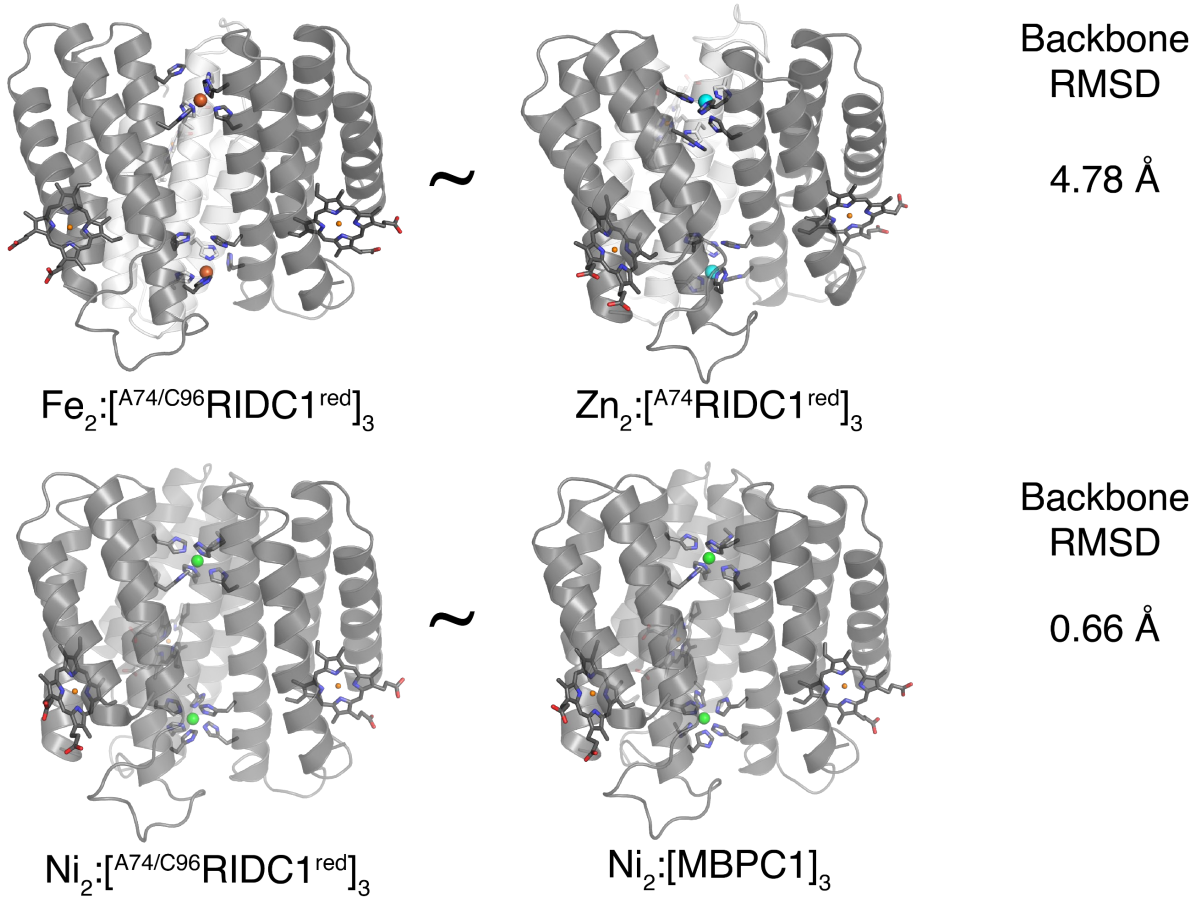

Figure 4.16 | Comparisons of metal-bound $^{A74/C96}$RIDC1^{red} structures with other cytochrome *cb*$_{562}$ variants. (a) Comparison of Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$ and Zn_2:[^{A74}RIDC1^{red}]$_3$, (b) Comparison of Ni_2:[$^{A74/C96}$RIDC1^{red}]$_3$ and Ni_2:[MBPC1]$_3$. Sequence-independent alignments were carried out in PYMOL.

Table 4.5 | Rosetta Interface Energy Calculations[a]

Structure	Trimer Energy (REU)	Monomer Energy (REU)	Trimer-Monomer Energy (REU)
Ni_2:[$^{A74/C96}$RIDC1]$_3$	-515.20	-539.44	24.25
Fe_2:[$^{A74/C96}$RIDC1]$_3$	221.29	246.37	-25.08

[a] Excludes energetic contribution of coordinating histidine residues and metal ions

In contrast to the Fe^{II} and Ni^{II}-directed assemblies, the crystal structure of the Zn^{II}-directed assembly revealed a tetrameric, D_4-symmetric architecture (Zn_4:[$^{A74/C96}$RIDC1^{red}]$_4$) in which the free Cys96 residues coordinate the metal ion (**Figure 4.13, 4.15d**). The assembly features four identical, tetrahedral $His_2GluCys$ coordination sites (**Figure 4.13a, 4.15d**). Each antiparallel dimer

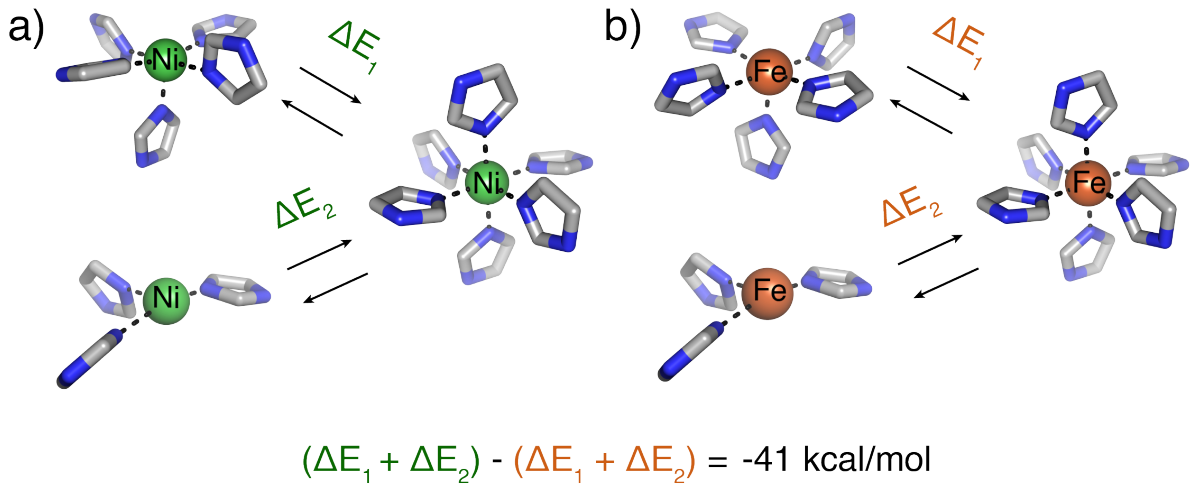

Figure 4.17 | DFT calculations on the relative stability of His_3, His_5, and His_6 coordination environments at (a) Ni^{II} and (b) Fe^{II} centers. Prior to the calculations, Ni^{II} was substituted into the Fe:His_3 and Fe:His_5 coordination sites extracted from the Fe_2: [$^{A74/C96}$RIDC1^{red}]$_3$ crystal structure, while Fe^{II} was substituted into one of the Ni^{II}:His_6 coordination sites extracted from the Ni_2: [$^{A74/C96}$RIDC1^{red}]$_3$ crystal structure. Calculations were performed without geometry optimization. The calculations predict that Ni^{II} is more stabilized by His_6 coordination than Fe^{II}, by 41 kcal/mol. We surmise that the relative stabilization of Ni^{II} in a His_6 coordination environment is enough to bias trimerization in the "up-up-up" arrangement.

Table 4.6 | DFT-computed energies of metal coordination sites[a]

Protein	Coordination	Multiplicity	Overall Charge	Energy (B3LYP)[b]
$^{A74/C96}$RIDC1^{red}	Fe^{II}:His_3	5	+2	-2093.28
$^{A74/C96}$RIDC1^{red}	Fe^{II}:His_5	5	+2	-2394.33
$^{A74/C96}$RIDC1^{red}	Fe^{II}:His_6	5	+2	-2620.57
$^{A74/C96}$RIDC1^{red}	Ni^{II}:His_3	3	+2	-2337.81
$^{A74/C96}$RIDC1^{red}	Ni^{II}:His_5	3	+2	-2638.86
$^{A74/C96}$RIDC1^{red}	Ni^{II}:His_6	3	+2	-2865.13
$^{A74/C96}$RIDC1^{red}	Ni^{II}:His_6 (alt. conf.)	3	+2	-2864.94

[a] Coordination sites extracted from crystal structures and without geometry optimization
[b] Energies reported in hartrees. 1 hartree = 627.50 kcal/mol

is oriented about 75° with respect to the other (**Figure 4.19**). This canted arrangement contrasts with the nearly collinear arrangement (θ = 21°) of antiparallel dimers in Zn_4: [$^{A74/C96}$RIDC1^{ox}]$_4$ (**Figure 4.19**). Taken together, the Zn_4:[$^{A74/C96}$RIDC1^{red}]$_4$ and Zn-bound [$^{A74/C96}$RIDC1^{ox}]$_4$ structures illustrate the dual functional role of cysteine as a metal coordinating ligand and covalent handle, which is critical to redox signaling in biological systems.[11, 42, 43] Our results demonstrate that $^{A74/C96}$RIDC1^{red} can adopt three unique and stable architectures in the presence of different metal ions (**Figure 4.3, 4.11-4.14**), exemplifying how metal coordination preferences and hydrophobic packing can collectively influence assembly outcomes.

4.4 Conclusions

The ability of a single protein construct to switch between multiple structural states upon exposure to different stimuli is indispensable to biological functions spanning signal transduction

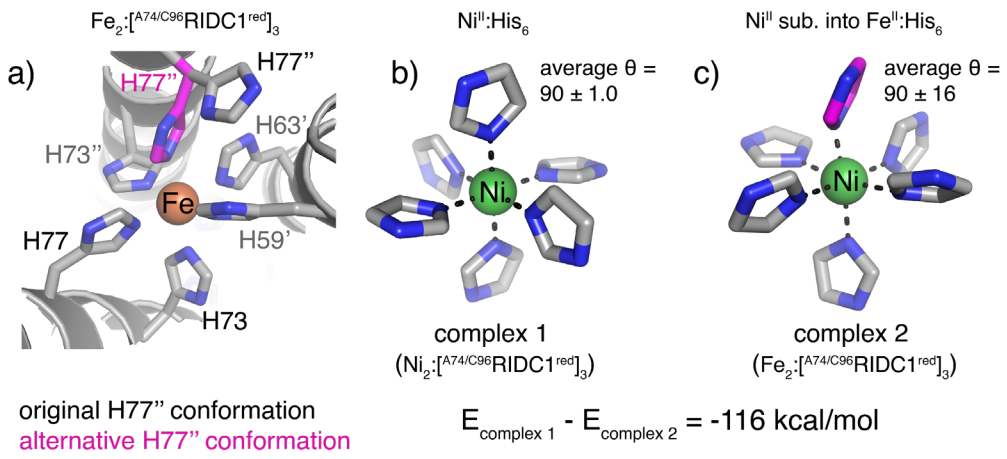

Figure 4.18 | Modeling and DFT calculations of a hypothetical His_6 coordination site in Fe_2:[$^{A74/C96}$RIDC1^{red}]$_4$. (a) Overlay of coordination sites in the Fe^{II}-directed trimer with and without the alternative H77’’ conformation that would give rise to a Fe^{II}:His_6 site. The alternative conformation of H77’’ (magenta) represents the highest probability rotamer within coordinating distance of Fe^{II} as predicted by the Dunbrack rotamer library. (b) Model of Ni^{II}:His_6 in the Ni^{II}-directed trimer. (c) Model of the hypothetical Fe^{II}:His_6 site in Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$ substituted with Ni^{II}. Significant deviations from ideal bond angles of an octahedral coordination geometry were observed in the hypothetical site, which led to a lower DFT-computed energy. This result suggests that the “up-up-down” trimer topology is less effective than the “up-up-up” topology at preorganizing octahedral His_6 coordination t

and gene regulation.[1, 2, 7, 8, 10] Herein, we have demonstrated that a single designed construct ($^{A74/C96}$RIDC1) can be subjected to redox- and metal-based stimuli to obtain five structurally distinct assemblies (**Figure 4.3**). The large structural diversity of $^{A74/C96}$RIDC1 assemblies can be attributed to an intricate interplay between metal-ligand, disulfide bonding, and hydrophobic interactions. While potentially serving as starting points for engineering downstream functions, the dynamic $^{A74/C96}$RIDC1 assemblies also pave the path to the generation of multistate protein switches.

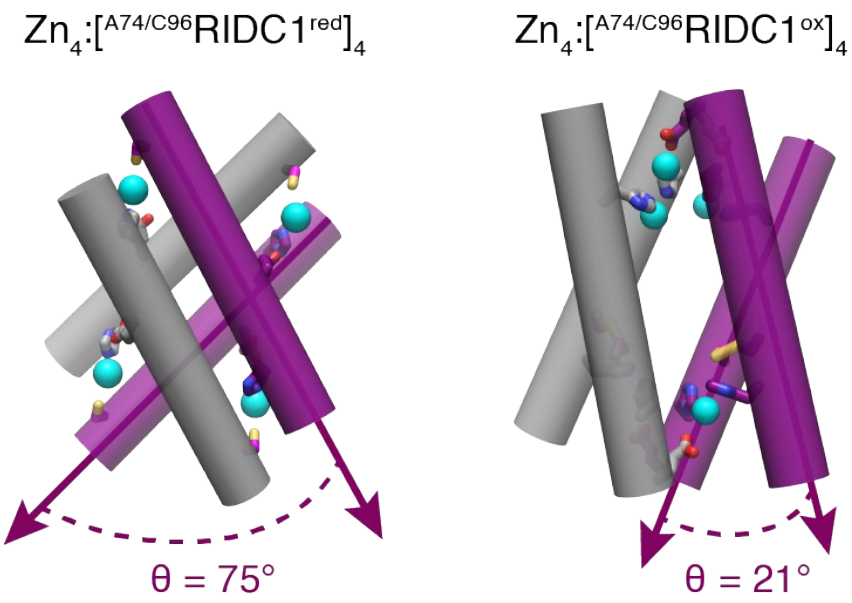

Figure 4.19 | Topviews of Zn$_4$:[$^{A74/C96}$RIDC1^{red}]$_4$ (left) and Zn$_4$:[$^{A74/C96}$RIDC1^{ox}]$_4$ (right) illustrating significant differences in the tetramer topologies. Each cartoon cylinder represents a protein monomer. Interhelical angles were calculated in PYMOL.

4.5 Materials and Methods

4.5.1 Mutagenesis, Expression, and Purification

$^{A74/C96}$RIDC1 is a variant of Rosetta interface design cytochrome-1 (RIDC1), which is itself a variant of cytochrome *cb*$_{562}$. [34] PCR-based site-directed mutagenesis of plasmids encoding RIDC1 (pET20b-[RIDC1]) was performed as previously described.[44] Purified plasmids were transformed into competent BL21 (DE3) *E.coli* cells containing the *ccm* (cytochrome c maturation) cassette plasmid, pEC86.[45] Colonies were allowed to grow for 20 hours on LB/agar plates containing ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL). Starter cultures were grown overnight for 16 hours at 37°C in LB media supplemented with the same antibiotic concentrations, diluted 100-fold into fresh, antibiotic supplemented LB media, and then grown at 37°C until the OD_{600} reached 0.6-1. Cultures were inoculated into 2.8 L glass flasks containing 1 L of LB media supplemented with antibiotics and shaken at 100 RPM for 20-24 hours at 37°C. Cells were pelleted

via centrifugation (5,000 x g, 4°C, 5 min) and the media discarded. The red cell pellets were resuspended in a 10 mM sodium acetate buffer solution (pH 5.0) and vigorously stirred until all pellets were resuspended. The resulting mixture was sonicated for 15 min in pulses of 30 s on and 60 s off (Qsonica). The lysate was titrated with sodium hydroxide to a pH of 10, and then acetic acid to a pH of 5.0, and then clarified by centrifugation (10,000 x g, 4°C, 20 min). The cleared lysate was applied to a CM Sepharose Fast Flow (Biorad) resin preequilibrated with a 10 mM sodium acetate buffer solution (pH 5.0) and eluted using a step-gradient of 0-500 mM NaCl. The visibly red eluate was pooled, concentrated, and exchanged into 10 mM NaP_i buffer solution (pH 8.0). The protein was then loaded onto a 5 mL High-Q cartridge column preequilibrated with the same buffer solution and eluted using a step-gradient of 0-1 M NaCl. Fractions with Reinheitszahl ratios (A_{415}/A_{280}) above 3 were pooled, concentrated, exchanged into a 20 mM MOPS buffer solution supplemented with 150 mM NaCl, and loaded onto a Superdex S75 size column. Fractions with Reinheitszahl ratios (A_{415}/A_{280}) above 5.5 were pooled, concentrated, and treated with 5 mM EDTA/DPA. Protein identity and purity were determined by ESI-MS and SDS-PAGE gels, respectively (**Figure 4.20-4.21**).

4.5.2 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

Oxidized protein samples (200 μM monomer) were mixed directly with metal salts. To obtain reduced protein samples (200 μM monomer), the samples were incubated with 5 equivalents Tris(3-hydroxypropyl) phosphine (THPP) for 15 minutes prior to metal addition. Co^{II}, Ni^{II}, Cu^{II}, and Zn^{II} additions were made at 1 equivalent metal/monomer under aerobic conditions, while Fe^{II} additions were made at 1 or 5 equivalents/monomer in an anaerobic chamber. Due to the possibility of irreversible electron transfer from non-heme Fe^{II} to the ferric-heme cofactors of $^{A74/C96}$RIDC1 under oxidized conditions, SV measurements of $^{A74/C96}$RIDC1^{ox} mixed with Fe^{II} are not reported.

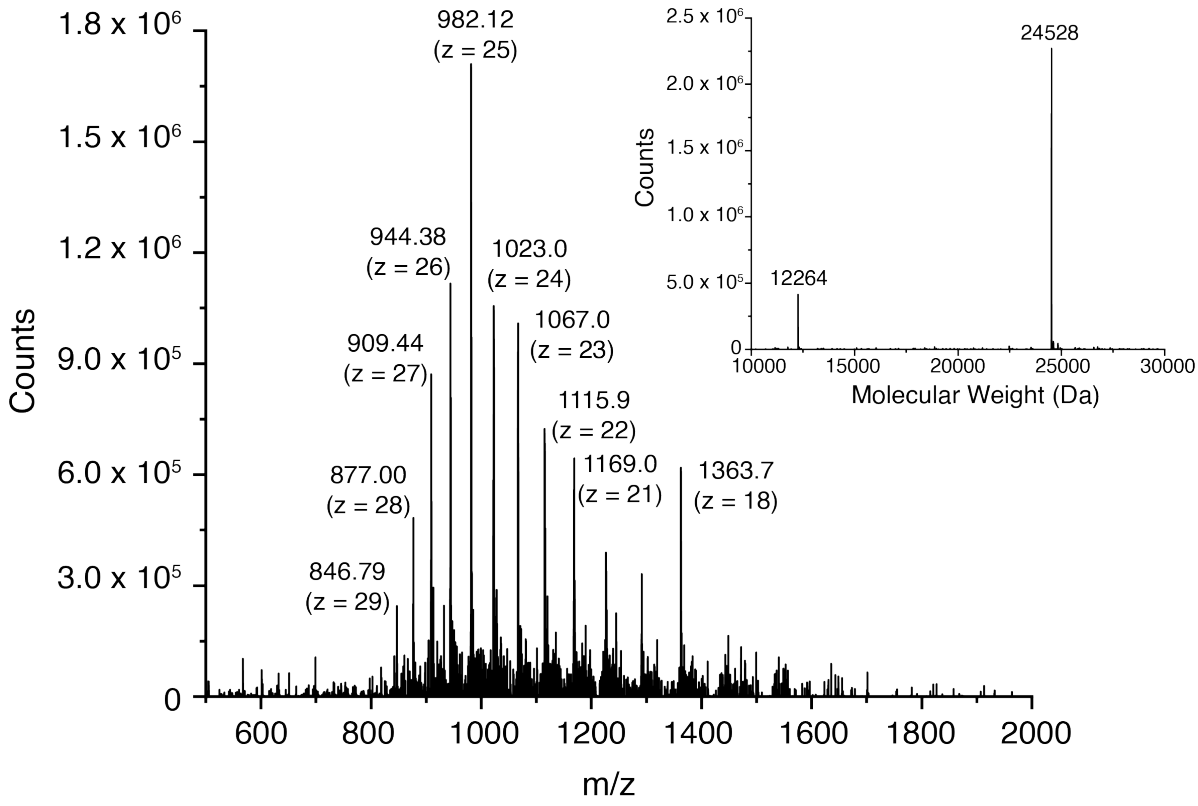

Figure 4.20 | Raw and deconvoluted (inset) ESI-MS spectra of $^{A74/C96}$RIDC1^{ox}. Charge states and m/z values in the raw spectrum correspond to the dimer.

Sedimentation velocity (SV) measurements were made in a solution of 20 mM TRIS (pH 7.5) at 25°C on a Beckman XL-A instrument equipped with a AN-60 Ti rotor and at 41,000 RPM. Samples were monitored at 570 nm (corresponding to a Q band of cytochrome) up to 12 h. Scans were processed and molecular weight distributions calculated using SEDFIT software.[46] Fitting parameters such as the buffer density (0.9988 g/mL), buffer viscosity (0.01007 poise), and partial specific volume (0.7313 mL/g) were calculated by SEDNTERP. SV profiles are shown at a confidence level of 95%.[46, 47] Oligomerization yields were estimated based on Riemann integrations of the peaks of the SV profiles, where the bounds of each discrete peak were defined by the full width at half maximum (FWHM).

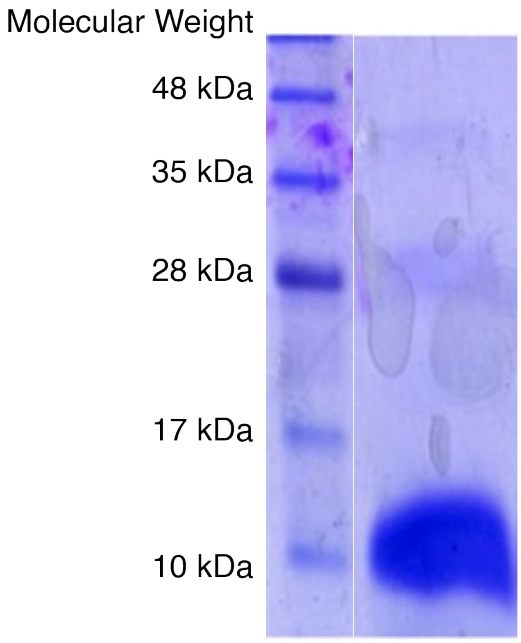

Figure 4.21 | Denaturing SDS-PAGE gel of $^{A74/C96}$RIDC1. The gel was run at 200 V for 45 minutes and under reducing conditions.

4.5.3 X-ray structure determination

Crystals of $^{A74/C96}$RIDC1 variants were obtained by sitting-drop vapor diffusion at 25°C. To obtain crystals of oxidized samples, 2-2.5 mM protein monomer was mixed with 1-5 equivalents of metal salts for at least one hour at 25°C. To obtain crystals of reduced samples, 2-2.5 mM protein monomer was mixed with 5-10 equivalents of either THPP or tris(2-carboxyethyl)phosphine (TCEP) prior to metal addition in an anaerobic chamber. Metal-loaded samples were mixed at a ratio of 1 µL: 1 µL or 2 µL: 1 µL with mother liquor. All crystals were transferred into perfluoro polyether (Hampton) for cryoprotection prior to freezing. Diffraction data were collected at 80-100 K on ALS Beamline 5.0.1 (Apo [$^{A74/C96}$RIDC1^{ox}]$_4$), SSRL Beamline 9-2 (Co_2:[$^{A74/C96}$RIDC1^{ox}]$_4$, Zn_2:[$^{A74/C96}$RIDC1^{ox}]$_4$, Fe_2: [$^{A74/C96}$RIDC1^{red}]$_3$, Ni_2: [$^{A74/C96}$RIDC1^{red}]$_3$, Zn_4: [$^{A74/C96}$RIDC1^{red}]$_4$), and ALS Beamline 8.3.1 (Cu_4: [$^{A74/C96}$RIDC1^{red}]$_4$). Diffraction data were processed using either iMOSFLM or XDS and scaled using SCALA.[48] Molecular replacement was carried out using Phaser with monomeric cytochrome *cb*$_{562}$ (PDB:

2BC5) as the search model.[49] Refinement was performed using phenix.refine while model building and placement of metal ions/water was performed using COOT.[50] Electron density maps were generated using Phenix and then converted into CCP4 map files using a Fast Fourier Transform algorithm (FFT, CCP4i).[48] All final models and CCP4 electron density maps were rendered in PYMOL (www.pymol.org). Surface calculations were performed using PISA.[51]

4.5.4 Rosetta interface energy calculations

To estimate the stability of Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$ and Ni_2:[$^{A74/C96}$RIDC1^{red}]$_3$ based on non-covalent interactions while excluding the contribution of metal-ligand interactions, Rosetta score calculations were performed on Fe_2:$^{A74/C96}$RIDC1$^{red}{}_3$ and Ni_2:$^{A74/C96}$RIDC1$^{red}{}_3$ crystal structures. In each case, the heme cofactors and all metal ions and water molecules were removed prior to loading the trimeric complex into Rosetta. After evaluating the Rosetta score of the trimer, the chains were moved 100 Å apart and the Rosetta score of the system was re-evaluated. Following these calculations, the sum of REU values for metal-binding histidine residues (H59, H63, H73, H77) was subtracted from each Rosetta score. Finally, the adjusted Rosetta score of the trimer was subtracted from that of the separated chains to obtain a ΔREU value that serves as a proxy for the ΔΔG of trimerization.[52]

4.5.5 DFT calculations

All metal complexes were extracted from crystal structures of RIDC1 variants. Input files were prepared using Avogadro software.[53] Single point calculations were performed at the level of B3LYP theory using the 6-31G basis set.[54] Calculations were performed without an initial geometry optimization step to avoid introducing model bias. To generate the hypothetical Ni^{II}:His_6 coordination site in Fe_2: [$^{A74/C96}$RIDC1^{red}]$_3$, the position of the His77 residue of the Fe^{II}:His_5 site was modified to be within coordinating distance of the metal center. This new rotamer was

predicted by the Dunbrack rotamer library to be the most stable metal-binding rotamer accessible.[55]

4.5.6 Native electrospray ionization (ESI) mass spectrometry

ESI-MS experiments were performed using an Elite hybrid linear ion trap-orbitrap mass spectrometer equipped with a heated-electrospray ionization (HESI) source (Thermo). All solutions for ESI-MS were prepared using LC-MS grade water (JT Baker). $^{A74/C96}$RIDC1 samples were prepared at 50 μM protein monomer in 20 mM NH_4HCO_3. High mass mode at 240,000 resolution mode was applied to obtain the best resolution limit. To prevent cross-contamination, peek tubing and steel capillary were sequentially washed using 0.1% formic acid and 20 mM NH_4HCO_3 between measurements. The estimated resolution (m/Δm) of the mass spectrometer was ~80,000 around the +11 charge state. Distributions were analyzed using Qual Browser (Thermo).

4.6 Acknowledgements

This work was funded by NIH (R01-GM138884 and T32-GM112584), and by NASA (80NSSC18M0093; ENIGMA: Evolution of Nanomachines in Geospheres and Microbial Ancestors (NASA Astrobiology Institute Cycle 8). E.G. acknowledges funding by EMBO (ALTF 1336-2015). Portions of this research were carried out at the Stanford Linear Accelerator Center (supported by the DOE, Office of Basic Energy Sciences contract DE-AC02-76SF00515 and NIH P30-GM133894) and the Advanced Light Source at the Lawrence Berkeley National Laboratory (supported by the DOE, Office of Basic Energy Sciences contract DE-AC02-05CH11231and NIH P30-GM124169-01). Coordinate and structure factor files for the crystal structures have been deposited into the Protein Data Bank (www.rcsb.org) with the following accession codes: 7RWV (Apo [$^{A74/C96}$RIDC1^{ox}]$_4$), 7SU2 (Co_2:[$^{A74/C96}$RIDC1^{ox}]$_4$), 7RWW (Zn_4:[$^{A74/C96}$RIDC1^{ox}]$_4$), 7RWY

(Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$), 7RWU (Fe_2:[$^{A74/C96}$RIDC1^{red}]$_3$), 7TEP (Cu_4:[$^{A74/C96}$RIDC1^{red}]$_4$), and 7RWX (Cu_4:[$^{A74/C96}$RIDC1^{red}]$_4$).

Chapter 4 is reproduced, in part, with permission, from: Kakkis, A.; Golub, E.; Choi, T. S.; Tezcan, F. A., "Redox- and metal-directed structural diversification in designed metalloprotein assemblies", *Chem. Commun* **58 (49)**, 6958-6961 (2022). The dissertation author was the primary author on all reprinted materials.

4.7 References

1. James, L. C.; Tawfik, D. S., Conformational diversity and protein evolution – a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **2003,** *28* (7), 361-368.
2. Tokuriki, N.; Tawfik, D. S., Protein dynamism and evolvability. *Science* **2009,** *324* (5924), 203-7.
3. Zhu, J.; Avakyan, N.; Kakkis, A.; Hoffnagle, A. M.; Han, K.; Li, Y.; Zhang, Z.; Choi, T. S.; Na, Y.; Yu, C. J.; Tezcan, F. A., Protein Assembly by Design. *Chem. Rev.* **2021,** *121* (22), 13701-13796.
4. Sanders, B. D.; Jackson, B.; Marmorstein, R., Structural basis for sirtuin function: what we know and what we don't. *Biochim. Biophys. Acta* **2010,** *1804* (8), 1604-16.
5. Chasteen, N. D.; Harrison, P. M., Mineralization in ferritin: an efficient means of iron storage. *J. Struct. Biol.* **1999,** *126* (3), 182-94.
6. Armstrong, R. N., Mechanistic diversity in a metalloenzyme superfamily. *Biochemistry* **2000,** *39* (45), 13625-32.
7. Rosenbaum, D. M.; Rasmussen, S. G.; Kobilka, B. K., The structure and function of G-protein-coupled receptors. *Nature* **2009,** *459* (7245), 356-63.
8. Groitl, B.; Jakob, U., Thiol-based redox switches. *Biochim. Biophys. Acta* **2014,** *1844* (8), 1335-43.
9. Chang, Y. G.; Cohen, S. E.; Phong, C.; Myers, W. K.; Kim, Y. I.; Tseng, R.; Lin, J.; Zhang, L.; Boyd, J. S.; Lee, Y.; Kang, S.; Lee, D.; Li, S.; Britt, R. D.; Rust, M. J.; Golden, S. S.; LiWang, A., A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **2015,** *349* (6245), 324-328.
10. Takken, F. L.; Albrecht, M.; Tameling, W. I., Resistance proteins: molecular switches of plant defence. *Curr. Opin. Plant Biol.* **2006,** *9* (4), 383-90.
11. Fillat, M. F., The FUR (ferric uptake regulator) superfamily: diversity and versatility of key transcriptional regulators. *Arch. Biochem. Biophys.* **2014,** *546*, 41-52.
12. Gay, N. J.; Gangloff, M.; Weber, A. N., Toll-like receptors as molecular switches. *Nat. Rev. Immunol.* **2006,** *6* (9), 693-8.
13. Ghanbarpour, A.; Pinger, C.; Esmatpour Salmani, R.; Assar, Z.; Santos, E. M.; Nosrati, M.; Pawlowski, K.; Spence, D.; Vasileiou, C.; Jin, X.; Borhan, B.; Geiger, J. H., Engineering the hCRBPII Domain-Swapped Dimer into a New Class of Protein Switches. *J. Am. Chem. Soc.* **2019,** *141* (43), 17125-17132.

14. Karchin, J. M.; Ha, J. H.; Namitz, K. E.; Cosgrove, M. S.; Loh, S. N., Small Molecule-Induced Domain Swapping as a Mechanism for Controlling Protein Function and Assembly. *Sci. Rep.* **2017,** *7*, 44388.
15. Salgado, E. N.; Lewis, R. A.; Mossin, S.; Rheingold, A. L.; Tezcan, F. A., Control of Protein Oligomerization Symmetry by Metal Coordination: C2 and C3 Symmetrical Assemblies through CuII and NiII Coordination. *Inorg. Chem.* **2009,** *48* (7), 2726-2728.
16. Bai, Y.; Luo, Q.; Zhang, W.; Miao, L.; Xu, J.; Li, H.; Liu, J., Highly ordered protein nanorings designed by accurate control of glutathione S-transferase self-assembly. *J. Am. Chem. Soc.* **2013,** *135* (30), 10966-9.
17. Signarvic, R. S.; DeGrado, W. F., De novo design of a molecular switch: phosphorylation-dependent association of designed peptides. *J. Mol. Biol.* **2003,** *334* (1), 1-12.
18. Harrington, L.; Fletcher, J. M.; Heermann, T.; Woolfson, D. N.; Schwille, P., De novo design of a reversible phosphorylation-dependent switch for membrane targeting. *Nat. Commun.* **2021,** *12* (1), 1472.
19. Ballister, E. R.; Lai, A. H.; Zuckermann, R. N.; Cheng, Y.; Mougous, J. D., In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (10), 3733-8.
20. Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A., Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **2016,** *533* (7603), 369-73.
21. Wei, K. Y.; Moschidi, D.; Bick, M. J.; Nerli, S.; McShan, A. C.; Carter, L. P.; Huang, P. S.; Fletcher, D. A.; Sgourakis, N. G.; Boyken, S. E.; Baker, D., Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117* (13), 7208-7215.
22. Cerasoli, E.; Sharpe, B. K.; Woolfson, D. N., ZiCo: a peptide designed to switch folded state upon binding zinc. *J. Am. Chem. Soc.* **2005,** *127* (43), 15008-9.
23. Langan, R. A.; Boyken, S. E.; Ng, A. H.; Samson, J. A.; Dods, G.; Westbrook, A. M.; Nguyen, T. H.; Lajoie, M. J.; Chen, Z.; Berger, S.; Mulligan, V. K.; Dueber, J. E.; Novak, W. R. P.; El-Samad, H.; Baker, D., De novo design of bioactive protein switches. *Nature* **2019,** *572* (7768), 205-210.
24. Alberstein, R. G.; Guo, A. B.; Kortemme, T., Design principles of protein switches. *Curr. Opin. Struct. Biol.* **2021,** *72*, 71-78.
25. Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y.; Baker, D., Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (19), 10687-10691.
26. Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D., Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003,** *331* (1), 281-299.
27. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y. E.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popovic, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P., ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth. Enzymol.* **2011,** *487*, 545-574.

28. Brodin, J. D.; Ambroggio, X. I.; Tang, C.; Parent, K. N.; Baker, T. S.; Tezcan, F. A., Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **2012,** *4* (5), 375-82.
29. Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A., Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **2015,** *137* (33), 10468-71.
30. Yang, M.; Song, W. J., Diverse protein assembly driven by metal and chelating amino acids with selectivity and tunability. *Nat. Commun.* **2019,** *10* (1), 5545.
31. Rao, S. R.; Schettler, S. L.; Horne, W. S., Metal-Binding Foldamers. *Chempluschem* **2021,** *86* (1), 137-145.
32. Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A., Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* **2007,** *129* (44), 13374-13375.
33. Radford, R. J.; Nguyen, P. C.; Ditri, T. B.; Figueroa, J. S.; Tezcan, F. A., Controlled protein dimerization through hybrid coordination motifs. *Inorg. Chem.* **2010,** *49* (9), 4362-4369.
34. Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A., Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (5), 1827-1832.
35. Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A., In vitro and cellular self-assembly of a Zn-binding protein cryptand via templated disulfide bonds. *J. Am. Chem. Soc.* **2013,** *135* (32), 12013-12022.
36. Rittle, J.; Field, M. J.; Green, M. T.; Tezcan, F. A., An efficient, step-economical strategy for the design of functional metalloproteins. *Nat. Chem.* **2019,** *11* (5), 434-441.
37. Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F. A., Metal-templated design of chemically switchable protein assemblies with high-affinity coordination sites. *Angew. Chem. Int. Ed. Engl.* **2020,** *59* (49), 21940-21944.
38. Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A., De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J. Am. Chem. Soc.* **2016,** *138* (40), 13163-13166.
39. Song, W. J.; Tezcan, F. A., A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014,** *346* (6216), 1525-1528.
40. Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A., Evolution of metal selectivity in templated protein interfaces. *J. Am. Chem. Soc.* **2010,** *132* (25), 8610-8617.
41. Ni, T. W.; Tezcan, F. A., Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (39), 7014-7018.
42. Maret, W., Molecular aspects of human cellular zinc homeostasis: redox control of zinc potentials and zinc signals. *Biometals* **2009,** *22* (1), 149-57.
43. D'Autreaux, B.; Pecqueur, L.; Gonzalez de Peredo, A.; Diederix, R. E.; Caux-Thang, C.; Tabet, L.; Bersch, B.; Forest, E.; Michaud-Soret, I., Reversible redox- and zinc-dependent dimerization of the Escherichia coli fur protein. *Biochemistry* **2007,** *46* (5), 1329-42.
44. Liu, H.; Naismith, J. H., An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **2008,** *8*, 91.
45. Faraone-Mennella, J.; Tezcan, F. A.; Gray, H. B.; Winkler, J. R., Stability and folding kinetics of structurally characterized cytochrome c-b562. *Biochemistry* **2006,** *45* (35), 10504-11.

46. Schuck, P., Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **2000,** *78* (3), 1606-1619.
47. Lebowitz, J.; Lewis, M. S.; Schuck, P., Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* **2002,** *11* (9), 2067-79.
48. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994,** *50* (Pt 5), 760-3.
49. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-21.
50. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004,** *60* (Pt 12 Pt 1), 2126-32.
51. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372* (3), 774-97.
52. Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J., The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **2017,** *13* (6), 3031-3048.
53. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012,** *4* (1), 17.
54. Hehre, W. J., Ab initio molecular orbital theory. *Acc. Chem. Res* **1976,** (9), 399-406.
55. Shapovalov, M. V.; Dunbrack, R. L., Jr., A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **2011,** *19* (6), 844-58.

Chapter 5: Conclusions

5.1 Introduction

Metalloproteins mediate a wide range of biological processes critical to Life.[1-5] Given their biological importance, metalloproteins have long inspired efforts to emulate and even expand on their natural functions through *de novo* design or repurposing.[6-8] While the vast majority of designed metalloproteins are monomeric, many complex functions performed by natural metalloproteins are achieved through the assembly of discrete protein domains into oligomeric structures hosting interfacial metal coordination sites.[4, 5] Given the critical role of protein assembly in rendering functional metalloprotein architectures, our work is oriented toward the design of metalloprotein assemblies using structurally simple, monomeric building blocks. We have demonstrated that via the application of MDPSA and MeTIR methodologies to a four-helix bundle (cytochrome cb_{562}), we can design metalloprotein assemblies with a wide variety of structures, metal coordination environments, and potential functions.

5.2 Metal-templated design of protein trimers with high affinity, His_6 coordination sites

Previously, all metalloprotein assemblies designed using the MDPSA methodology formed assemblies whose oligomeric states were dependent on the coordination preferences of the exogenous metal ions. TriCyt1, an engineered variant of cytochrome cb_{562}, crystallized into a trimer in the presence of Ni^{II} or Cu^{II}.[9] Crystal structures of the metal-protein complexes revealed that both metal ions, despite possessing distinct coordination preferences, were bound in mononuclear, octahedral, His_6 coordination environments. In correlation with the crystallography data, TriCyt1 formed trimeric assemblies in solution in the presence of all mid-to-late first row transition metal ions (Mn^{II}-Zn^{II}), with trimer yields ranging from 12% (+Mn^{II}) to 89% (+Co^{II}). Calprotectin, an immune protein involved in metal ion sequestration,[1] is the only known

metalloprotein to coordinate Mn^{II} at a His_6 coordination site with nanomolar affinity. Inspired by this natural example and the rarity of high affinity Mn^{II} coordination sites in designed proteins,[10] we implemented MeTIR to redesign TriCyt1 into more stable trimers which could bind Mn^{II}-Zn^{II} with at least nanomolar affinity (**Figure 5.1**).

The first round of MeTIR converted TriCyt1 into TriCyt2 through the computationally prescribed installation of hydrophobic packing and salt-bridge interactions at the C_3 interface. TriCyt2 trimerized with near-quantitative yield in the presence of Mn^{II} but remained monomeric in the absence of metal ions. A second round of MeTIR involved the rationally guided installation

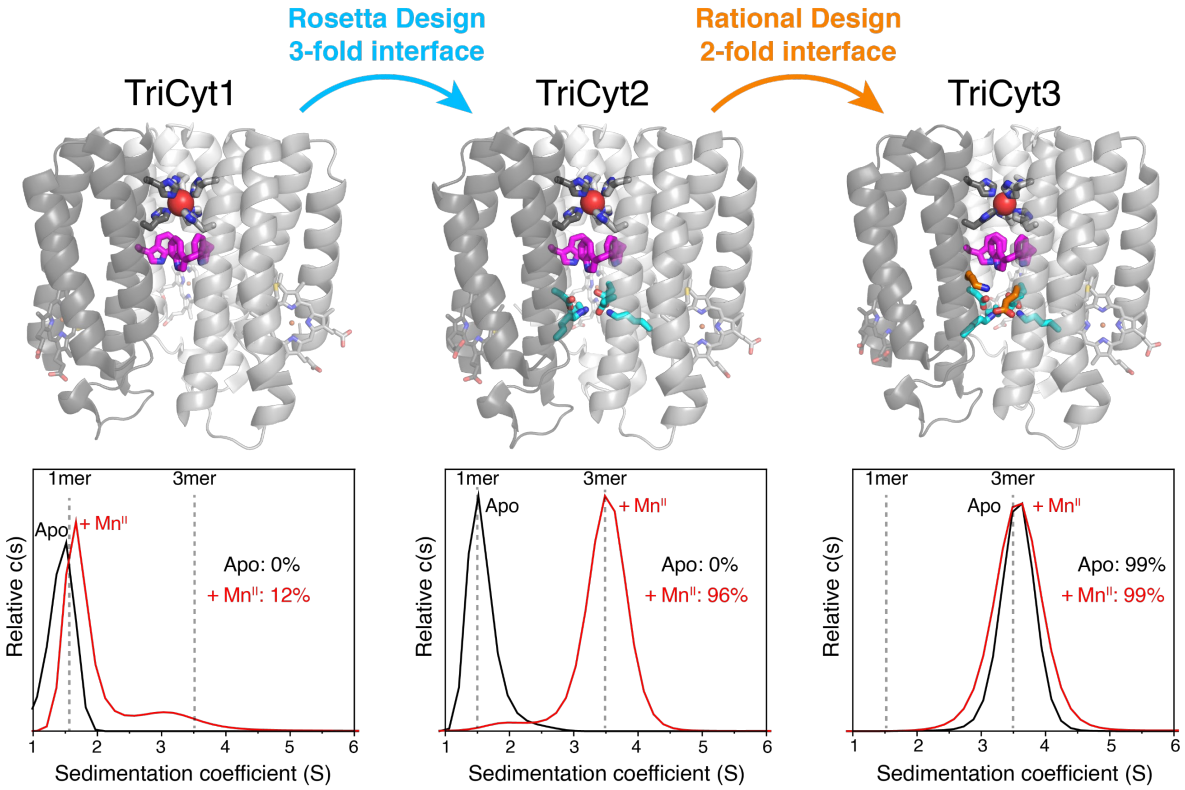

Figure 5.1 | Workflow to convert TriCyt1 (left) into TriCyt3 (right) via sequential interface redesign. The His_6 site is shown in gray, Trp70 residues are shown in magenta, select mutations resulting from Rosetta Design are shown in cyan, and select mutations resulting from the rational design step are shown in orange. Metal ions are shown as red spheres. For each SV-AUC graph, percentages indicate trimerization yields as estimated from numerical integration of the distributions. One advantage of the sequential design process was the ability to isolate both Mn^{II}-switchable (TriCyt2) and Mn^{II}-independent (TriCyt3) trimeric constructs, which could each provide unique downstream functions.

of salt-bridge interactions at the two-fold interfaces of TriCyt2, resulting in TriCyt3. TriCyt3 formed a metal-independent trimer whose oligomerization could be reversibly triggered by changes in solution pH. Competitive binding titrations, X-ray crystallography, and X-band EPR spectroscopy collectively showed that TriCyt3 coordinates Mn^{II} at a highly symmetrical, His_6 coordination site with near-nanomolar affinity ($K_d \sim 50$ nM).[9] Mutagenesis studies and X-ray crystallography experiments demonstrated our ability to modulate the secondary coordination sphere around the His_6 site via the engineering of new hydrogen bonding interactions with primary sphere residues.

TriCyt represents a promising starting point for engineering metal-based functions, as it binds a wide range of metal ions at a single, well-defined coordination site while being amenable to secondary sphere modifications. Current studies are oriented toward the diversification of the primary and secondary spheres of TriCyt in pursuit of functions encompassing metal sequestration, electron transfer, and catalysis.

5.3 Primary and secondary sphere modifications of TriCyt in pursuit of metal-based functions

The potential functions of natural and designed metalloproteins are inextricably tied to their coordination site/primary sphere characteristics. For example, the mononuclear, His_6 primary sphere of TriCyt3 could be utilized for high affinity metal binding and sequestration due to coordinative saturation around the metal ion. Mononuclear, coordinatively saturated primary spheres represent just one of many metal coordination motifs that are utilized by metalloproteins for metal-based functions. Therefore, expanding the potential functionality of the TriCyt scaffold requires diversification of the primary coordination sphere. In pursuit of protein constructs which could bind lanthanide ions with high affinity and selectivity, mediate hydrolytic and redox-

dependent transformations, and stabilize multinuclear metal coordination sites, we carried out rationally guided primary sphere mutations of TriCyt2 and TriCyt3 (**Figure 5.3a**).

First, we aimed to establish whether high-throughput screening of TriCyt2 variants featuring amino acid residues with O-based ligands could lead to the discovery of selective lanthanide binding proteins. To this end, we carried out SSM of residue positions 73 and 77 of TriCyt2, using VAN codon randomization to generate a sequence library encompassing four amino acid residues which have previously been observed in coordination sites of lanthanide binding proteins: Gln, Glu, Asn, and Asp. Armed with a library of TriCyt2 variants, we combined a periplasmic protein extraction method with SEC and tryptophan fluorescence quenching experiments to rapidly monitor protein oligomerization in the presence of lanthanide ions, hypothesizing that protein oligomerization would be coupled to lanthanide coordination. Control experiments demonstrated significant flaws in this approach. An SEC chromatogram of TriCyt2 extracted from the periplasm indicated metal-*independent* assembly, while fluorescence spectroscopy measurements were not reproducible when carried out on protein extracts. Future protocols to efficiently screen for lanthanide binding proteins would involve direct measurements of metal-protein complex formation via mass spectrometry and expression of library constructs in the media, which should be a less heterogenous milieu than the periplasm.

Another motivation for primary sphere modification of TriCyt was to obtain metalloprotein constructs which mediate redox-dependent or hydrolytic transformations. As a first step toward the design of such metalloproteins, we replaced His77 residues of TriCyt3 with Gly/Ala to obtain a mononuclear, His_3 coordination site reminiscent of the primary sphere environments of cysteine dioxygenase and carbonic anhydrase (**Figure 5.3a**). We also mutated these constructs into their heme-less variants (ncTriCyt3),[11] enabling reaction monitoring via UV-vis spectroscopy and

abrogating potential background catalysis by the heme moiety. We then screened for emergent peroxidase and hydride transfer activity using UV-vis and NMR spectroscopies, respectively. In our initial screens, ^{A77}ncTriCyt3 did not exhibit any metal-dependent catalytic activity. By contrast, Cu^{II}-loaded ^{G77}ncTriCyt3 exhibited an enhancement in peroxidase activity relative to apo protein for three substrates: syringol, catechol and ODA. However, the protein-Cu^{II} complex did not have enhanced catalytic activity compared to free Cu^{II}, with the difference in rates being as high as 0.0003 AU x s^{-1} (substrate = ODA). With the aim of improving the Cu^{II}-dependent ODA peroxidase activity of ^{G77}ncTricCyt3, we implemented a high-throughput screening protocol. Prior to screening, we generated a sequence library via SSM of residue positions 80 and 81. We utilized NYT codon randomization to obtain 64 mutants which should have larger substrate binding pockets than ^{G77}ncTriCyt3. *In vivo* and *in vitro* screens revealed two mutants with oxidation rates on par or exceeding that of free Cu^{II}, GESL and GEAL. SV-AUC experiments revealed that GESL was nearly quantitatively assembled in the presence of Cu^{II} and at protein concentrations used in the catalytic assays, suggesting that the observed activity is largely attributable to protein-bound Cu^{II}. Meanwhile, the Cu^{II}-dependent trimer yield of GEAL was only 72%, implying that a smaller fraction of the catalytic activity is mediated by protein-bound Cu^{II}. Moving forward, GESL represents a promising starting point for various catalytic screens and directed evolution of emergent catalytic functions.

Our third path of primary sphere modification was to mutate His77 of ncTriCyt2 to Cys77 to obtain metalloprotein constructs which could stabilize [4Fe-4S] clusters (**Figure 5.3a**). We envisioned that ^{C77}ncTriCyt2 could simultaneously coordinate [4Fe-4S] at a $(Cys77)_3$ site and a non-cluster metal ion at a $(His73)_3$ site, rendering metalloprotein constructs which host two electronically coupled, compositionally distinct metal centers. Such metal centers are commonly

featured in functionally complex metalloenzymes[4, 12, 13] but are rarely obtained through design.[14] The successful construction of proteins hosting multinuclear, heterometallic coordination sites could thus increase the scope of natural metalloprotein functions which we can emulate or expand on via design.

The formation of protein-cluster adducts following the reaction of ^{C77}ncTriCyt2 with a synthetic cluster ($[Fe_4S_4Cl_4]^{2-}$, or $Fe_4S_4^{syn}$) was probed in solution using UV-vis and EPR spectroscopies. Taken together, the solution data suggested 60% incorporation of an unusually low potential, protein-Fe_3S_4 complex. Crystal structures did not corroborate the formation of such a complex, instead revealing a dinuclear Fe complex with Fe-Fe distances of 2.5-3.2 Å. Inspired by the prospect of stabilizing similar dinuclear sites in solution, we assessed the metal coordination behavior of a TriCyt3 variant termed ^{C77}CIT, which features Cys77 and His81 coordination layers. We expected that ^{C77}CIT would be a more suitable scaffold for a multinuclear site, as PYMOL modeling predicted that the S_γ atoms of the Cys77 layer would be in closer proximity than those of ^{C77}ncTriCyt2. Crystallization of ^{C77}CIT in the presence of excess Fe^{II} led to a serendipitous discovery: instead of a dinuclear site, we observed a mononuclear, trigonal planar Fe:Cys_3 coordination complex with distortions toward a trigonal pyramidal geometry (**Figure 5.3b**). This particular Fe coordination environment is not only unprecedented in biology[15], but resembles that of synthetic Fe coordination complexes which bind and activate a wide variety of small molecules (**Figure 5.3b**).[16-18] In future studies, we will focus not only on assessing reactivity at the Fe:Cys_3 site but also establishing more effective and predictable methods for stabilizing multinuclear sites/metal clusters in designed metalloproteins. One potential method involves using the target

metal cluster as a structural template for designing a protein assembly that more effectively pre-organizes the cluster coordination environment, analogous to MeTIR.

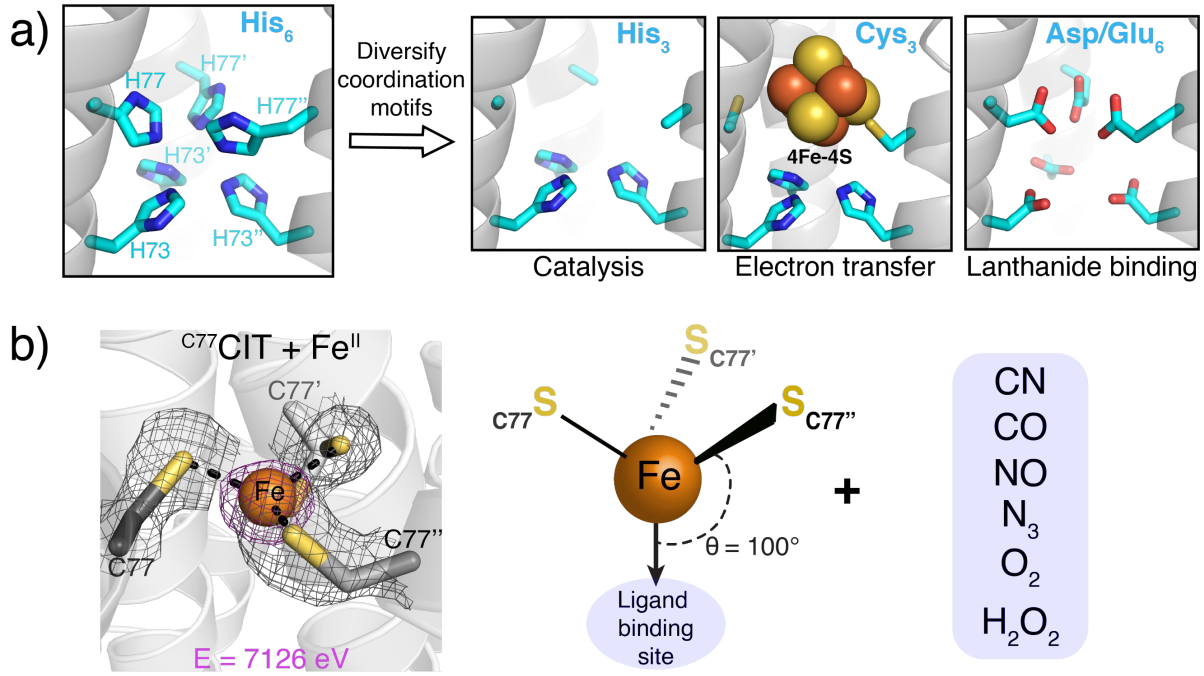

Figure 5.2 | Primary sphere modifications of TriCyt2 and TriCyt3 mutants in pursuit of diverse metal-based functions. (a) Scheme illustrating three distinct modifications of the TriCyt coordination site (originally His_6) in pursuit of metal-dependent catalysis, electron transfer via Fe-S clusters, and lanthanide binding. (b) Crystal structure of ^{C77}CIT-Fe^{II} (left) and schematic of the mononuclear, Fe:Cys_3 coordination site (right). The pyramidalization angle (θ-90° = 10°) is intermediate between that of ideal trigonal planar (θ-90° = 0°) and trigonal pyramidal (θ-90° = 19.5°) complexes. Given its low coordination number and non-zero pyramidalization angle, the Fe:Cys_3 complex may bind to or react with a wide array of small molecules. Reactivity will likely vary depending on the resting redox state of the Fe ion.

5.4 Design of a metalloprotein that forms multiple redox- and metal-directed assembly states

Most protein design strategies involve the implementation of extensive noncovalent interactions in a protein building block to mediate its assembly into a single, stable structure ("one sequence/one assembly"). However, the propensity to form *multiple* distinct conformations or assembly states from a single protein sequence plays a critical role in biological function.[19-22]

While there are some examples of designed proteins whose assembly can be triggered by a single type of stimulus,[23-26] the ability to design proteins that respond to more than one stimulus type to access more than two structurally distinct states has been limited. Previously, we characterized the assembly states of two RIDC1 variants, ^{C96}RIDC1 and ^{A74}RIDC1. We observed that ^{C96}RIDC1 formed a metal-independent tetramer upon the oxidation of Cys96 to form disulfide bonds, while ^{A74}RIDC1 assembled into a trimer upon the addition of Zn^{II}.

Motivated by the prospect of obtaining a protein construct which can access multiple oligomeric states in response to both redox- and metal-based stimuli, we generated a new protein construct incorporating both mutations ($^{A74/C96}$RIDC1).[27] We found that $^{A74/C96}$RIDC1 formed five structurally distinct assemblies upon the addition of redox- and metal-based stimuli (**Figure 5.3**). In its disulfide-oxidized form ($^{A74/C96}$RIDC1^{ox}), the construct assembled into two different but closely related tetrameric conformations, one corresponding to the metal-free state and the other to a Co^{II}/Zn^{II}-bound state.[27] Evidently, the covalent preorganization induced by Cys96/Cys96 disulfide bonds helped restrict the oligomeric/conformational diversity of $^{A74/C96}$RIDC1^{ox}. By contrast, the disulfide-reduced form of the protein ($^{A74/C96}$RIDC1^{red}) could assemble into three highly distinct architectures upon the addition of metal ions: an "up-up-down" trimer ($+Fe^{II}$), an "up-up-up" trimer ($+Ni^{II}$), and a canted tetramer ($+Zn^{II}$).[27] Rosetta interface and DFT calculations led us to surmise that the large structural diversity of $^{A74/C96}$RIDC1^{red} assemblies is due to a complex interplay between the formation of favorable hydrophobic packing interactions and metal-ligand interactions which satisfy the coordination preferences of the exogenous metal ion. $^{A74/C96}$RIDC1 assemblies serve as potential starting points for engineering downstream functions encompassing metal sequestration, electron transfer, and catalysis. Because of its multi-stimuli-responsiveness, $^{A74/C96}$RIDC1 also serves as a promising building block for obtaining protein

constructs whose redox- and metal-directed assembly states are interconvertible. Achieving a redox- and metal-gated switch would likely require computational redesign of $^{A74/C96}$RIDC1 such that the construct could access redox- and metal-directed architectures representing equally shallow energy minima.

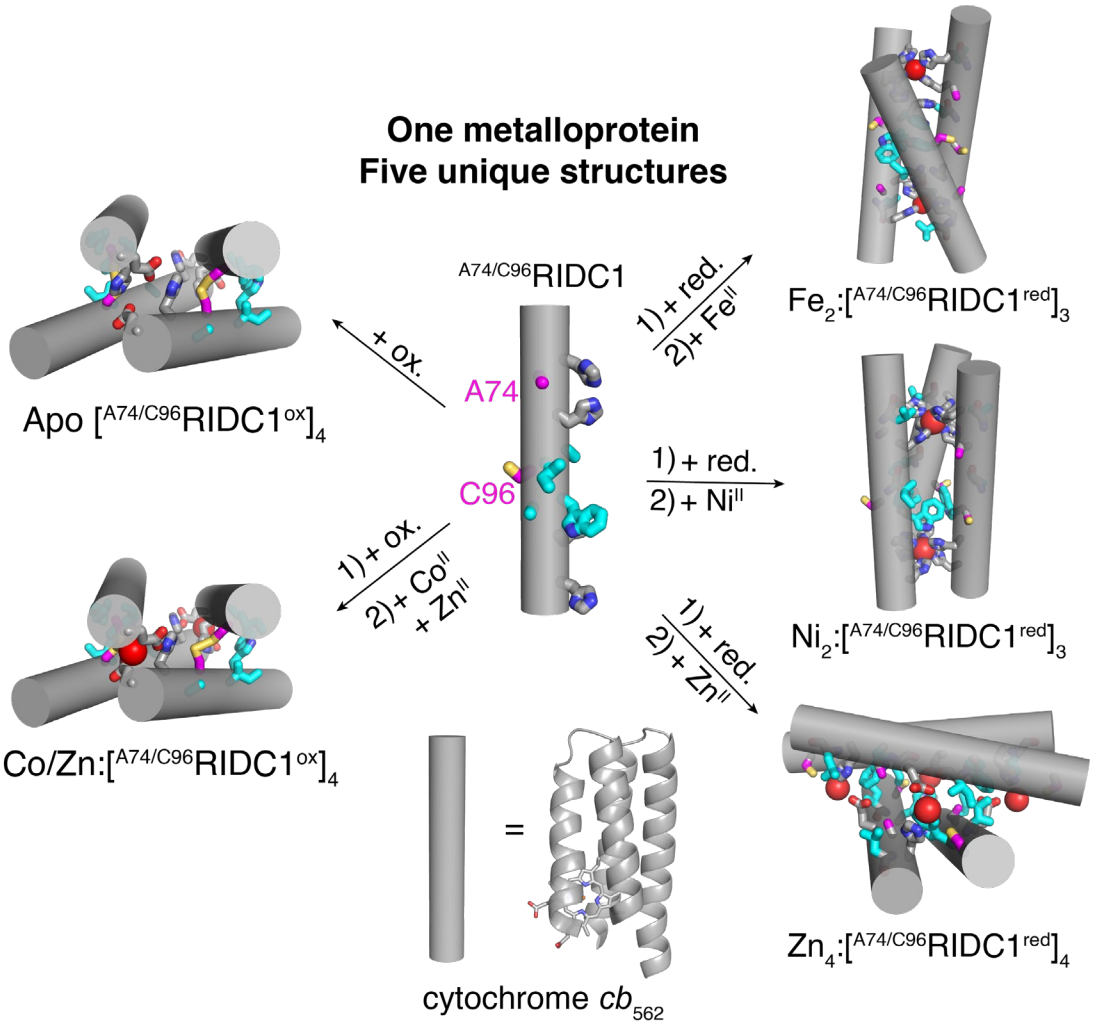

Figure 5.3 | Structural states of $^{A74/C96}$RIDC1 obtained through the addition of redox and/or metal-based stimuli. Hydrophobic mutations are highlighted in cyan. Ox. = oxidant, red. = reductant. The oligomeric states of the assemblies were determined using SV-AUC and X-ray crystallography.

5.5 References

1. Zygiel, E. M.; Nolan, E. M., Transition Metal Sequestration by the Host-Defense Protein Calprotectin. *Annu. Rev. Biochem.* **2018,** *87*, 621-643.
2. Subramanian Vignesh, K.; Deepe, G. S., Jr., Metallothioneins: Emerging Modulators in Immunity and Infection. *Int. J. Mol. Sci.* **2017,** *18* (10).
3. Sly, W. S.; Hu, P. Y., Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu. Rev. Biochem.* **1995,** *64*, 375-401.
4. Rutledge, H. L.; Tezcan, F. A., Electron Transfer in Nitrogenase. *Chem. Rev.* **2020,** *120* (12), 5158-5193.
5. Nelson, N.; Ben-Shem, A., The complex architecture of oxygenic photosynthesis. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (12), 971-82.
6. Chalkley, M. J.; Mann, S. I.; DeGrado, W. F., De novo metalloprotein design. *Nat. Rev. Chem.* **2021,** *6* (1), 31-50.
7. Davis, H. J.; Ward, T. R., Artificial Metalloenzymes: Challenges and Opportunities. *ACS Cent. Sci.* **2019,** *5* (7), 1120-1136.
8. Miller, D. C.; Athavale, S. V.; Arnold, F. H., Combining chemistry and protein engineering for new-to-nature biocatalysis. *Nat. Synth.* **2022,** *1* (1), 18-23.
9. Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F. A., Metal-Templated Design of Chemically Switchable Protein Assemblies with High-Affinity Coordination Sites. *Angew. Chem., Int. Ed. Engl.* **2020,** *59* (49), 21940-21944.
10. Rittle, J.; Field, M. J.; Green, M. T.; Tezcan, F. A., An efficient, step-economical strategy for the design of functional metalloproteins. *Nat. Chem.* **2019,** *11* (5), 434-441.
11. Hoffnagle, A. M.; Eng, V. H.; Markel, U.; Tezcan, F. A., Computationally Guided Redesign of a Heme-free Cytochrome with Native-like Structure and Stability. *Biochemistry* **2022,** *61* (19), 2063-2072.
12. Crane, B. R.; Siegel, L. M.; Getzoff, E. D., Sulfite reductase structure at 1.6 A: evolution and catalysis for reduction of inorganic anions. *Science* **1995,** *270* (5233), 59-67.
13. Sendovski, M.; Kanteev, M.; Ben-Yosef, V. S.; Adir, N.; Fishman, A., First structures of an active bacterial tyrosinase reveal copper plasticity. *J. Mol. Biol.* **2011,** *405* (1), 227-37.
14. Klein, A. S.; Zeymer, C., Design and engineering of artificial metalloproteins: from de novo metal coordination to catalysis. *Protein Eng. Des. Sel.* **2021,** *34*.
15. Andreini, C.; Cavallaro, G.; Lorenzini, S.; Rosato, A., MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res.* **2013,** *41* (Database issue), D312-9.
16. Davies, S. C.; Durrant, M. C.; Hughes, D. L.; Richards, R. L.; Sanders, J. R., Iron, cobalt and vanadium complexes of the N(CH2CH2S)33− ligand with chloride, azide, cyanide and carbonyl co-ligands. *J. Chem. Soc., Dalton Trans.* **2000,** (24), 4694-4701.
17. Lu, T. T.; Chiou, S. J.; Chen, C. Y.; Liaw, W. F., Mononitrosyl tris(thiolate) iron complex [Fe(NO)(SPh)3]- and dinitrosyl iron complex [(EtS)2Fe(NO)2]-: formation pathway of dinitrosyl iron complexes (DNICs) from nitrosylation of biomimetic rubredoxin [Fe(SR)4]2-/1- (R = Ph, Et). *Inorg. Chem.* **2006,** *45* (21), 8799-806.
18. Chambers, M. B.; Groysman, S.; Villagran, D.; Nocera, D. G., Iron in a trigonal tris(alkoxide) ligand environment. *Inorg. Chem.* **2013,** *52* (6), 3159-69.
19. Rosenbaum, D. M.; Rasmussen, S. G.; Kobilka, B. K., The structure and function of G-protein-coupled receptors. *Nature* **2009,** *459* (7245), 356-63.

20. Groitl, B.; Jakob, U., Thiol-based redox switches. *Biochim. Biophys. Acta* **2014,** *1844* (8), 1335-43.
21. Chang, Y. G.; Cohen, S. E.; Phong, C.; Myers, W. K.; Kim, Y. I.; Tseng, R.; Lin, J.; Zhang, L.; Boyd, J. S.; Lee, Y.; Kang, S.; Lee, D.; Li, S.; Britt, R. D.; Rust, M. J.; Golden, S. S.; LiWang, A., Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **2015,** *349* (6245), 324-8.
22. Fillat, M. F., The FUR (ferric uptake regulator) superfamily: diversity and versatility of key transcriptional regulators. *Arch. Biochem. Biophys.* **2014,** *546*, 41-52.
23. Ghanbarpour, A.; Pinger, C.; Esmatpour Salmani, R.; Assar, Z.; Santos, E. M.; Nosrati, M.; Pawlowski, K.; Spence, D.; Vasileiou, C.; Jin, X.; Borhan, B.; Geiger, J. H., Engineering the hCRBPII Domain-Swapped Dimer into a New Class of Protein Switches. *J. Am. Chem. Soc.* **2019,** *141* (43), 17125-17132.
24. Salgado, E. N.; Lewis, R. A.; Mossin, S.; Rheingold, A. L.; Tezcan, F. A., Control of protein oligomerization symmetry by metal coordination: C2 and C3 symmetrical assemblies through Cu(II) and Ni(II) coordination. *Inorg. Chem.* **2009,** *48* (7), 2726-8.
25. Signarvic, R. S.; DeGrado, W. F., De novo design of a molecular switch: phosphorylation-dependent association of designed peptides. *J. Mol. Biol.* **2003,** *334* (1), 1-12.
26. Ballister, E. R.; Lai, A. H.; Zuckermann, R. N.; Cheng, Y.; Mougous, J. D., In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (10), 3733-8.
27. Kakkis, A.; Golub, E.; Choi, T. S.; Tezcan, F. A., Redox- and metal-directed structural diversification in designed metalloprotein assemblies. *Chem. Commun.* **2022,** *58* (49), 6958-6961.

Appendix 1: Script and derivation used for generating fits of metal-binding isotherms

A.1 Script 1: This Dynafit script[1] describes the competition between Fura-2 and TriCyt3 for Co^{II} assuming a single metal binding site on the protein. Scripts identical to this one were used to generate fits of Mn^{II}-, Ni^{II}-, Cu^{II}-, and Zn^{II}-binding isotherms. Annotations are in bold.

```
[task]
task = fit
data = equilibria
; L ... Fura-2
; P ... TriCyt3
; M ... Co

[mechanism]
L + M <==> LM : Kd1 dissoc
P + M <==> PM : Kd2 dissoc

[constants]
Kd1 = 8.6E-9        Published Kd of Co(II):Fura-2 complex
Kd2 = 1E-7 ?        Initial estimate of the Kd of Co(II):TriCyt3 complex

[concentrations]
L = 10e-6
[responses]
L = 30000 ? LM = 50000 ?

[data]
directory ./data_directory
extension txt
variable M
file | concentration P = 10e-6
```

[output]

directory ./output_directory

[end

A.1 Derivation 1: Competitive binding for metal (Co^{II} in this example) between Fura-2 and TriCyt3 can be described by a cubic equation whose derivation is described below.[2] The cubic fit of metal-binding isotherms can only be used when assuming a single metal binding site on the protein.

The cubic equation can be derived based on the following initial equations:

(1) $K1 = [M] * \frac{[F]}{[MF]}$

(2) $K2 = [M] * \frac{[P]}{[MP]}$

(3) $[Mtot] = [M] + [MF] + [MP]$

(4) $[MP] = [Ptot] - [P]$

(5) $[MF] = [Ftot] - [F]$

where M represents Co^{II}, P represents TriCyt3, and F represents Fura-2
From equations (2) and (4):

(6) $[MP] = [Ptot] * \frac{[M]}{[M]+ K2}$

From equations (1) and (5):

(7) $[MF] = [Ftot] * \frac{[M]}{[M]+ K1}$

Substituting (6) and (7) into (5) and forming this into a cubic equation for [M] gives:

(8) $[M]^3 + [M]^2 * a + [M] * b + c = 0$ with

(9) $a = [Ptot] + K1 + K2 + [Ftot] - [Mtot]$

(10) $b = K1 * K2 + K1 * ([Ptot] - [Mtot]) + K2 * ([Ftot] - [Mtot])$

(11) $c = -K1 * K2 * [Mtot]$

We then can solve for [M]:

(12) $[M] = -\frac{a}{3} + \frac{2}{3} * \sqrt{(a^2 - 3b)} * \cos\left(\frac{\theta}{3}\right)$ with

(13) $\theta = arcos\left(\frac{(-2a^3 + 9ab - 27c)}{2 * \sqrt{(a^2 - 3b)^3}}\right)$

Substituting (12) in (7) gives

(14) $[MF] = \frac{[Ftot] * (2 * \sqrt{a^2 - 3b} * \cos\left(\frac{\theta}{3}\right) - a)}{3 * K1 + \left(2 * \sqrt{a^2 - 3b} * cos\left(\frac{\theta}{3}\right) - a\right)}$

In our setup the absorbance of Co^{II} alone can be assumed to be 0 at 342 nm and similarly the absorbance of TriCyt3 at 342 nm can be assumed to be constant with or without metal bound. Therefore, any change in absorbance (ΔA) during the titration experiment can be attributed to the change in absorbance at 342 nm for Mag-Fura/Mag-Fura2 in its free and metal-bound state, respectively.

(15) $\Delta A = [MF] * (\epsilon_{MF} - \epsilon_F)$

With K1, [Ptot], and [Ftot] known and constant, the sum of squared deviations between experimental and calculated values for [Mtot] can be minimized by adjusting the dissociation constant for the Metal:TriCyt3 complex (K2).

References

1. Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A., Evolution of Metal Selectivity in Templated Protein Interfaces. *J. Am. Chem. Soc.* **2010,** *132* (25), 8610-8617.
2. Wang, Z.-X., An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.* **1995,** *360* (2), 111-114.