UNIVERSITY OF CALIFORNIA, SAN DIEGO

Uncovering Allostery in a Uniquely Folded Metalloprotein

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

in

Chemistry and Biochemistry

by

Elizabeth Leigh Baxter

Committee in Charge:

Professor Patricia A. Jennings, Chair
Professor Cliff Kubiak
Professor Andrew Mccammon
Professor Jose N. Onuchic
Professor Stanley J. Opella
Professor Wei Wang

2013

The Dissertation of Elizabeth Leigh Baxter is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2013

DEDICATION:

I would like to dedicate this thesis to my family;

To my dad, for sparking my passion for science with dinner table conversations about space, theoretical physics, and orbital mechanics;

To my mom, for believing in me, and teaching me to never quit and to never settle;

To my sister, for all of our adventures, past and future, and for being one of the funniest people I know;

Finally, to Aurash, for all of his support and encouragement, and for inspiring me with his capacity for friendship and love.

EPIGRAPH

The best way out is always through.

Robert Frost

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

β1	beta strand 1
CISD	CDGSH iron sulfur domain
Csk	C-terminal Src kinase
$E_{m,7}$	midpoint/redox potential
E. coli	Escherichia coli
EDTA	ethylene diamine tetraacetic acid
ENaC-γ	Epithelial Na+ Channel γ
Fe-S	iron sulfur
GROMACS	Groningen Machine for Chemical Simulations
IL-1β	interleukin-1 beta
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	luria broth
L1	loop 1
L2	loop 2
MD	molecular dynamics
MP	metalloprotein
Naf-1	nutrient-deprivation autophagy factor-1
NMR	Nuclear Magnetic Resonance
OMM	Outer Mitochondrial Membrane
PEG	polyethylene glycol
PDB	protein data bank

PPAR-γ	peroxisome proliferator-activated receptor subtype gamma
PsTZDs	PPARy-sparing TZDs (PsTZDs)
Q	fraction of native contacts
RPM	revolutions per minute
SBM	Structure-Based Model
SHE	standard hydrogen electrode
SMOG	Structure-based MOdels in Gromacs
SSRL	Stanford Synchrotron Radiation Lightsource
T_f	Folding Temperature
WHAM	Weighted Histogram Analysis Method
WT	wild type

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my Advisors, Jose Onuchic and Patricia Jennings. The completion of this dissertation would not have been possible without their support and guidance, and perhaps most importantly, their patience and good humor.

I feel privileged to have been a part of the Jennings Lab and the Onuchic Research Group. It's rare to have the chance to work with so many warm, intelligent, and talented individuals.

Chapter 3, in part, is a reprint as it appears in Proceedings of the National Academy of Sciences 2011. Baxter EL, Jennings PA, Onuchic JN. 2011. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is a reprint as it appears in Proceedings of the National Academy of Sciences 2011. Baxter EL, Jennings PA, Onuchic JN. 2011. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in part, is a reprint as it appears in Proceedings of the National Academy of Sciences, 2012. Baxter EL, Zuris JA, Wang C, Vo PLT, Axelrod HL, Cohen AE, Nechushtai R, Onuchic JN, Jennings PA. 2012. The dissertation author was the primary investigator and author of this paper.

VITA

2006 B.S., Biochemistry and Molecular Biology, University of California, Santa Cruz

2009 M.S., Chemistry, University of California, San Diego

2013 Ph.D., Chemistry, University of California, San Diego

Publications

Baxter EL, Zuris JA, Wang C, Vo PL, Axelrod HL, Cohen AE, Nechushtai R, Onuchic JN, Jennings PA, Allosteric control in a metalloprotein dramatically alters function. *Proc. Natl. Acad. Sci. USA* **2013,** 110 (3), 948-953;

Baxter EL, Jennings PA, & Onuchic JN, Strand swapping regulates the 2Fe-2S cluster in the diabetes drug target mitoNEET. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (6), 1955-1960;

Baxter EL, Jennings PA, & Onuchic JN, Interdomain communication revealed in the diabetes drug target mitoNEET. *Proc. Natl. Acad. Sci. USA* **2011**, *108* (13), 5266-5271;

ABSTRACT OF DISSERTATION

Uncovering Allostery in a Uniquely Folded Metalloprotein

by

Elizabeth Leigh Baxter

Doctor of Philosophy in Chemistry

University of California, San Diego, 2013

Professor Patricia A Jennings, Chair

Proteins are linear chains of amino acids that self-assemble, or "fold", into 3-dimensional structures which perform tasks in organisms. Nature has selected for sequences which fold quickly and efficiently into their functional structure. While it is important for organisms to have proteins which fold efficiently, it is essential that

proteins also perform their function. These two evolutionary pressures are sometimes in conflict, and perfect optimization of both folding and function may not always be possible. For example, a mutation which slows folding may be selected for if it enhances protein function, therefore increasing the fitness of the organism. As a result, regions of a protein that contribute to slow folding may be critical for function.

Identifying regions that contribute to slow folding may be an effective way to predict and identify sites that are critical to a proteins function. This dissertation uses this approach to characterize mitoNEET, a protein which is implicated in diabetes, aging, cancer, and obesity. MitoNEET was discovered because it unexpectedly binds the commonly prescribed diabetes drug Pioglitazone. The protein is a uniquely folded homodimer, and each protomer coordinates two [2Fe-2S] clusters. These iron-sulfur clusters are capable of electron transfer, and the cluster itself may be transferred to acceptor partner proteins. Understanding how this protein regulates its metal centers is critical for better drug design.

We used structure-based models to simulate the folding of mitoNEET, and observed that a loop far removed from the metal center creates a constraint which slows the folding of the protein. We predicted that this region was evolutionarily conserved, and that mutations to this site would disrupt the function of the protein. To test this theory, we used mutagenesis to introduce perturbations at this site. We observed that changes to this loop alter the rate of cluster transfer, cluster decay, and the redox potential. This result is striking because properties of metal centers are traditionally thought to be controlled by the small fraction of amino acids which directly surround

them. Our work challenges this paradigm, and we feel it opens the door to more intelligent drug design for this class of proteins.

Chapter 1

Introduction

DIABETES AND MITONEET

As of 2011, The Center for Disease Control estimates that 25.8 million people in the United States have diabetes, with type 2 diabetes comprising between 90-95% of these cases (1). Type 2 diabetes is typically caused by obesity and a lack of exercise, and is characterized by high blood glucose levels as a result of insulin resistance (2, 3). While there are many drug treatments available for type 2 diabetes, a class of insulin sensitizing agents known as Thiazolidinediones (TZDs) are some of the most commonly prescribed.

TZDs are high affinity ligands for the transcription factor peroxisome proliferator-activated receptor subtype γ (PPARγ) (4), and for years the predominantly held paradigm was that the beneficial effects of these drugs are exerted through activation of PPARγ (5-7). Unfortunately many of the side effects of TZD's are also PPARγ mediated. Activation of PPARγ in collecting duct tissue directly increases expression of the epithelial Na+ channel γ (ENaC-γ), resulting in the increased fluid retention commonly observed with TZD treatment (8), and collecting duct PPARγ knockout mice are protected against increases in body weight and plasma volume expansion associated with TZDs (9). Tissue specific PPARγ knockout mice show resistance to the increased adiposity (10, 11) and hepatic steatosis (12) associated with the TZD rosiglitazone.

To further complicate things, recent evidence suggests that some of the insulin sensitizing effects of TZDs are exerted in a PPARγ independent manner. Some of the effects of TZDs occur too quickly to be dependent on PPARγ -mediated transcription events (13), and the ability of TZDs to activate PPARγ is not predictive of its efficacy as an insulin sensitizing agent (14-16). Additionally, studies with muscle, liver, and adipose

tissue PPARγ knockout mice show that TZDs are able to exert beneficial effects independent of PPARγ binding, and TZDs designed to be PPARy-sparing TZDs (PsTZDs) are able to ameliorate insulin resistance and inflammation in obese mice (17). These studies suggest alternate molecular targets exist for TZDs, and by targeting these specifically many of the side effects associated with PPARγ activation can be avoided (17).

In 2004 an alternate target was identified for the commonly prescribed TZD pioglitazone, or Actos as it is more commonly known. Colca *et al.* identified a 17kda protein which specifically cross-linked a radiolabeled photoaffinity derivative of pioglitazone (18). This protein was named mitoNEET based on the fact that it was localized to the mitochondrial fraction of rodent brain, liver, and skeletal muscle, and because it has a rare "NEET" sequence in its C-terminus (18). The protein has a predicted transmembrane sequence in its N terminus, and mitochondrial fractionation and protease sensitivity experiments show that the protein is localized to the outer mitochondrial membrane (OMM), with a soluble domain exposed to the cytosol (19). Optical and electron paramagnetic spectroscopy studies demonstrated that the protein contains a [2Fe-2S] cluster, making it the first discovered iron containing protein localized to the OMM (20).

Unsurprisingly, mitoNEET and its homologues are critical in health and disease. MitoNEET is strongly implicated in diabetes; in addition to binding a type 2 diabetes drug target, mouse studies have demonstrated that varying expression levels of mitoNEET regulates insulin sensitivity as well as lipid homeostasis (21). In mice fed a high fat diet, overexpression of mitoNEET in adipocytes resulted in greater expansion of

adipose tissue and weight gain than in their nontransgenic littermates (21). Despite the marked obesity observed in the overexpressing mice, insulin sensitivity was preserved when compared to their counterparts (21). Additionally, overexpression of mitoNEET blocked transport of iron in to the mitochondria, and by doing so inhibited electron transport (21). Reduction of mitoNEET expression enhances mitochondrial respiratory capacity and reduces the weight gain associated with a high fat diet. While these mice are thinner, they experience heightened oxidative stress and glucose intolerance (21).

In addition to expressing MitoNEET (encoded by the gene CISD1) humans express two other mitoNEET like proteins, Naf-1, (encoded by CISD2) and Miner2 (encoded by CISD3). Mice with a deficiency in Naf-1 show signs of accelerated aging, including blindness, muscular degeneration, and mitochondrial degeneration and autophagy (22). A mutation in CISD2, the gene encoding Naf-1, is causative of wolframs syndrome (23). The symptoms of wolframs syndrome include diabetes insipidus, diabetes mellitus, optic atrophy, and deafness (24, 25). A single base pair transversion in the encoding gene CISD2 causes a splicing error which results in a truncated form of Naf-1 (23). The truncated form of Naf-1 lacks the domain which coordinates the [2Fe-2S] cluster, resulting in an apo-protein.

When the crystal structure of mitoNEET was solved, it was revealed that the protein had a completely unique fold among the pdb database (26-28). (Figure 1.1) As such, the protein became the defining member of the CDGSH iron sulfur domain (CISD) family of proteins, of which the consensus sequence is C-X-C-X_2-(S/T)-X_3-P-X-C-D-G-(S/A/T)-H. This family of proteins has been conserved throughout evolution and members are seen in eukaryotes, archaea, and bacteria. MitoNEET is homodimeric, with

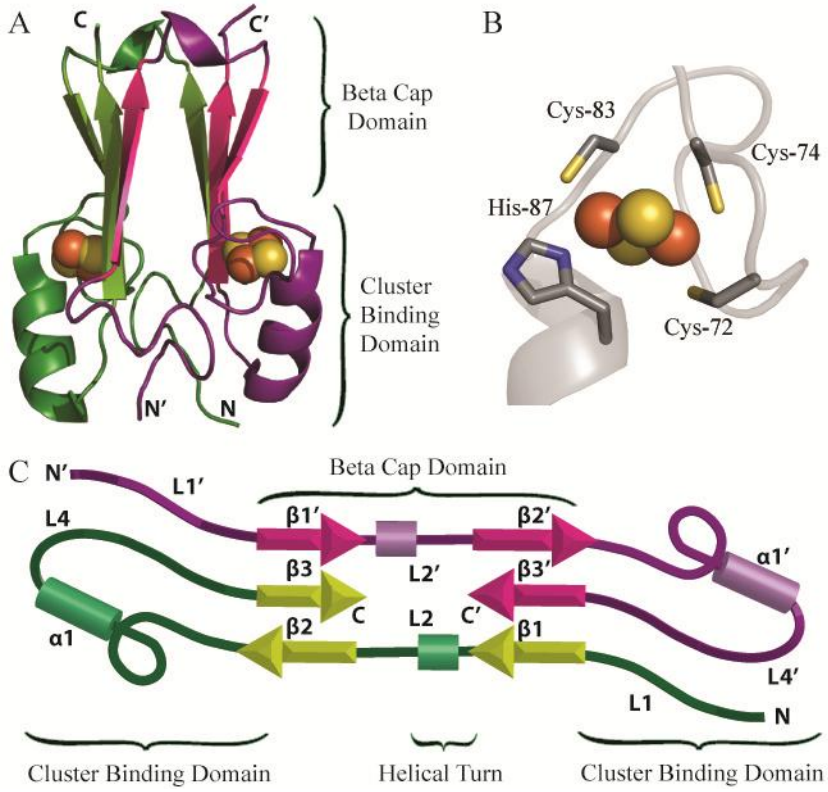

Figure 1.1: Structural organization and domain topology of mitoNEET
A) Ribbon diagram of mitoNEET colored by chain, showing the location of the beta cap domain and the cluster binding domain. B) The [2Fe-2S] cluster is coordinated by 3 cysteines and one histidine. C) Splay diagram of mitoNEET colored by chain. The beta cap domain consists of a beta sandwich composed of two three stranded beta sheets. Each beta sheet is composed of two strands from one protomer and a third strand from the second protomer.

each protomer coordinating a [2Fe-2S] cluster (Figure 1.1). The two protomers intertwine via an unusual strand swap to form two domains, a beta cap domain which consists of two three stranded beta sheets sandwiched together, and a cluster binding domain which coordinates the [2Fe-2S] clusters. (Figure 1.1) Other members of the CDGSH family of proteins exhibit variations to this topology (29). For example, the arrangement of strands and the number of strands in the beta cap domain may vary, and one monomeric structure has been characterized. However, all structures characterized to date have two conserved, opposing cluster binding pockets in the cluster binding domain, and a beta sandwich (26-31).

The metal centers in the CDGSH family of proteins are of particular interest. The majority of [2Fe-2S] clusters have either a rieske type 2Cys-2His coordination, or a ferredoxin like 4Cys coordination. MitoNEET and other CDGSH family members coordinate their [2Fe-2S] clusters with 3 cysteines and 1 solvent exposed histidine (26-31). (Figure 1.1B) MitoNEET's [2Fe-2S] clusters have been shown to be transferrable (32, 33), redox active, and labile, suggesting multiple possible functional roles for mitoNEET. The protein is capable of cluster transfer *in vitro* and *in vivo*, and TZD binding blocks this process (32, 33). The clusters have a redox potential of +25 mV at pH 7.0, however the redox potential decreases with increasing pH (34, 35). This sensitivity to the cellular environment suggests that mitoNEET may play a role as a sensor. The pH sensitivity may be attributable to the solvent exposed coordinating histidine, which can be protonated at low pH (36, 37). Strikingly, TZD binding also modulates the redox potential of the protein (35).

Since mitoNEET is not only a new drug target, but also has a completely unique fold, it is essential to understand how its structure regulates the properties of its [2Fe-2S] clusters, and thus, its function. In order to accomplish this, we turn to energy landscape theory and Structure-Based Models.

ENERGY LANDSCAPE THEORY

It's been estimated that if a random amino acid chain were to fold by sampling all possible configurations, it would take roughly the age of the universe for it to find its correctly folded structure (38). In reality, many proteins are observed to fold on the order of seconds. Energy landscape theory addresses this discrepancy by suggesting that organisms have evolved to express proteins which fold in a funneled fashion with minimal frustration (39-41). (Figure 1.2A, B) In minimally frustrated sequences, interactions between amino acids in the native structure are mutually attractive and thus stabilizing to the native state (40). As proteins organize in to structures in the search for the native state, structures that are similar to the native state are stabilized by the attractive force between native amino acid pairs and the energy of the system is reduced. Non-native interactions are kept to a minimum, so that structures dissimilar to the native state have fewer attractive interactions and are higher energy. Thus, native interactions guide the folding process to the correct native state resulting in a funneled landscape with few energetic traps.

Energy landscape theory provides the foundation for a class of molecular dynamics hamiltonians called Structure-Based Models (SBMs). In a SBM, structural data derived from a crystal structure or NMR structure is included in the Hamiltonian to provide a bias towards the native basin. Bonds, angles, and dihedrals in the model are set

so that the lowest energy state is that which is present in the native state. Depending on the type of SBM, pairs of atoms or residues that are in close proximity in the native structure are assumed to be native interactions and are assigned an attractive force, while non-native interactions are assigned a short range repulsive term. By imposing a native bias and excluding non-native interactions, these simplified models effectively use the principal of minimal frustration to model a funneled energy landscape. Because SBM's ignore energetic frustration, they are able to capture geometric constraints in folding and reflect the intrinsic difficulty in folding to a particular shape. These models have successfully captured the free energy barriers to folding and the folding mechanism of many systems (42-47).

SBMs have also proven applicable in furthering our understanding of the functional dynamics of proteins (48-51). This is because protein folding and protein function are linked, and functional information is contained in the ruggedness of the folding funnel (Figure 1.2 C). A simple way to understand the link between folding and function is from an evolutionary standpoint. While it is important for organisms to have proteins that fold quickly and efficiently, it is essential to the survival of the organism that proteins also perform their function. These two evolutionary pressures are sometimes in conflict, and perfect optimization of both efficient folding and function many not always be possible. The resulting energy landscape is still funneled, but may include frustration in the form of local energetic traps or roughness in the funnel. (Figure1.2C)

As a result, residues in a proteins sequence that impact folding efficiency may have been evolutionarily conserved, or even selected for. For example, a random

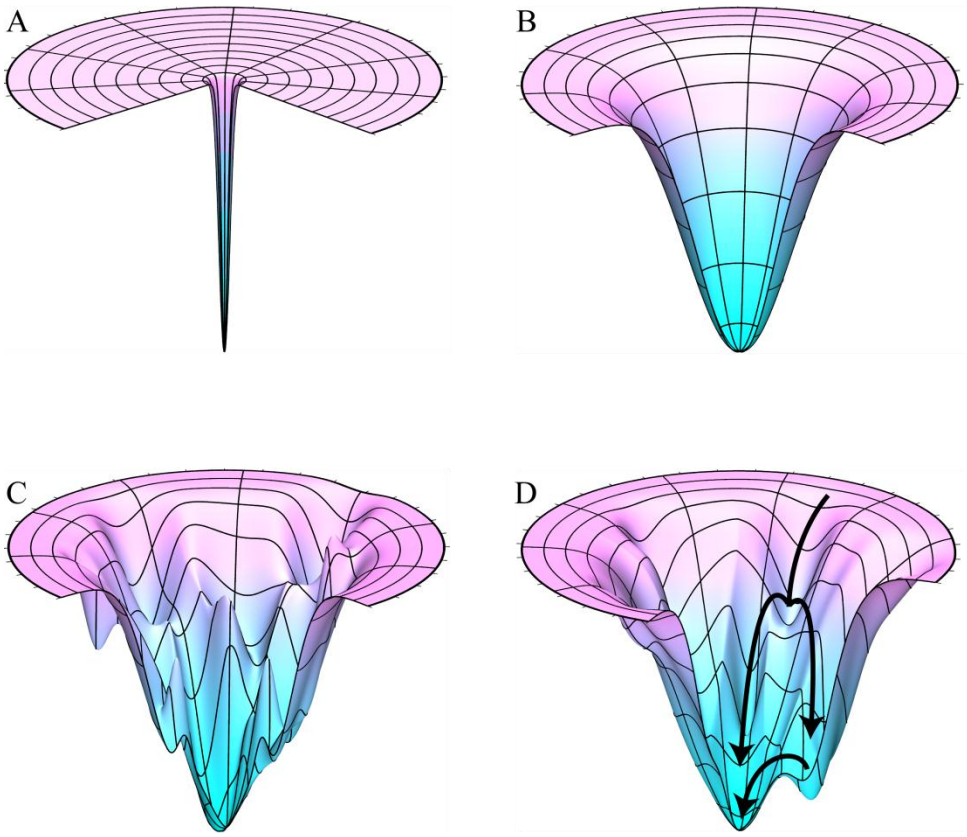

Figure 1.2: Funneled energy landscapes
A) A representation of a random search of conformational space in which the sampled configurations do not funnel the protein closer to the native state. B) A perfectly funneled energy landscape, in which all sampled configurations guide the folding towards the native basin. C) A rugged funneled landscape, in which some of the sampled configurations exist as traps, or local energetic minima. D) A funneled landscape in which a higher energy functional state is accessible under native conditions. These states may sometimes be accessed during folding as part of an alternate pathway.

mutation in a protein sequence may result in a protein which folds faster, but the same mutation may also compromise the function of the protein, and therefore the fitness of the organism. Alternately, a mutation which compromises folding may be selected for if it enhances the function of the protein and the fitness of the organism. A recent study showed that energetic frustration in proteins co-localizes with cofactor binding sites, suggesting that stability in these regions was compromised for function (52). Another example of this phenomenon is seen in WW domain proteins. Mutations to functional loops are able to speed up folding at the expense of function, and in some cases remove the barrier to folding completely (53, 54).

An alternate way to understand the link between folding and function is to realize that proteins perform their function on the same landscape that they fold on. In energy landscape theory, the native state exists as an energetic minimum at the bottom of the folding funnel. The "native state" of a protein is a bit of a misnomer because it implies that there is only one accessible state under native conditions, when in fact many proteins have multiple functional states accessible under native conditions. In some systems these functional states exhibit much greater degrees of disorder than the energetic minimum of the protein and can be represented as higher energy local minima which may be accessed from the lowest energy state (55) (Figure 1.2D). After a protein makes its way to the native state, it performs its functional motions at the bottom of this funnel, and periodically accesses higher energy functional states. Because these states may be short lived and disordered they may be difficult to characterize using traditional techniques such as circular dichroism or x-ray crystallography. In many systems these functional states are also accessed during the folding process, and may show up as folding

intermediates (55) or as alternate pathways in folding. Therefore characterizing the folding landscape of a protein can give us important clues about the functional motions and functional states of a protein.

Identifying regions which contribute to slow folding may be an effective way to predict and identify sites which are critical to a proteins function. Furthermore, a thorough characterization of the folding landscape of a protein, including the traps, intermediates, and alternate pathways may provide information about the functional states and functional motions of the folded protein. This dissertation seeks to make predictions about the structure function relationship in mitoNEET using energy landscape theory and structure based models, and validates these predictions with experimental data.

Chapter 2

General Methods

INTRODUCTION

In order to characterize the relationship between folding and function in mitoNEET, we use a combined theoretical and experimental approach. All simulations were performed with SBMs run with the GROMACS, however the Hamiltonian used varies between a coarse grained C_α model or an all-atom model depending on the type of simulations. To simulate the folding of mitoNEET, we used a coarse grained C_α model with the [2Fe-2S] cluster modeled implicitly. Simulations of the native basin dynamics and the functional motions of the protein were performed with an all-atom Hamiltonian.

Experimental work includes site directed site directed mutagenesis of mitoNEET, and expression and purification of mitoNEET and apo-ferredoxin. Crystallographic studies were used to examine whether mutagenesis to mitoNEET resulted in structural changes. UV-visible spectroscopy was used to characterize changes to the properties of mitoNEET's [2Fe-2S] clusters upon mutagenesis.

GENERAL METHODS

Coarse Grain Structure-Based Modeling

All MD simulations were performed using the GROMACS software package (56). Simulations were performed using a SBM in which the crystal structure of the protein is the lowest energy state. In the coarse grained C_α model, each residue is treated as a single bead located at the C_α position, and interactions occur between residues, rather than individual atoms. The functional form of the C_α Hamiltonian is:

$$V = \sum_{bonds} \varepsilon_r \, (r - r_0)^2 + \sum_{angles} \varepsilon_\theta \, (\theta - \theta_0)^2 + \sum_{backbone} \varepsilon_{BB} F_D(\phi - \phi_0)$$

$$+ \sum_{contacts} \varepsilon_C \left[5 \left(\frac{\sigma_{ij}}{r} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r} \right)^{10} \right] + \sum_{non-contacts} \varepsilon_{NC} \left(\frac{\sigma_{NC\,ij}}{r} \right)^{12}$$

Native contacts are determined using the SBMs in GROMACS (SMOG) Web server (57) with the Contacts of Structural Units (CSU) software (58). The distance between beads is given by r, and the native distance between contacts is given as σ_{ij}. Residues that are not determined to be in contact are defined as non contacts, and are assigned a hard wall interaction with a radius of σ_{NC} = 4Å. Native bond lengths between C_α beads are represented by r_0. Three body angles are represented by θ and four body angles are represented by ϕ. The backbone potential F_D is:

$$F_D(\phi) = [1 - \cos(\phi)] + \frac{1}{2}[1 - \cos(3\phi)]$$

The coefficients for each term are scaled so that $\varepsilon_r = 100\varepsilon$, $\varepsilon_\theta = 40\varepsilon$, $\varepsilon_D = \varepsilon_C = \varepsilon_{NC} = \varepsilon = k_B$. For simplicity, the [2Fe-2S] cluster was not modeled in with beads in the C_α model, instead the cluster was included implicitly by the contacts present between the cluster coordinating residues.

All-Atom Structure Based Modeling

In the all-atom potential, all atoms are modeled as beads, with the exception of hydrogen atoms. The functional form of the all-atom Hamiltonian is:

$$V = \sum_{bonds} \varepsilon_b \, (r - r_0)^2 + \sum_{angles} \varepsilon_\theta \, (\theta - \theta_0)^2 + \sum_{\frac{impropers}{planars}} \varepsilon_\chi \, (\chi - \chi_0)^2$$

$$+ \sum_{backbone} \varepsilon_{BB} F_D(\phi) + \sum_{sidechain} \varepsilon_{SC} F_D(\phi) + \sum_{contacts} \varepsilon_C \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r} \right)^{6} \right]$$

$$+ \sum_{non-contacts} \varepsilon_{NC} \left(\frac{\sigma_{NC\,ij}}{r} \right)^{12}$$

In this equation, r_0 represents the distances between atoms considered "in contact" in the native state, as defined by the crystal structure of the protein. In the all-atom model, atoms are said to be in contact if they are within 6Å from each other. Additionally, our model employs the Shadow algorithm; to be considered in contact, there must not be any other atoms in between the contact pair, they must be able to "see" each other (59). Atoms considered not in contact are assigned a hard wall interaction with a radius of σ_{NC} = 2.5Å. Native bond lengths between atoms are represented by r_0. Three body angles are represented by θ and four body angles are represented by ϕ. Improper dihedrals, represented by χ, maintain side chain planarity and backbone chirality. The backbone and side chain terms are scaled so that the ratio $\frac{\varepsilon_{BB}}{\varepsilon_{SC}}$ is set to 2. Similarly, the dihedral and contact terms are scaled so that the ratio $\frac{\sum \varepsilon_C}{\sum \varepsilon_{BB} + \sum \varepsilon_{SC}}$ is set to 2. Finally, the stabilizing energy is set so that $\sum \varepsilon_C + \sum \varepsilon_{BB} \sum \varepsilon_{SC} = \varepsilon N_{atoms}$. Additional coefficient scaling is $\varepsilon_b = 100\varepsilon$, $\varepsilon_\theta = 20\varepsilon$, $\varepsilon_\chi = 10\varepsilon$, $\varepsilon_{NC} = \varepsilon = k_B T^* = 1$. For ring planarity, $\varepsilon_\chi = 10\varepsilon$.

Bacterial Growth and Expression of MitoNEET

Plasmids containing the soluble domain of mitoNEET or Naf-1 in kanamycin resistant bacterial expression vector pET-28a(+)(Novagen) were obtained from Andrea

Conlan. The expression vectors contain an N-terminal thrombin cleavable histidine tag. Point Mutations were made with full plasmid PCR using overlapping primers (IDT Technologies) containing the mutation of interest.

Plasmids were transformed in to BL21-CodonPlus(DE3)-RIL cells (Stratagene). Cell stocks containing the expression vector of interest were grown in LB supplemented with 30 µg/mL kanamycin and 34 µg/mL chloramphenicol at 32°C. At an OD600 of 0.3, 10mM $FeCl_3$ was added to a final concentration of 0.75mM. Protein expression was induced with 1mM IPTG at an OD600 of 0.8 and cells were allowed to grow at 25°C for an additional 10-15 hours. Cells were harvested by centrifugation at 4,000 RPM for 15 minutes. Harvested cells were resuspended in sonication buffer (25mM Tris-HCl pH8, 5mM Imidazole, 300mM NaCl) and lysed by sonication. The cell lysate was cleared by centrifugation for 30 min at 13000 rpm and the supernatant was added to Ni-NTA resin (Qiagen) and allowed to batch bind for 30 minutes at 4°C. The resin was washed with 10x volume of wash buffer (25mM Tris-HCl pH 8, 30mM Imidazole, 300mM NaCl) and then equilibrated in thrombin cleavage buffer (25mM Tris-HCl pH 8.0, 150mM NaCl, 2.5mM CaCl2). The His-tag was cleaved with thrombin for 10-15 hours at 4°C, and an additional 4 hours at 25°C. The protein was eluted with wash buffer and purified with cation exchange chromatography (HiTrap, GE Healthcare).

Spectroscopy and Crystal Growth

All UV-Visible absorption spectroscopy was performed on a Cary50 spectrometer (Varian Inc, Palo Alto CA) equipped with a temperature controlled cell (T = 35°C).

Crystals were grown at 19.6°C using the vapor diffusion method on CrystalClear D Strips 96 well sitting drop plates (Douglas Instruments). PEG 3000 was used as the

precipitating agent and the buffer used was 100mM Tris-HCl pH 8 and 100mM NaCl. Crystal samples were soaked for 1 minute in 100mM Tris-HCl pH 8, 40% PEG 3000, then flash frozen to 77K in liquid nitrogen. Samples were shipped at 77K to SSRL in a SSRL supplied cassette system.

All Structures were visualized with Pymol.

Expression and Purification of Apo-Ferredoxin

Cell stocks containing the ferredoxin expression vector in BL21-BE3 E. Coli cells were obtained from Rachel Nechushtai. The petF gene encoding mFd – the ferredoxin encoding gene of Cyanobatrium Mastigocladus laminosus (60) was contained in bacterial expression vector pET-20b. The expression vectors contain an N-terminal thrombin cleavable histidine.

Cells were grown in TB supplemented with 100 ug/mL ampicillin at 37°C. Protein expression was induced with 1mM IPTG at an OD600 of 0.6 and cells were allowed to grow at 25°C for an additional 10-15 hours. Cells were harvested by centrifugation at 4,000 RPM for 15 minutes. Harvested cells were resuspended in sonication buffer (20mM Tris-HCl pH8, 5mM Imidazole, 5mM $MgCl_2$) and lysed by sonication. The cell lysate was cleared by centrifugation for 30 min at 13000 rpm and the supernatant was added to Ni-NTA resin (Qiagen) and allowed to batch bind for 30 minutes at 4°C. The resin was washed with 10x volume of wash buffer (20mM Tris-HCl pH 8, 15mM Imidazole, 50mM NaCl). The resin was then resuspended in elution buffer (20mM Tris-HCl pH 8.0, 150mM Imidazole, 300mM NaCl) and the suspension was left to precipitate on a nutator for 30 min at 4°C. The protein was then eluted. The elution was diluted 3X with 20mM Tris-HCl pH 8.0, and then purified with anion exchange

chromatography (HiTrap, GE Healthcare). The apo-protein eluted as the second peak and was collected and repurified using a Sephacryl S-100 size exclusion column. (GE Healthcare)

Chapter 3

Interdomain Communication Revealed in the Diabetes Drug Target MitoNEET

ABSTRACT

MitoNEET is a recently identified drug target for a commonly prescribed diabetes drug, Pioglitazone. It belongs to a previously uncharacterized ancient family of proteins for which the hallmark is the presence of a unique 39 amino acid CDGSH domain. In order to characterize the folding landscape of this novel fold, we performed thermodynamic simulations on MitoNEET using a SBM. Additionally, we implement a method of contact map clustering to partition out alternate pathways in folding. This cluster analysis reveals a detour late in folding and enables us to carefully examine the folding mechanism of each pathway rather than the macroscopic average. We observe that tightness in a region distal to the iron–sulfur cluster creates a constraint in folding and additionally appears to mediate communication in folding between the two domains of the protein. We demonstrate that by making changes at this site we are able to tweak the order of folding events in the cluster binding domain as well as decrease the barrier to folding.

INTRODUCTION

MitoNEET is a recently identified outer mitochondrial membrane protein that unexpectedly binds the commonly prescribed type II diabetes drug Pioglitazone (18-20). It is now recognized as a new drug target in diabetes therapy as opposed to the traditional PPARγ therapeutics (61). Mis-splicing of Naf-1, the structural homolog of mitoNEET, results in the rare disease Wolfram syndrome that initially presents with diabetes and rapidly progresses to blindness and early death (23). In addition, Naf-1 appears to play a significant role in aging and associated diseases. MitoNEET and Naf-1 possess a unique homodimeric fold with a CDGSH iron–sulfur cluster binding domain and a strand swapped beta cap (20, 26-28, 30, 62). Because regulating the activity of this new drug target is an area of high interest, investigation of the folding and possible allosteric modulation of function in this family is now a major research focus.

Energy landscape theory indicates that proteins have evolved to fold in a funneled fashion with minimal frustration (39-41). Because energetic frustration is sufficiently small, much of the heterogeneity in folding is dominated by the geometric constraints of the native structure. As a result, SBMs are capable of capturing the main features of the transition state and intermediates formed during folding for many proteins (42-47). In addition, our analysis of the bottlenecks in folding have led to a deeper understanding of regulatory mechanisms operating in specific proteins. This led us to the hypothesis that functional regions in proteins may add roughness to the landscape because they are under separate evolutionary pressure than areas used for efficient folding. For example, structure-based simulations with adenylate kinase demonstrated that the introduction of frustration induced conformational transitions associated with enzymatic catalysis

through specific unfolding, or cracking (48, 49). Folding simulations with Csk and IL-1β successfully captured long range communication to functional sites (50, 51). Therefore folding studies provide a unique approach to explore the functional landscape of biomolecules.

As a first step toward defining not only the folding but also potential points of regulation in the NEET family, we initiated theoretical structure-based folding studies. A common approach to analyzing large quantities of data from structure-based simulations is to examine macroscopic averages at a point along a reaction coordinate, such as Q. However, many proteins can access multiple routes to folding (63-67). In these systems, this approach can become problematic because averaging over multiple pathways can obscure the actual folding events in each route. In these cases it is useful to examine reaction coordinates transverse to Q, but because the folding funnel is highly multidimensional, it is difficult to know which reaction coordinates to examine. We implement a scheme to investigate the dimeric transition state in SBMs by clustering contact maps. Clustering provides us with an efficient way of compacting and visualizing this high-dimensional space in two dimensions, enabling us to efficiently sort data and identify alternate folding routes that make up this ensemble. It also proves especially effective for handling averaging over duplicate pathways that result from symmetry in multimeric systems. We show that in mitoNEET, the average transition state ensemble is misleading when compared with the transition state ensemble after clustering. Additionally, clustering reveals the presence of a detour late in folding.

Clustering enables us to carefully examine the folding mechanism of mitoNEET. We observe that a helical turn in the top of the beta cap domain introduces frustration in

folding. Additionally, we see communication between the beta cap domain and the cluster binding domain. Rigidity in the beta cap domain creates a constraint for how the rest of the protein folds, and this results in backtracking in loop 1 in the cluster binding domain. We demonstrate that we can modulate the behavior of the cluster binding domain by making changes at the top of the beta cap domain. By destabilizing a set of contacts in loop 2 at the top of the beta cap domain, we can relieve backtracking in the cluster binding domain. Additionally, this drops the barrier to folding, suggesting that this structural feature introduces frustration in folding.

RESULTS

Structure of MitoNEET and Nomenclature

MitoNEET is a homodimeric Fe-S protein with a novel fold (26-28). Each protomer consists of three β-strands (β1, β2, and β3), an alpha helix (α1), and four loops (L1, L2, L3, and L4). Moving from the N to C terminus they are ordered L1, β1, L2, β2, L3, α1, L4, β3. Together the two protomers intertwine to form two domains, a beta cap domain and a cluster binding domain (Figure 3.1). The cluster binding domain coordinates two redox active 2Fe–2S cluster (20, 34, 35). Each of the two cluster cradles is formed from loop 3 and helix 1 of a single protomer. The coordination is unusual in that each of the clusters are coordinated by three cysteines and one histidine. The beta cap domain consists of two beta sheets. Each beta sheet is composed of β2 and β3 from one protomer and β1′ from the other protomer, arranged in the order β1′ β3 β2. This strand swap of β1′ intertwines the two protomers and creates a helical turn across the top of the beta cap domain in L2 (Figure 3.1). In an effort to understand the functional

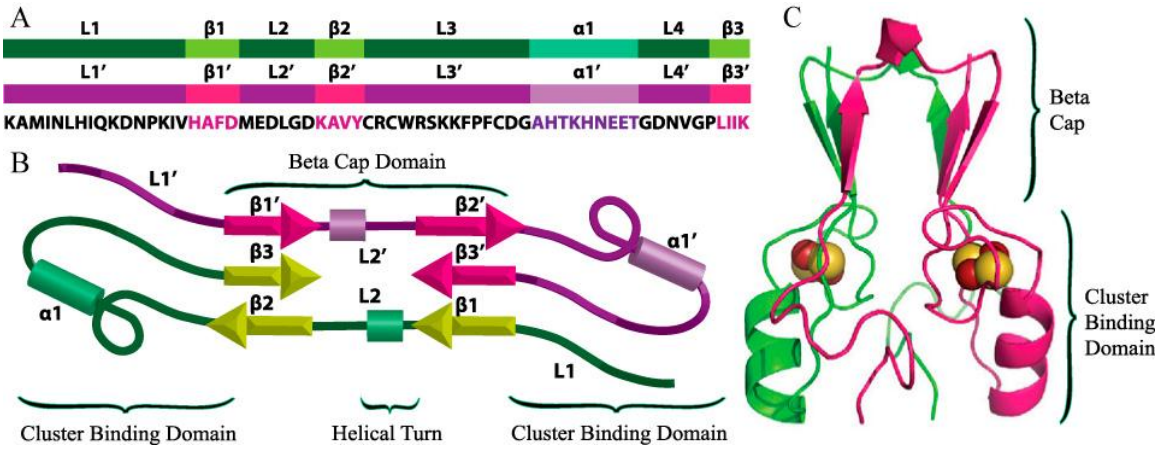

Figure 3.1: Structural organization and domain topology of mitoNEET
(A) Protomers aligned with the sequence. Highlighted regions correspond to the loops, helices, and beta strands in mitoNEET. (B) Splay diagram showing the location of the helical turn in the beta cap domain. (C) Ribbon diagram of mitoNEET colored by chain. The beta cap domain consists of a beta sandwich composed of two three-stranded beta sheets. Each beta sheet is composed of two strands from one protomer, and a third strand from a second protomer.

implications of the strand swap we now explore the mechanism of assembly of this protein.

Mechanism of Assembly—Coupled Folding and Dimerization.

SBMs have proven to be informative in the mechanism of assembly of a number of proteins (45, 68, 69). For example, we explain the rop dimer switch between syn and anti structures as a dual basin landscape that corresponds to distinct but related structures (70, 71). The unusual strand swap in mitoNEET creates a large interface surface between the two protomers. This led us to ask: Does the strand swap introduce constraints into the mechanism of assembly? In order to determine the mechanism of assembly of mitoNEET, we performed molecular dynamics simulations using a SBM.

The free energy surface of assembly for the mitoNEET dimer is projected onto three reaction coordinates; two corresponding to the folding of each protomer, and the third corresponding to protomer association (Figure 3.2A). Q_A represents the number of contacts formed in monomer *A*, $Q_A^{'}$ represents contact formation in monomer $A^{'}$, and Q_I represents the number of interface contacts formed. Moving along the reaction coordinate Q_A (or $Q_A^{'}$) we see that when Q_I is low, only a subset of Q_A or $Q_A^{'}$ contacts can be formed. It is only when a significant number of interface contacts are formed that an increase in the number of monomer contacts can be seen. The dimerization transition is broad, suggesting that there may be more than one population of structures in the transition state. It may be that multiple pathways to dimerization are present but are unable to be resolved using this particular set of reaction coordinates. This will be investigated in the next section. Analysis of the free energy barrier to folding (Figure 3.2B) indicates that folding and dimerization are highly cooperative with no obvious populated intermediates.

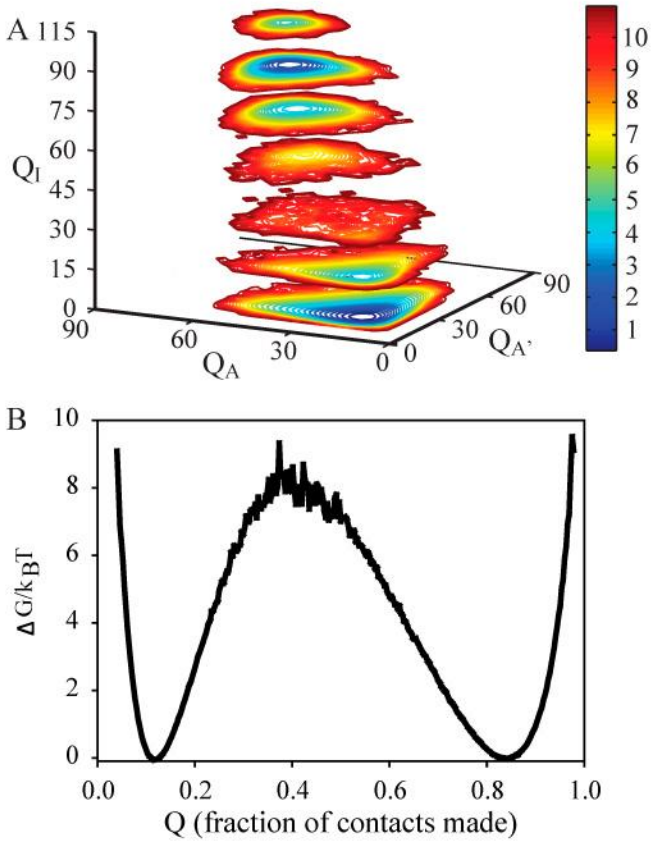

Figure 3.2: Free energy surface for the assembly of mitoNEET
(A) The free energy is projected on the reaction coordinates $Q_{\text{protomer A}}$, $Q_{\text{protomer A'}}$, and $Q_{\text{Interface}}$. A contour of the free energy as a function of Q_A and $Q_{A'}$ is plotted at eight different values of Q_I. (B) Free energy is plotted as a function of Q at T_f

Taken together this initial analysis suggests that the folding and assembly of mitoNEET is coupled and dimerization is obligatory. That is, the formation of individual protomers A and A' is only allowed upon binding. In addition, the 8 kt barrier is high for a protein of this size and suggests that traps along folding routes exist.

Clustering Uncovers The Folding Route.

The presence of traps during folding and assembly was investigated by performing a detailed analysis of the transition state ensemble. The probability of specific native contact formation in the transition state ($Q = 0.4$) is shown Figure 3.3A. This ensemble average view suggests that the transition state is diffuse and unstructured. However, examination of individual trajectories between the folded and unfolded basins indicates that transitions appear to choose between one of two pathways, and the averaging over these two pathways is responsible for the appearance of an unstructured transition state.

An accurate description of the folding process requires efficient separation of alternative routes. To this end, we implement a method for clustering structures in the transition state by their associated contact maps (described in detail in *Methods*). The resulting similarity network for $Q = 0.4$ is shown in Figure 3.3B. Two distinct clusters are observed, suggesting that two different ensembles are populated in the transition state. The probability of native contact formation for these two clusters at $Q = 0.4$ are provided in Figure 3.3 C and D. In contrast to the initial contact map generated for the ensemble where the two alpha helices appeared to form together and all beta sheets appeared relatively unstructured, clustering reveals contact formation between β2′ and β3′, β1 and β3′, as well as contacts within helix α1′. In the second cluster, the same contacts are

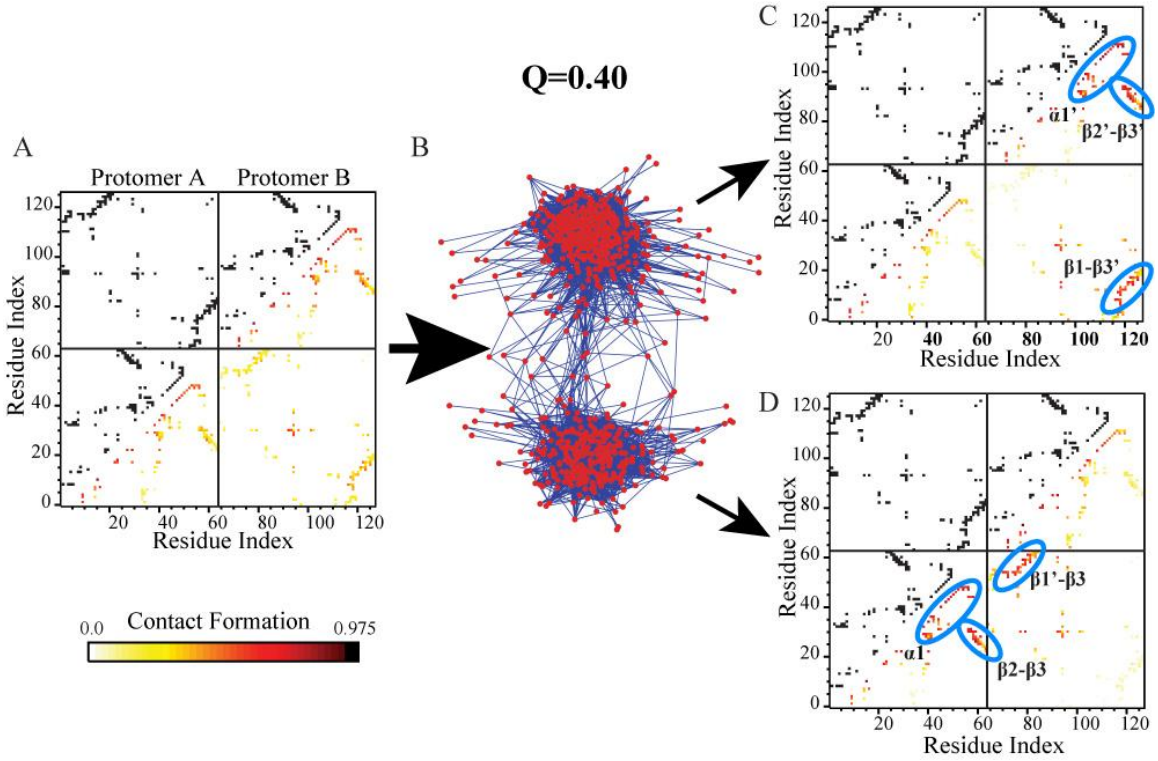

Figure 3.3: Cluster analysis of mitoNEET at Q = 0.4
(A) Probability of specific contact formation at Q = 0.4, represented with color on a contact map. White and yellow represent a lower probability of contact formation and orange and dark read represent a higher probability of contact formation. All native contacts are plotted in black across the diagonal for reference. (B) Similarity network for Q = 0.4. Red nodes represent contact maps for individual snapshots in completed transitions at Q = 0.4. Blue edges represent similarity between nodes. Shorter edges represent higher similarity between snapshots. (C) Probability of specific contact formation for each cluster. Structural features are circled in blue. In the top cluster, α1', β2'-β3', and β1-β3' form. (D) In the bottom cluster, α1, β2-β3, and β1'-β3 form.

formed but in the complementary protomer (β2 and β3, β1′ and β3, and α1). In fact, at $Q = 0.4$ only a single dominant route is present; however, the symmetry of the multimeric system creates a duplicate pathway distinguishable only through naming of the individual protomers. A structural representation of the fraction of native contact formation by residue at $Q = 0.4$ is plotted in Figure 3.4. After clustering the transition state is composed of one sheet from the beta sandwich and one half of the cluster binding domain. Half of the protein forms by the transition state and structural elements from both protomers are involved in the nucleation step. At $Q = 0.6$ (Figure 3.5) a third cluster appears that represents an alternate folding route. Here residue contact formation between the two protomers is symmetric. L1 and L1′ within the cluster binding domain are disordered, while the entire beta cap domain is well formed (β2 and β3, β1′ and β3, β2′ and β3′, β1 and β3′, L2, and L2′).

Rigidity in the Beta Cap Domain Forces Backtracking in Distal Sites.

The formation of native contacts within the protein is not uniform across the molecule. We therefore partitioned and analyzed the behavior of specific subsets of contacts (Figure 3.6A) within the protein as it folds to more carefully examine the order of events. We then plot the progression of these subsets of contacts (Q_{part}) as a function of the total number of contacts in the native fold (Q_{total}) (Figure 3.6 B and C). We observed that formation of contacts in a helical turn in L2, L1, and the late-forming beta sheet (β2 and β3, β1′ and β3) are interdependent. That is, contacts found in the helical turn (plotted in red) form early in the folding process, while contacts found in L1 (plotted in blue) form in a nonmonotonic manner, and those within the beta sheet (plotted in green) begin to populate after contact formation in L1 peak. These latter contacts (blue) exhibit a

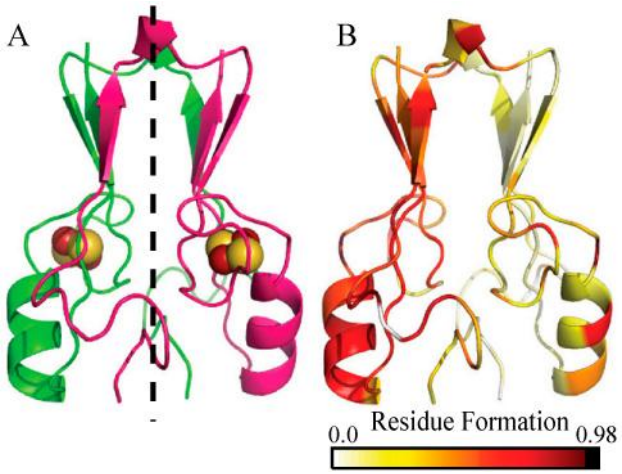

Figure 3.4: Transition state structure of mitoNEET
(*A*) MitoNEET colored by chain for reference. (*B*) Residue contact formation for mitoNEET at $Q = 0.4$. $P(Qi,Qca)$ is the probability that the set of contacts involving residue *i*, *Qi*, are formed at a given *Qca*. $P(Qi,Qca)$ was calculated for each cluster and represented with a color scale on the structure of mitoNEET. White and yellow residues represent decreased residue formation, and red and black represent higher residue formation. In the transition state half the beta sandwich forms (β1, β3′, and β2′) and one half of the cluster binding domain are formed.

behavior called backtracking, a signal of frustration in folding. Backtracking is the forming, breaking, and reforming of native contacts as the protein proceeds to the folded state. The formation of contacts in the late forming beta sheet appears to be coupled to backtracking. As L1 contacts begin to break, the formation of these contacts begins, and then plateaus just as contacts in L1 begin to form again. In the similarity network at $Q = 0.6$ (Figure 3.5) we see a third cluster appear that represents an alternate folding route present in 29% of complete folding/unfolding transitions. This detour in folding is the backtracking route. The second beta sheet comes together so that both beta sheets in the beta cap domain and the helical turns are well formed; however, in this process L1 twists away from the helix and cluster-binding region, breaking contacts that had formed earlier in folding.

To test if early formation of the helical part of L2 creates a constraint for how the second half of the protein folds, we performed a second set of simulations in which we destabilized this structural feature by removing three contacts in the helical turn and relaxing five dihedral angles. Relaxing this helical turn reduces the barrier to folding (Figure 3.6D). We looked for changes in the folding mechanism by again plotting Q_{part} vs Q_{total} for the same groups of contacts in this new set of simulations. Formation of L1 plateaus at $Q = 0.45$, at which point beta sheet 2 begins to form. However, the backtracking observed in WT mitoNEET is abolished. Taken together, these data indicate that stabilization of the helical turn (beta cap domain) impedes final formation of the native homodimer.

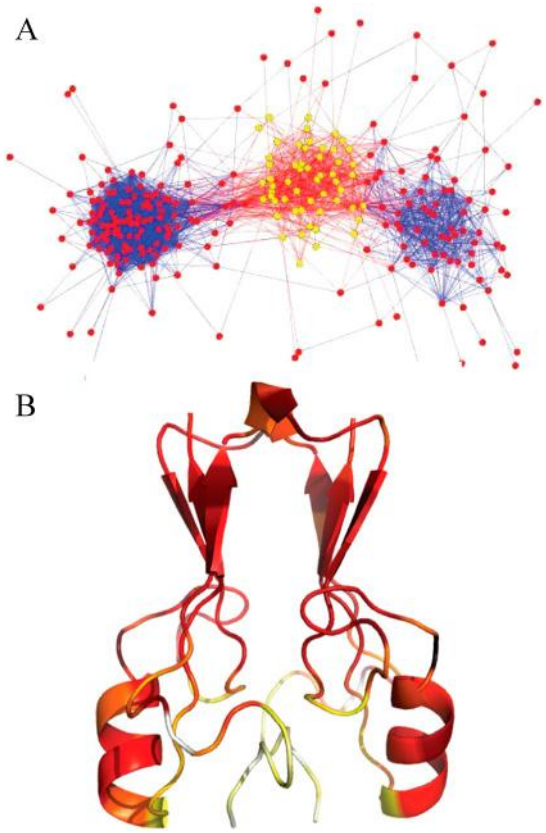

Figure 3.5: Cluster analysis reveals late detour in folding
(*A*) Similarity network at $Q = 0.6$. The presence of a third cluster highlighted in yellow represents a detour in folding. (*B*) Residue contact formation for middle cluster at $Q = 0.6$.

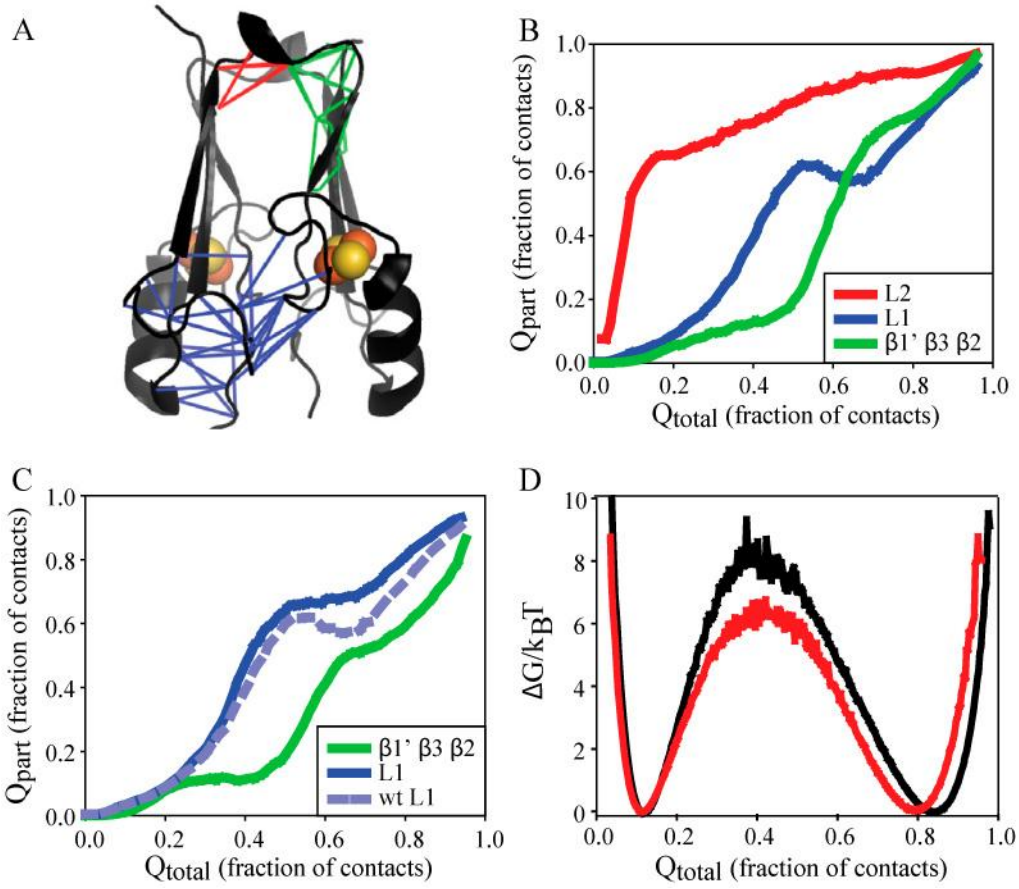

Figure 3.6: Dynamic coupling between the beta cap domain and cluster binding domain.

(*A*) Contacts to be examined are drawn on to the structure and grouped by color. Red lines represent contacts within and near the helix in L2, green lines represent contacts in the beta sheet (β1, β3′, β2′), and blue lines represent contacts between L1 and the cluster binding domain. (*B*) The average fraction of these subsets of contacts formed as a function of Q_{total}. Contacts in and near the helix in the turn form early. Contacts between L1 and the cluster binding domain begin to form but backtrack. Contacts in L2 and the beta sheet are dynamically coupled to contacts in L1 and begin to form as L1 begins to reform contacts after backtracking. (*C*) The average fraction of groups of contacts formed as a function of *Q* after destabilization of contacts in L2. Contacts between L1 and the cluster binding domain begin to form but backtrack. For reference, contacts between L1 and the cluster binding domain are plotted before destabilization as a light blue dotted line. Contacts in the beta sheet are still coupled to contacts in L1 and begin to form as contacts in L1 plateau. (*D*) Overlay of the free energy barriers for mitoNEET (in black) and mitoNEET with contacts (shown in red in *A*) and dihedrals relaxed (in red).

DISCUSSION

Cluster Analysis of the Transition State.

Often the approach to analysis of the transition state in molecular dynamics simulations is examination of the average transition state ensemble. Although useful for comparing to experiments, this approach can obscure the presence of multiple routes and the actual order of folding events. This is not only problematic for proteins that can access multiple pathways but also for symmetric multimeric systems. In addition, the amount of data generated for adequate sampling of transition states and intermediates during folding requires an efficient method for categorizing the multiple states sampled over hundreds of trajectories. To tackle this problem we implemented a method for clustering snapshots of populated states by scoring the similarity of associated contact maps. This method results in an efficient and robust way of visualizing and understanding complex folding landscapes. In our current study clustering revealed heterogeneity in the transition state and population of multiple routes in the folding of mitoNEET. This complexity in folding is obscured by the standard analysis. Thus, this methodology is not only superior for folding studies but also will be highly useful in characterizing local cracking and functional dynamics associated with regulation of activity in biomolecules.

Synchronization in Folding, Backtracking, and Frustration.

Despite being a symmetric homodimer, folding is asymmetric with each protomer contributing to early formation of one structural half. That is, when viewing the protein with a plane bisecting the beta sandwich, this plane crosses a loop (L2) with a helical turn that creates a swapped loop (L1) and strand (β1) between the two beta sheets. This swap tethers together the two protomers within each structural half, coupling folding and

dimerization. An unswapped structure would necessarily have more intra vs inter beta strand contacts within each sheet, allowing for more independent folding of each monomer. The swap forces contacts between residues that are on different protomers, therefore not close in sequence, to be part of the transition state for folding. This leads to increased cooperativity. It may be that the geometry and short length of L2 helps to enforce this strand swap by preventing the association of strand β1 with β3, and β1′ with β3′, thus preventing the monomers from forming independently.

Immediately following the transition state, as the downhill process initiates, rearrangement of the remaining structural features occurs. Sometimes these pieces pack in the wrong order and require backtracking as they search for the final structure. Interestingly, we observe that the swapped loops are involved in trapping and repacking during backtracking. That is, early formation of contacts in L1 and L2 result in a trap in folding where it is geometrically difficult for the remainder of contacts to form. In order for the second half of the beta cap domain to finish folding, the cluster binding domain must partially open up, and the swapped loop must break contacts with the cluster-binding domain. Weakening contacts in L2 relaxes some of the tightness of the beta cap domain. We see in these simulations that the barrier to folding drops and backtracking in the cluster-binding domain is relieved. This suggests that L2 is responsible for frustration in folding and that changes there can translate to the cluster-binding domain. The rop in barrier height is not a result of a decrease in backtracking, both are a result of decreased trapping. Thus, the swapped strands and associated loops contribute to both the nucleation process (transition state) and subsequent backtracking during final packing.

Allosteric Coupling of the Beta Cap Domain and Cluster-Binding Domain.

While evolution selects for robust folders, it must do this while conserving and selecting for function. This competition between selection for efficient folders and function can introduce frustration and prevent a landscape from becoming perfectly funneled (72, 73). It's been observed that functional regions of many proteins do not aid in folding and may in fact interfere with it. For example, functional loop mutations in WW domain proteins speed up folding at the expense of function, and in some cases they remove the barrier to folding completely (53, 54). It is possible, then, that frustration in the folding landscape can give us important clues about which structural features are important for function. Previous work done with the beta trefoil family of proteins demonstrated that a functionally important beta bulge was involved in backtracking and responsible for the slow folding of the IL-1β family of proteins (47, 50, 74). It is possible that in mitoNEET evolution has kept this frustration in L2 because this structural feature is functionally important. We observe that rigidity in the beta cap domain forces backtracking of the swapped loop in the cluster binding domain, demonstrating that there is communication between the two domains. The beta cap domain could function as an allosteric control site, modulating cluster insertion, assembly, or electron transfer. It would be interesting to determine experimentally if this element of geometric frustration in L2 is linked to functional regulatory properties of mitoNEET.

CONCLUSION

We used a SBM to characterize the folding landscape of mitoNEET. The folding mechanism uncovered in this landscape reveals communication between distal regions of the protein. We see that because of the strand swap, folding and assembly is cooperative

and dimerization precedes final folding. The individual monomers cannot independently fold then associate. We partition alternate routes in folding by clustering transition state contact maps. This efficiently separates out duplicate pathways caused by the symmetry of the dimeric system, and it reveals an alternate route late in folding. A careful examination of the folding mechanism reveals a region of frustration in loop 2 at the top of the beta cap domain, and that folding of the beta cap domain is dynamically coupled to the cluster binding domain. Tightness in the beta cap domain creates a constraint for how the rest of the protein folds, which results in backtracking in L1 in the cluster-binding domain. We demonstrate that by destabilizing a set of contacts in loop 2 at the top of the beta cap domain, we can relieve backtracking in the cluster-binding domain and drop the barrier to folding by a few kT.

SPECIAL METHODS

The contact map gives all possible interactions between a given residue and the other residues in a given structure. Contacts are identified using the Contacts of Structural Units software package (CSU) (58) on the crystal structure of mitoNEET stored in PDB ID 2QH7 (26). Because of slight asymmetries in the crystal structure, 14 contacts were generated that were not symmetrical between the two monomers. To simplify the model we remove these 14 unsymmetrical contacts. The resulting contact map consisted of 98 intramonomer contacts for each of the two monomers and 132 interface contacts for a total of 324 contacts.

A coarse grained SBM is used to represent the protein as described previously (42). In this model, each residue is represented by its Ca atom, and only interactions present in the contact map (between residues in the native state) are considered.

Nonnative interactions are not considered, and so energetic frustration is not included in this model. In our coarse grained folding studies the iron–sulfur cluster is implicitly included in the contacts present between cluster binding residues. Simulations were performed using Version 3.3.3 of the GROMACS software package (56). The integrator used was stochastic dynamics. The Berendsen algorithm was used with the coupling constant of 2. The time step τ was 0.0005. Each monomer was temperature coupled separately. Simulations were performed at folding temperature. Half of the simulations were started from the folded dimeric conformation, and half were started from unfolded and unbound monomers. A harmonic potential with an offset of 17 Å was applied to the center of mass of each monomer to hold the two monomers together.

We use the fraction of native contacts formed in a given snapshot of the protein as the reaction coordinate. ($Q_{C\alpha}$ is the fraction of natively interacting residue pairs whose ca atoms are within 1.2 times their native distance.) A contact is formed between $C\alpha$ atoms I and j if $rij < 1.2r^0_{ij}$ where r^0_{ij} is the pair distance in the native state. Q_A represents the contacts formed in protomer A, Q_A' represents contacts formed in protomer A', and Q_I represents interface contacts between the two monomers. Q_{part} is calculated as the fraction of contacts formed within a specific subset of contacts at a given value of Q_{total}.

Cluster analysis was performed as follows. All snapshots in the transition state at $Q = 0.4$ were represented with a contact map. Each representative contact map was internally compared using the logic gate xnor. If two structures had the same native contact formed or the same native contact not formed, a point was added to the similarity score between the two structures. If one structure had a contact formed that the other did not, this was considered dissimilar and no point was added to the score. Higher scores

indicate higher similarity between contact maps, and therefore higher similarity between snapshots of the transition state. The highest possible score was 324, one point for each possible contact in the native state. A threshold value of 200 was set, meaning that in order for two structures to be considered similar, they must have 200 out of 324 native contacts in common. Similarity networks were created for different values of *Q* near the transition state and were visualized using the edge weighted spring embedded layout algorithm in Cytoscape (75, 76).

ACKNOWLEDGEMENTS

E.L.B. thanks Jeffery Noel for helpful discussions. This work was supported in part by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (NSF; Grant PHY-0822283) and NSF Grant MCB-1051438, and National Institutes of Health Grant GM-54038. E.L.B. was also supported by a San Diego Fellowship.

Chapter 3, in part, is a reprint of the material as it appears in Proceedings of the National Academy of Sciences of the United States of America (2011), Baxter EL, Jennings PA, Onuchic JN. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Strand Swapping Regulates the [2Fe-2S] cluster in MitoNEET

ABSTRACT

MitoNEET is a recently identified diabetes drug target that coordinates a transferable [2Fe-2S] cluster, and additionally contains an unusual strand swap. In this manuscript, we use a dual basin SBM to predict and characterize the folding and functionality of strand swapping in mitoNEET. We demonstrate that a strand unswapped conformation is kinetically accessible and that multiple levels of control are employed to regulate the conformational dynamics of the system. Environmental factors such as temperature can shift route preference toward the unswapped pathway. Additionally we see that a region recently identified as contributing to frustration in folding acts as a regulatory hinge loop that modulates conformational balance. Interestingly, strand unswapping transfers strain specifically to cluster-coordinating residues, opening the cluster-coordinating pocket. Strengthening contacts within the cluster-coordinating pocket opens a new pathway between the swapped and unswapped conformation that utilizes cracking to bypass the unfolded basin. These results suggest that local control within distinct regions affect motions important in regulating mitoNEET's [2Fe-2S] clusters.

INTRODUCTION

The mitoNEET family of [2Fe-2S] proteins are key regulators of health and disease (18-20). Targeting these proteins for drug design has gained considerable interest in the last few years. An important challenge in targeting these proteins is that they do not contain a classical binding pocket and their mechanisms of regulation are only now being uncovered. We recently reported that mitoNEET can act as a cluster transfer protein (33) and that long-range communication within the protein can regulate the cluster-binding domain (77). Uncovering the physical basis for this long-range communication is an important first step toward understanding the structure-function relationships within this protein family.

MitoNEET is composed of two protomers (shown in purple and green, Figure 4.1) which intertwine via a swapped strand (β1) and loop (L1) to form a β-cap domain and a cluster-binding domain (26-28, 62) (Figure 4.1). Loop 2 (L2) leads in to the swapped strand and tethers the two structural halves of the protein together. We reported that the structural elements involved in strand swapping contribute to geometric frustration in folding and mediate communication between the β-cap and cluster-binding domain (77). Specifically, early formation of contacts in L2 slows folding. This behavior led us to ask, could L2 regulate strand swapping, and is the frustration we observe a consequence of strand swapping in dimeric mitoNEET?

In this system, geometric constraints dominate the landscape over energetic effects, and SBMs capture the overall characteristics of the landscape (77). Multiple basin SBMs are powerful tools that simulate energy landscapes in systems that undergo conformational changes or multimeric association reactions. These models revealed

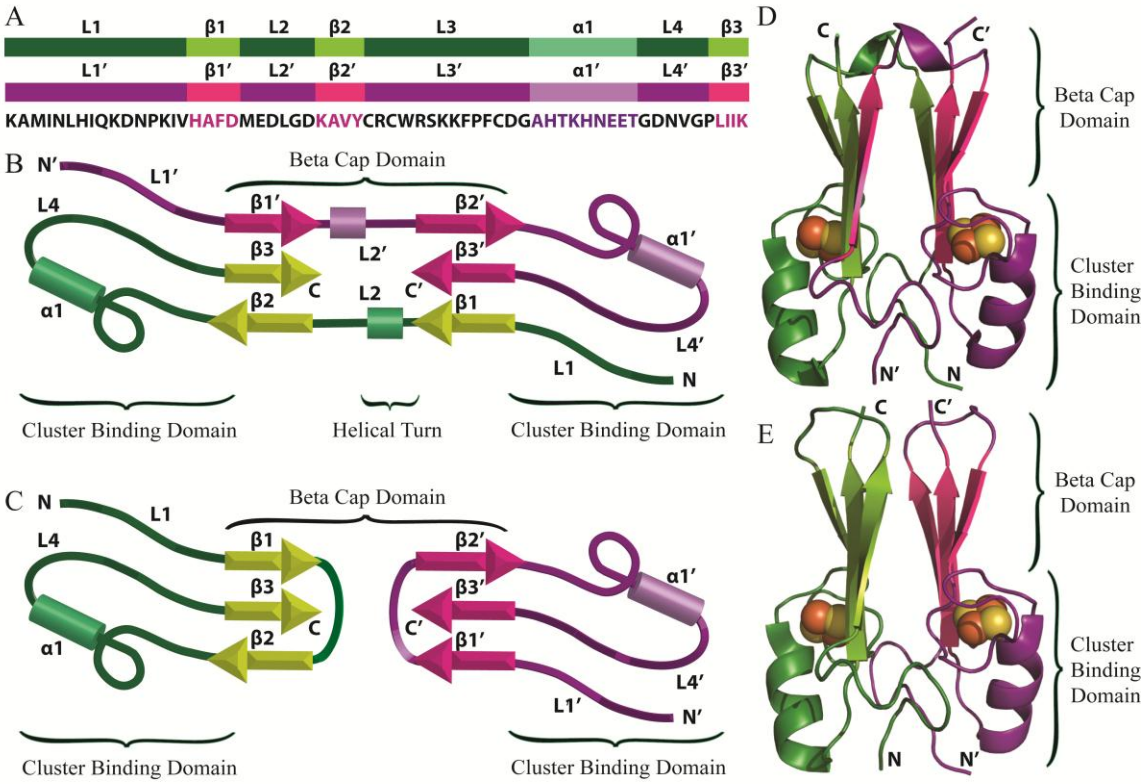

Figure 4.1: Structural organization and domain topology of mitoNEET
Protomers are individually colored in purple and green for clarity. (*A*) Protomers aligned with the sequence. Highlighted regions correspond to the loops, helices, and β-strands in mitoNEET. (*B*) Splay diagram of native mitoNEET. (*C*) Splay diagram of simulated unswapped mitoNEET showing the swapping of strand 1 and loop 1 mediated by the rearrangement of L2. Ribbon diagram of native mitoNEET (*D*) and simulated unswapped mitoNEET (*E*).

the underlying mechanisms of conformational transitions (49, 78), protein aggregation (79), and domain swapping (80), and additionally have been shown to successfully predict domain swapped structures (46). Recently, we used a dual funneled landscape to explain the unusual kinetic behavior of the rop-dimer system resulting from two distinct and competing conformations (71, 81). Drawing from this success, we undertook a study of the exciting new diabetes drug target mitoNEET that is known to transfer its cluster under oxidative stress conditions. In this study, we test the hypothesis that cluster-binding properties affect quaternary structure via long-range communication. One advantage of dual basin SBMs is that they allow us to construct and examine the behavior of conformations that are functionally important but may not be fully amenable to experimental structural characterization.

We used a dual basin SBM to investigate the role of strand swapping in mitoNEET. We demonstrate that although the unswapped conformation is kinetically accessible, it is thermodynamically less stable. The loop responsible for frustration in folding plays a significant role in thermodynamically destabilizing the unswapped structure relative to the native swapped structure, thus regulating strand swapping. Additionally, we see that the unswapped conformation transfers strain to the cluster-binding pocket that results in opening of the cluster-binding pocket and possibly facilitates cluster release. Tightening the contacts between cluster-coordinating residues results in switching from the unswapped conformation to the native swapped structure. These results suggest that not only is a strand unswapped structure accessible, but that it may play an important functional role in cluster transfer and release.

RESULTS

MitoNEET is a homodimer in which the two protomers intertwine to form a β-cap domain and a cluster-binding domain. Each protomer (colored in green and purple for clarity) contains three β-strands, an alpha helix, and four loops (Figure 4.1). The cluster-binding domain is composed of L1, L3, α1, and L4, and contains two [2Fe-2S] clusters, each with a 3Cys-1His coordination. The β-cap domain consists of two β-sheets sandwiched together. Each β-sheet is composed of three β-strands; two β-strands from one protomer (β2 and β3), and the third β-strand from the second protomer (β1′) leading to a strand swapped configuration. L2 follows the swapped strand (β1′) and loop (L1′) in sequence, tethering the two β-sheets into the β-sandwich, and contributes to frustration in the folding of this protein. This topology raises the question of whether L2 plays a regulatory role in strand swapping, and could the frustration we observe be a result of this regulation?

To address these questions, we created a dual funneled landscape comprised of both the native swapped structure and constructed unswapped structure using a SBM. In our model, we introduced additional contacts into strand 1 and L1 to allow for an alternate strand unswapped structure. Specifically, residues in strand 1 and L1 that make intraprotomer contacts are now allowed to make the corresponding interprotomer contacts, and residues that make interprotomer contacts are allowed to make intraprotomer contacts. In the native state of mitoNEET, strand 1 and L1 are swapped between the two structural halves so that they make interprotomer contacts with strand 3′ and loop 4′ (Figure 4.1 B and D). The additional contacts allow strand 1 to make intraprotomer contacts with strand 3 and L1 to make intraprotomer contacts with loop 4

(Figure 4.1 C and E), allowing each β-sheet to form entirely from one protomer, in the order β1, β3, β2. β-Strand 1 is no longer required to swap between the two structural halves, and L2 has the option of wrapping underneath a β-sheet instead of stretching between the two β-sheets. In addition, we created a single basin model in which only the unswapped conformation is accessible in order to better examine the folding of this protein.

Mechanism of Assembly

It is informative to first compare the folding of the strand unswapped conformation with that of the native, which we recently reported (77). Results from simulations in which only the strand unswapped conformation is accessible is plotted as the free energy as a function of Q_{total} in Figure 4.2A. At the folding temperature of the native swapped dimer ($T_{f(N)}$), the barrier for unswapped conformation is higher than that for the native swapped conformation, and the folded basin is destabilized. At the folding temperature of the unswapped conformation ($T_{f(U)}$), the barrier to folding is significantly lower than that observed for the native structure at $T_{f(N)}$ and peaks at a higher Q_{total}. In Figure 4.2B, a free energy surface of assembly for the unswapped conformation is presented. Q_A represents contact formation in protomer A, Q_A' represents contact formation in protomer A', and Q_I represents protomer–protomer interface contact formation. At $Q_I = 0.4$ interface contact formation, two basins are visible, one corresponding to high Q_A and low Q_A', and the second corresponding to low Q_A and high Q_A'. This signature indicates that one protomer may begin folding independently of the other, and the formation of interface contacts between the two protomers helps nucleate the folding of the second protomer. It is only when a number of interface contacts are

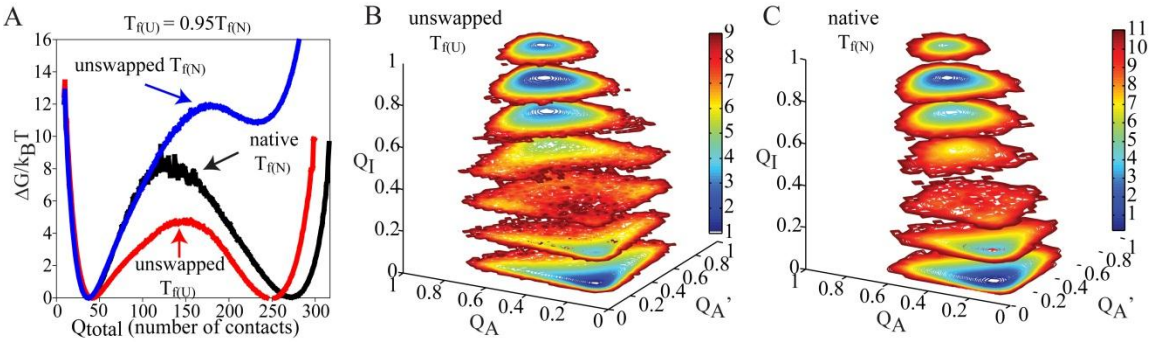

Figure 4.2: Single basin modeling

(*A*) Free energy is plotted as a function of Q at $T_{f(N)}$ for native mitoNEET (black) and unswapped mitoNEET at $T_{f(N)}$ (blue) and at $T_{f(U)}$, or 0.95 $T_{f(N)}$ (red). The unswapped structure is kinetically accessible, but thermodynamically unstable at the folding temperature of the native protein. Free energy surface for the assembly of unswapped mitoNEET at $T_{f(U)}$ (*B*) and native mitoNEET at $T_{f(N)}$ (*C*). The free energy is projected on to the reaction coordinates Q protomer A, Q protomer A', and Q interface. A contour of the free energy as a function of Q_A and $Q_{A'}$ is plotted at seven different values of Q_I.

formed that the folding of the second protomer begins. This folding mechanism differs from that of the native swapped dimer, where dimerization is coupled to protomer formation and no folding of individual protomers is observed. (Figure 4.2C).

Dual Basin Simulations

Figure 4.3A shows the free energy surface of the dual basin simulations at $T_{f(N)}$ as a function of two coordinates, Q_N and Q_U. Q_N represents the total number of contacts formed in the native configuration, and Q_U represents the total number of contacts formed in the unswapped conformation. Two major populations are observed. The basin at low Q_U and Q_N represents the unfolded ensemble, and the basin at intermediate Q_U and high Q_N corresponds to the native swapped configuration. At high Q_U and intermediate Q_N, a small population is present that represents the unstable unswapped configuration. In an effort to explore the formation of the unswapped configuration, we performed simulations at 0.98 $T_{f(N)}$. The resulting free energy surface is plotted in Figure 4.3B. Under these conditions we enhance the population of species at high Q_U and low Q_N, corresponding to the unswapped configuration. Below 0.98 $T_{f(N)}$, it is difficult to get sufficient thermodynamic sampling, however as temperature decreases, the unswapped structure becomes more kinetically accessible. At 0.9 $T_{f(N)}$, 45% of trajectories starting from the unfolded state transition to the unswapped configuration. Taken together these results indicate that, although the unswapped structure is thermodynamically less stable, it is kinetically accessible.

Destabilization of Geometrically Frustrated L2 Increases the Population of the Unswapped Configuration.

Geometric frustration in L2 mediates communication between the β-cap and the

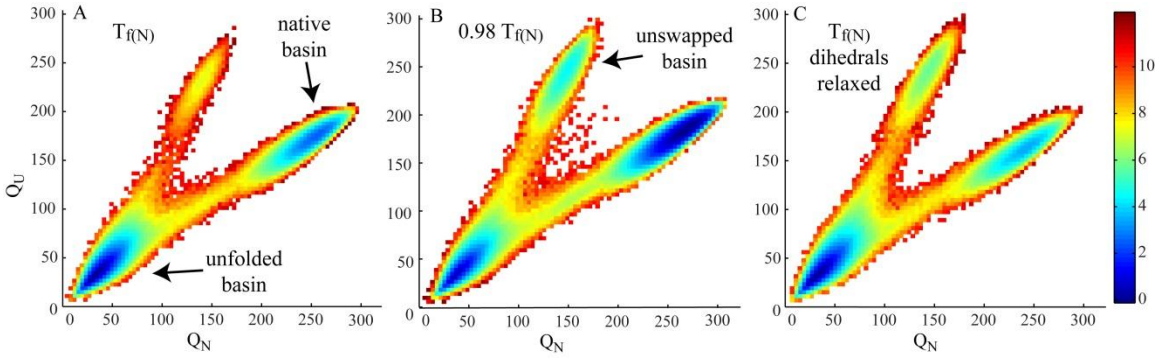

Figure 4.3: Dual basin modeling
Free energy profiles for dual basin simulations of mitoNEET. (*A*) Simulations performed at the folding temperature of native mitoNEET with all native contacts and dihedrals. Two basins are present corresponding to native mitoNEET and the unfolded state. The unswapped configuration is accessible, but is unstable at $T_{f(N)}$. (*B*) Analysis of folding of native and unswapped mitoNEET with a dual basin simulations performed at 0.98 $T_{f(N)}$ of native mitoNEET. A third basin is present corresponding to the unswapped configuration. (*C*) Dual basin simulations performed at $T_{f(N)}$ with relaxation of contacts and dihedrals in L2 allow the formation of the unswapped configuration.

cluster-binding domain in the native dimer (77). This frustration leads to communication with distal residues that backtrack, or locally unfold, and contribute to regulation of global motions. To examine the effect that frustrated L2 has on the landscape of the native and unswapped structures, we removed tertiary contacts and relaxed dihedrals in this region and then repeated dual basin simulations at $T_{f(N)}$. Our results indicate that the native configuration is still preferred, but relaxing L2 increases the accessibility and stability of the unswapped configuration (Figure 4.3C).

Strand Unswapping Activates an Alternate Site of Communication in the Cluster-Binding Region.

The protein must completely unfold to move between the native swapped and unswapped structures in the SBM where all contacts are uniform, therefore, it is essential to understand the mechanism of unfolding of the respective structures. To address this question, we examined the mechanism of unfolding for the unswapped (Figure 4.4 A–C) and native (Figure 4.4 D–F) conformation in single basin landscapes. As described previously (77), we analyzed the behavior of individual contacts within one protomer and then partitioned them based on the order in which they unfold. For clarity we show a splay diagram of one protomer colored by groups of contacts to be examined with the second protomer shown in gray (Figure 4.4 A and D) and a ribbon diagram colored by groups of contacts to be examined with a surface rendition of the second protomer in gray (Figure 4.4 B and E). Regions shown in black are not plotted and are given for reference. We then plot the fractional population of these subsets of contacts (Q_{part}) as a function of the total number of contacts in the protein during unfolding (Q_{total}) (Figure 4.4 C and F).

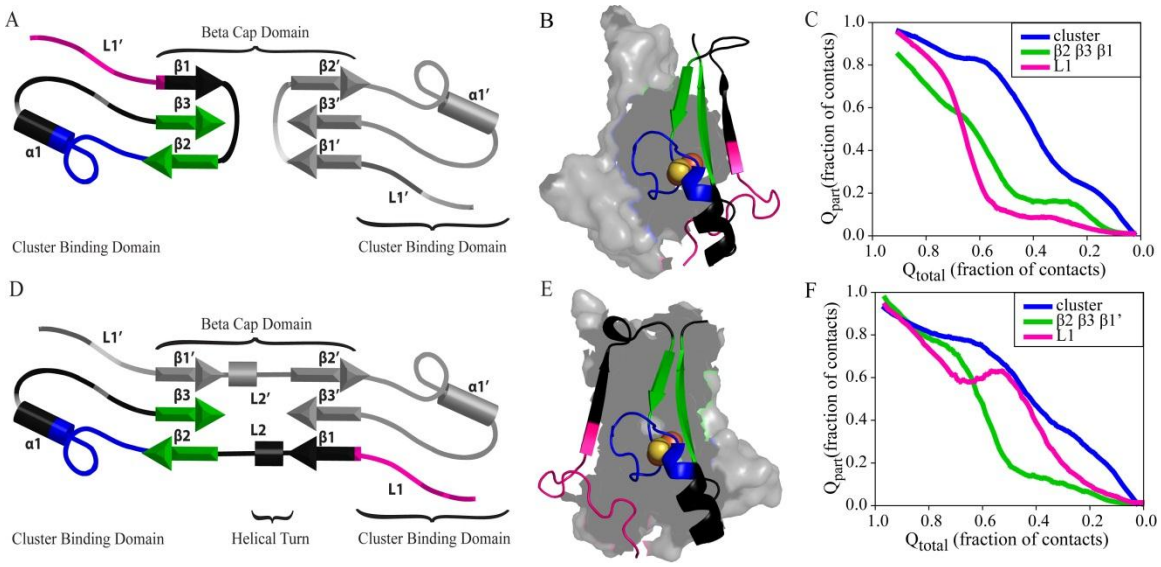

Figure 4.4: Single basin modeling: the order of unfolding for the unswapped and native configuration of mitoNEET

Order of unfolding for the unswapped and native configurations of mitoNEET within one protomer. The unswapped configuration of mitoNEET (*A* and *B*) and native configuration (*D* and *E*) are colored by groups of contacts examined. Pink represents contacts present in loop1, green represents contacts present in the β-sheet (β2, β3, β1, and β2, β3, β1′), blue represents contacts within the cluster-binding pocket and between the two pockets. Regions in black are not plotted and are shown for reference. The average fraction of these subsets of contacts formed as a function of Q_{total} for the unswapped configuration (*C*) and the native configuration (*F*). In both cases, the cluster-binding pocket is the last to unfold. In the native configuration, contacts in L1 exhibit backtracking. These contacts break early but begin to reform as the β-sheet breaks, and finish breaking late in unfolding. In the unswapped configuration, backtracking is lost. Contacts in L1 are the first to unfold followed by contacts in the β-sheet.

Contacts within the cluster-binding pocket and between the two cluster-binding pockets (plotted in blue) are the last to unfold in both models. As reported previously (77), we observe backtracking in L1 (plotted in pink) in the native conformation. That is, L1 begins to break early in unfolding, then reforms contacts, then subsequently breaks again later in unfolding. The backtracking in the cluster-binding domain is coupled to the unfolding of the β-sheet (plotted in green) in the β-cap domain, demonstrating communication between the two domains. In the unfolding of the unswapped conformation, backtracking is lost in L1. Instead, L1 appears to initiate unfolding followed closely by the unfolding of the β-sheet.

The loss of a backtracking signal in the unswapped conformation does not necessarily suggest that communication between domains has been lost. Although the formation of all individual contacts are possible within our model, we observe that strain from strand unswapping introduces competition between contact formation in different regions. Specifically, the short length of L2 prevents the simultaneous formation of all contacts between β1 and β3, and β2 and β3 in the unswapped configuration. In the case that all contacts are formed between β1 and β3, L2 pulls on β2, thus destabilizing some contacts between β2 and β3, resulting in destabilization of the β-sheet (plotted in green) in the unswapped simulations, which on average has only 85% contact formation in the folded basin. This strain appears to translate directly in to the cluster-binding pocket. The majority of contacts in the cluster-binding region are able to form completely in the folded state, however, in the unswapped configuration, the two contacts between the cluster-coordinating residues are only formed 80% of the time in the folded basin. These

results suggest that strand unswapping shifts the site of communication in the cluster-binding domain from L1 to the cluster-coordinating residues.

Stabilizing Cluster-Binding Contacts Triggers Switching.

In an effort to address the mechanism of interconversion between the unswapped and the native swapped conformation, we examined the time-dependent evolution of structure within each individual trajectory. Example trajectories plotted as *Q* as a function of time are shown in Figure 4.5. Contacts specific to the unswapped and native swapped structures are plotted in red and black, respectively. Contacts that are not involved in strand swapping and are common between the two structures are given in blue. Plotted in Figure 4.5A is an example trajectory in which mitoNEET starts in the unswapped conformation (red), completely unfolds, and then refolds into the native swapped conformation (black). The shared contacts (blue) that describe the global structure must completely unfold to move from the unswapped to the native swapped conformation. That is, only after complete unfolding of the unswapped conformation will the native strand swapped structure form. Therefore complete unfolding is required to move between the native and unswapped basins.

In our single basin simulations, we observed communication between the β-cap and the cluster-coordinating residues in the unswapped conformation. The majority of contacts in the cluster-binding region are able to form completely in the folded state, however, two contacts between the coordinating residues are only formed 80% of the time in the folded basin. These residues follow β2 in primary sequence and are positioned perpendicular to β2 in the tertiary structure. Therefore, optimal placement of β2 likely influences the orientation of the cluster-binding residues. Likewise, cluster coordination

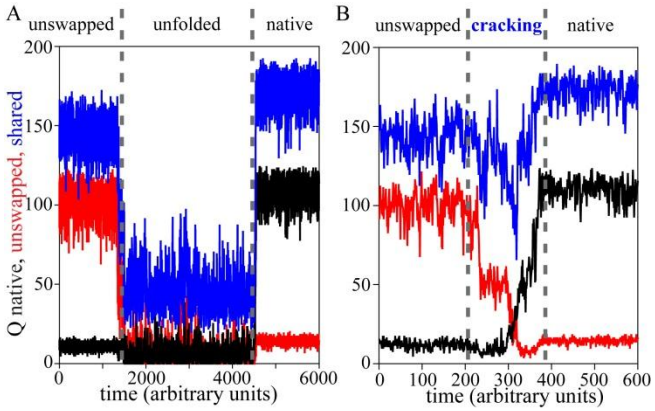

Figure 4.5: Tightening cluster contacts triggers switching by cracking
Sample trajectory of dual basin simulations at $T_{f(U)}$ (*A*). Contacts specific to the native configuration are plotted in black, and unswapped contacts are plotted in red. Contacts that define the global fold and are shared between the two configurations are plotted in blue. The unswapped conformation must completely unfold before the native conformation can form. (*B*) Sample trajectory of dual basin simulations at $T_{f(U)}$ with cluster-binding contacts tightened. MitoNEET is able to switch from the unswapped conformation to the native without completely unfolding. Cracking in shared contacts relieves strain and preserves the global fold.

geometry may influence packing of strands within the β-sheet. To test the role of these cluster-coordinating contacts in strand swapping, we introduce heterogeneity into our dual basin model by strengthening the cluster-coordinating contacts. At the folding temperature of the unswapped conformation $T_{f(U)}$, or 0.95 $T_{f(N)}$, we observe switching without fully unfolding between the native and unswapped structures. Specifically, 29% of trajectories that fold in to the unswapped configuration first are now able to switch into the native configuration without completely unfolding. An example of a switching event is shown in Figure 4.5B. Contacts shared between the two structures (plotted in blue) do not completely unfold, rather they exhibit a behavior called cracking, or local unfolding in response to local stress in conformational transitions (49, 82, 83). Here cracking preserves the global fold by relieving stress associated with the repositioning of β1 and L1. For comparison, low temperature simulations (0.9 $Tf_{(N)}$, 0.95 $T_{f(N)}$, 0.98 $T_{f(N)}$) were run without these contacts strengthened, and switching behavior is not observed at any temperature. Therefore, changes in stability of the [2Fe-2S] cluster may be an essential regulator of conformational dynamics within the protein.

DISCUSSION

Frustration in Folding a Necessary Consequence of Regulating Conformational Balance.

In a previous study, we identified L1 and L2 as interesting regions of frustration in mitoNEET. We reported that, for folding directly into the native structure of mitoNEET, a subset of contacts exhibit backtracking (77). That is, these contacts form early in folding, but to complete folding they break to relieve geometric frustration, and reform later. More importantly, we demonstrated that L2 in the β-cap domain controls the

backtracking behavior of L1 in the cluster-binding domain, thus mediating communication between the two domains. By relaxing L2, we were able to remove backtracking in L1 and decrease the barrier to folding. An important question is why has mitoNEET evolved this frustration in L2?

In our present study, we introduce the same relaxation of contacts and dihedrals in L2 that previously decreased the barrier to folding, however, we do this in a dual basin landscape. Relaxing this loop destabilizes the native configuration and allows the unswapped conformation to be significantly populated at $T_{f(N)}$ (Figure 4.3C). In contrast, when L2 is geometrically constrained, the unswapped configuration is significantly destabilized relative to the native conformation (Figure 4.3A), suggesting that L2 functions as a regulatory hinge loop for the strand swapping of β1 and L1, thus controlling the association of β1 with β3, and β1 with β3′. A more relaxed loop would result in a faster folding protein, but the additional geometric constraint imposed by L2 is necessary for regulating the balance between strand swapped conformers.

Although the structuring of L2 biases the folding of mitoNEET toward the native structure, the unswapped structure is not forbidden. At 0.98 $T_{f(N)}$, we see that the unswapped conformation becomes accessible, and in low-temperature kinetic runs we demonstrate that route preference can be significantly shifted toward the unswapped folding pathway, suggesting that the route preference and conformational balance are dynamic and sensitive to environmental control.

Tightening Cluster-Binding Pocket in Unswapped Conformation Opens Switching Pathway.

In contrast to the native folding pathway, which requires coupled folding and dimerization, the unswapped folding pathway allows protomers to form more independently of each other prior to dimerization (Figure 4.2 B and C). Because proteins are monomeric emerging from the ribosome, and unfolded proteins are more vulnerable to degradation, prefolding of the unswapped monomer may provide a degree of protection until a partner is produced. In our perfectly homogenous SBM, if the unswapped dimer is formed first, the protein must fully unfold before it can find the native swapped configuration (Figure 4.5A). However real proteins have heterogeneity in contact strength, and mitoNEET likely experiences dynamic heterogeneity near the [2Fe-2S] cluster. In vivo tightening in the cluster pocket may arise as a result of the formation of a disulfide bond between C72 and C83, protein or drug binding, or cluster assembly and insertion. Tightening cluster-binding contacts in our model opens a switching pathway previously unavailable. This route provides an additional level of conformational control by facilitating strand swapping without complete unfolding. As strand 1 and L1 switch from the unswapped position in to the native swapped position, the global fold exhibits cracking but remains intact (Figure 4.5B). An intriguing possibility is that the strain in L2 facilitates cluster assembly or insertion into the unswapped structure by pulling open the cluster-binding pocket via β-strand 2, thus increasing the accessibility of the cluster-coordinating residues. Switching may then provide a convenient mechanism for preserving the coordination of the [2Fe-2S] cluster while moving from the unswapped structure to the more thermodynamically stable native

structure (Figure 4.2A). Alternatively, the preferred pathway between the unswapped and the native conformation may be through complete unfolding, and switching may be a fail-safe mechanism activated in response to accidental cluster insertion into the unswapped conformation. In this case, switching would be a better alternative to complete unfolding which could release toxic free iron into the cell.

Evolution of Functional Control Through Geometric Frustration.

MitoNEET and Naf-1 are distinct drug targets (18, 23, 84), and their crystal structures reveal a dimeric and strand swapped topology in which each protomer contains one CDGSH domain that coordinates a [2Fe-2S] cluster (26-28, 30, 62). Recently, homologous structures from archaea and bacteria have been released that show significant permutations in the β-cap domain. These structures include a monomeric structure containing tandem CDGSH domains and two strands per β-sheet, a dimeric strand unswapped structure with four strands per β-sheet, and a dimeric strand swapped structure with five strands per β-sheet (29). These permutations raise the question of how and why dimerization and strand swapping evolved in this class of proteins. It has been proposed that frustration in folding is the result of conservation of functional residues (50, 53, 54, 74, 85). Local energetic frustration is present near the binding sites of many proteins (52), and the functional regions of some proteins have been shown to be responsible for decreased stability (72, 73, 86), folding traps (50, 74), long-lived partially folded intermediates (87), and misfolding (88). An intriguing idea is that frustration has been engineered into proteins throughout evolution. The NEET family of proteins may have evolved different strand swapped folds to maintain careful control of [2Fe-2S]

cluster release and conformational balance in the presence of changing environmental conditions.

CONCLUSION

We used a dual basin SBM to predict and characterize the folding and functionality of strand swapping in mitoNEET. Our results show that multiple levels of control regulate the conformational dynamics of this system. A loop which was previously identified as contributing frustration to folding acts as a regulatory loop destabilizing the unswapped conformation relative to the native conformation. Local tuning of contacts in this loop reverses this effect, increasing the accessibility of the unswapped conformation. Additionally, route preference is shifted toward the unswapped folding pathway by modulating temperature. We demonstrate that not only is a strand unswapped structure accessible, but also that strand unswapping preferentially destabilizes contacts opening the cluster-binding pocket. Tightening these local contacts opens an alternate route between the unswapped conformation and the native swapped conformation that bypasses the unfolded basin. This local control of global structure and dynamics offers a powerful mechanism to evolve multiple functional states with a single protein sequence.

SPECIAL METHODS

A coarse-grained SBM was used to model the protein as described previously (34), in which each residue is represented by its C^{α} atom. Our model consists of native interactions in addition to some nonnative interactions to model domain swapping as described below. These nonnative interactions introduce some additional frustration into the system not normally present in purely SBMs. In our coarse-grained studies, the iron-

sulfur cluster is implicitly included in the contacts present between cluster-binding residues. Simulations were performed using version 3.3.3 of the GROMACS software package (56). The integrator used was stochastic dynamics. The Berendsen algorithm was used with the coupling constant of two. The time step τ was 0.0005. Each protomer was temperature coupled separately. A harmonic potential with an offset of 20 Å was applied to the center of mass of each protomer to hold the two protomers together.

Native contacts were identified using the SBMs in GROMACS (SMOG) Web server (57) with a Contacts of Structural Unit contact map (58) and were assigned interaction distances based on the crystal structure of mitoNEET stored in PDB ID 2QH7 (26). Slight asymmetry in the crystal structure results in 14 unsymmetric contacts between the two protomers. To simplify our model, we removed these 14 contacts from the set of native contacts. To model a dual funneled landscape with domain swapping, additional contacts were introduced in strand 1 and L1 in both chains. For each intraprotomer contact present between residues *i* in β1 and L1 and *j* in chain A (i' in β1′ and L1′ and j' in chain B), the identical interprotomer contacts between residues *i* in β1 and L1 of chain A and j' in chain B were created using the original intraprotomer interaction distances. Likewise, for all interprotomer contacts present between residues *i* in β1 and L1 of chain A and j' in chain B, the corresponding intraprotomer contacts between residues *i* in β1 and L1 of chain A and *j* in chain A were added. The resulting contact map consisted of 92 interprotomer contacts and 34 intraprotomer contacts specific to strand unswapping, 92 interprotomer contacts and 34 intraprotomer contacts specific to native strand swapping, and 40 interprotomer contacts and 158 intraprotomer contacts common to both the native strand swapped and strand unswapped structure.

The symmetry in multimeric systems such as mitoNEET can create duplicate pathways distinguishable only through naming of the individual protomers, and averaging over these pathways can obscure the actual order of folding events. To address this problem, we used Cytoscape (75) to cluster transition state structures as described previously (77) and to examine the order of folding events within a cluster.

ACKNOWLEDGEMENTS

The authors thank Jeff Noel, Michael Jamros, and Mark Paddock for helpful discussions. This work was supported in part by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (NSF, Grant PHY-0822283) and by NSF MCB-1051438, and by National Institutes of Health Grant GM-54038. E.L.B. was also supported by a San Diego Fellowship.

Chapter 4, in part, is a reprint of the material as it appears in Proceedings of the National Academy of Sciences of the United States of America (2012), Baxter EL, Jennings PA, Onuchic JN. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Allosteric control in a metalloprotein dramatically alters function

ABSTRACT

Metalloproteins (MPs) comprise one-third of all known protein structures. This diverse set of proteins contain a plethora of unique inorganic moieties capable of performing chemistry that would otherwise be impossible using only the amino acids found in nature. Most of the well-studied MPs are generally viewed as being very rigid in structure, and it is widely thought that the properties of the metal centers is primarily determined by the small fraction of amino acids that make up the local environment. Here we examine both theoretically and experimentally whether distal regions can influence the metal center in the diabetes drug target mitoNEET. We demonstrate that a loop (L2) 20Å away from the metal center exerts allosteric control over the cluster binding domain, and regulates multiple properties of the metal center. Mutagenesis of L2 results in significant shifts in the redox potential of the [2Fe-2S] cluster, and orders of magnitude effects on the rate of [2Fe-2S] cluster transfer to an apo-acceptor protein. These surprising effects occur in the absence of any structural changes. An examination of the native basin dynamics of the protein using all-atom simulations shows that twisting in L2 controls scissoring in the cluster binding domain, and results in perturbations to one of the cluster coordinating histidines. These allosteric effects are in agreement with previous folding simulations that predicted L2 could communicate with residues surrounding the metal center. Our findings suggest that long-range dynamical changes in the protein backbone can have a significant effect on the functional properties of MPs.

INTRODUCTION

Metalloproteins (MPs) comprise nearly one third of all known protein structures and are classified by their unique inorganic moieties. Examples are the copper-containing cupredoxins, iron-sulfur proteins, nickel-containing hydrogenases, and numerous others (89-91). The incorporation of metals allows for thousands of new biological catalysts, capable of performing chemistry that would otherwise be impossible using only the amino acids found in nature. The widely-held view of MPs is that the properties of the metal centers are primarily dictated by the "inner sphere" which consists of the directly coordinating ligands and to a lesser extent local residues which can influence both the electrostatic environment as well as the hydrophobic residues which can act like wires and funnel electron density (92). The redox properties of MPs tend to be influenced by only a small fraction of the amino acids in the protein, generally those directly bonded to, or in close proximity to, the metal center (93). Indeed, it has been shown that the local scaffold of a metallic redox center can by chemically synthesized to mimic the active site of the protein and in many cases the electron transfer properties of these metalloproteins can be reproduced and better understood in these simplified ligand environments (92).

Iron-sulfur (FeS) cluster-containing proteins make up the largest class of MPs and are major players in human health and disease (94). They play critical roles as electron transfer proteins in processes such as photosynthesis, cellular respiration, nitrogen fixation (95, 96), and catalysis (89). The newest member of the FeS cluster protein family, mitoNEET, is a uniquely folded homodimeric [2Fe-2S] protein with each monomer bearing a single [2Fe-2S] cluster (26-28) (Figure 5.1A). The clusters are ligated in a rare 3Cys-1His coordination sphere (Figure 5.1B), where the single His is critical for

function (33, 62) and drug binding (97). MitoNEET is a target of the thiazolidinedione (TZD) class of anti-type II diabetes drugs and is the first known FeS protein to be targeted by drug binding (18). We discovered that mitoNEET functions as a cluster transfer protein and can donate its cluster to an apo-acceptor protein and into mitochondria under oxidative stress conditions (33). These stress conditions are commonly found in patients with diabetes (98, 99). TZD binding blocks cluster transfer in vitro and iron overload in vivo, a condition commonly associated with type II diabetes (33). An interesting hypothesis is that abrogation of mitochondrial iron overload by treatment of cells with small molecules such as TZDs is a result of changes in mitoNEET's cluster properties and is an active area of inquiry.

We recently used energy landscape theoretical studies to investigate the factors that govern cluster properties in mitoNEET (77, 100). Briefly, energy landscape theory indicates that proteins fold in a funneled fashion with minimal frustration, with the native state and functional fluctuations occurring towards the bottom of this funnel (39, 41). Because proteins are active on the same landscape that they fold on (55), their functional motions may introduce ruggedness in to the folding landscape. For example, functional loop mutations in WW domain proteins speed up folding at the expense of function, and in some cases remove the barrier to folding completely (53, 54). Additionally, energetic frustration in proteins co-localize with cofactor binding sites (52). Simulations with the beta trefoil family of proteins demonstrated that strain in a functionally important beta bulge was responsible for the slow folding of the family (47, 50, 101). Thus, identifying residues that contribute frustration in folding may be an effective way to predict and

identify important sites for protein-protein interactions as well as new binding regions for potential drug targets.

Whether energy landscape theory can predict functional control in metalloproteins is an open question. We used this theory together with SBMs (42, 102, 103) to investigate the landscape of mitoNEET and predicted a loop *distal* to the [2Fe-2S] cluster (L2) constrains folding and controls the motions of cluster binding domain. We hypothesized that, despite being ~20 Å removed from the [2Fe-2S] cluster, this frustrated loop region may function as an allosteric control site, regulating functional properties of the [2Fe-2S] moiety (77). In our current study we test this hypothesis by experimentally introducing perturbations into this distal loop (Figure 5.1A) and analyze the properties of the [2Fe-2S] cluster. Point mutations as well as insertion of new residues in L2 significantly reduce cluster stability, and can accelerate mitoNEET cluster transfer to apo-acceptor proteins by a factor of 15 fold. These perturbations in L2 also shift the redox potential by up to ~60mV indicating that long-range effects regulate cluster stability, cluster transfer, and electron transfer potential in this system. A common interpretation of these results would be that mutation alters the structure of the protein, especially the cluster-binding region. Strikingly, the crystal structures of the mutant proteins show no changes in overall fold or in the cluster-binding domain. The only difference observed is in a single mutant protein with the necessary elongation of the distal loop upon insertion mutagenesis. Taken together with native-basin simulations we suggest correlated motions between L2 and the protein scaffold provides distal allosteric control of the [2Fe-2S] cluster and is an interesting site for targeting with drug design.

RESULTS

Perturbations in the L2 Region Induce Long-range Changes in the Redox Potential of MitoNEET's [2Fe-2S] Cluster

Mutagenesis of several residues in the L2 region (Figure 5.1A, blue spheres) led to significant changes in the redox potential ($E_{M,7}$) of the protein's cluster (Figure 5.1C). We used optical spectroelectrochemical titrations (34) to measure redox potentials of various L2 mutants. We mutated Met62 and Asp67 to Gly in order to relax L2 and allow greater flexibility in the beta cap domain. These M62G and D67G mutants showed $E_{M,7}$ decreases of -12 mV and -14 mV, respectively, from the WT $E_{M,7}$ value of +26 ± 3 mV (34). Importantly, these shifts are ionic strength independent and suggest that factors other than electrostatics control the redox properties. Therefore, we introduced an aromatic group in the L2 region that would potentially stabilize L2 with increased hydrophobic packing between protomers. We found that this mutation induced an even greater shift in $E_{M,7}$ value which is 28 mV less than WT. We also found that reducing flexibility of residue 66 by replacing Gly with Ala (G66A) shifted the $E_{M,7}$ ~ -25 mV. The G66A/D67A double mutant was designed to increase the helicity of the L2 region and interestingly lead to the largest $E_{M,7}$ shift from the WT protein of ~44mV (-18 ± 3 mV for G66A/D67A compared to +26 ± 3 mV). Finally, opposite changes in $E_{M,7}$ were induced by the choice of residue inserted at position 68 in the amino acid sequence. Insertion of alanine (A68 Insert) led to a positive increase in the $E_{M,7}$ value to +43 ± 7 mV. Alanine is known to be a helix maker and as a result we wanted to determine whether the insertion of a helix breaker, such as threonine, would induce similar or opposite changes in the $E_{M,7}$. We found that insertion of threonine (T68 Insert) caused

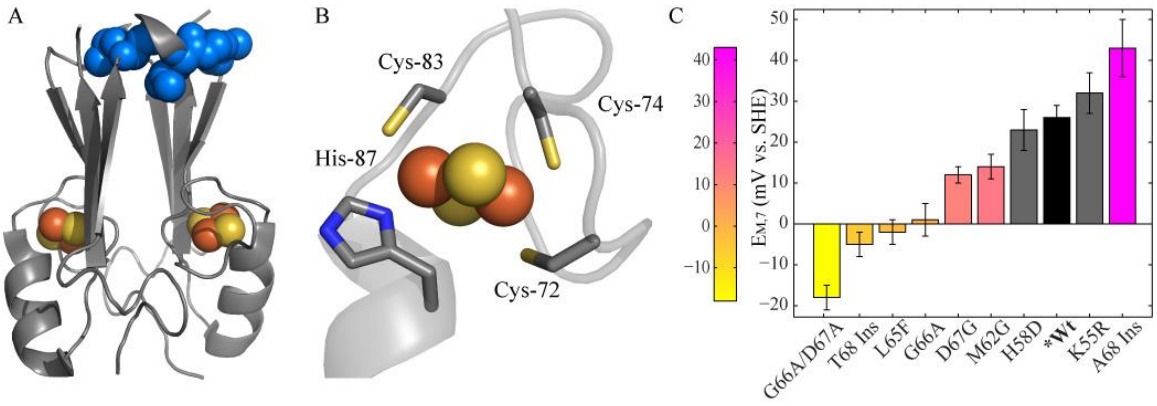

Figure 5.1: Mutagenesis of distal residues in L2 result in both positive and negative shifts in the [2Fe-2S] cluster redox potential at pH 7 ($E_{M,7}$)

(A) Mutated residues in L2 are colored in blue on the structure of mitoNEET. (B) The [2Fe-2S] cluster is coordinated by 3 cysteines and 1 histidine. (C) $E_{M,7}$ values for WT mitoNEET and L2 mutants span a range of ~60 mV. $E_{M,7}$ values are adjusted to standard hydrogen electrode (SHE) values, with errors indicated by cross bars. Wild type is shown in black, and controls are shown in grey.

decrease in $E_{M,7}$, to a new value of -5 ± 4 mV, which is 31 mV less than the WT. These long range redox changes are observed despite the fact that the mutated residues are between 18A and 26 Å away from the [2Fe-2S] cluster, as measured both intraprotomer and interprotomer. To test whether mutagenesis of any region of the protein is capable of resulting in redox changes, we introduced two additional mutations as controls, K55R and H58D. In both cases, the shift in redox potential is under 5mV, despite the fact that these mutations are 7.5, and 9.7 Å away from the [2Fe-2S] cluster (Figure 5.1). Taken together, our results show that residues in the L2 region (Figure 5.1A shown in blue) can communicate with the distally located cluster binding region of the protein and change cluster redox properties.

The L2 Region Strongly Influences the Cluster Transfer Function of MitoNEET as well as the Innate Stability of its [2Fe-2S] Cluster

We discovered that mitoNEET functions as a [2Fe-2S] cluster transfer protein (33). Having shown that allosteric changes in the L2 region are able to affect redox properties (Figure 5.1), we hypothesized that L2 may also play an important role in cluster transfer. Both the M62G and L65F show significant increases in cluster transfer (Figure 5.2A) compared to that observed for the WT protein. The largest shift was seen in the M62G mutant, which transferred over 10-fold faster than the WT (Figure 5.2A). Conversely, cluster transfer is slowed in the G66A, D67G, and G66A/D67A mutant proteins. Finally, both the A68 and T68 insertion mutations showed very similar transfer rates to one another (Figure 5.2A), even though they induce opposite shifts in $E_{M,7}$ (Figure 5.1C). We conclude that the L2 region is able to dramatically affect protein function through allosteric effects on the distal [2Fe-2S] clusters.

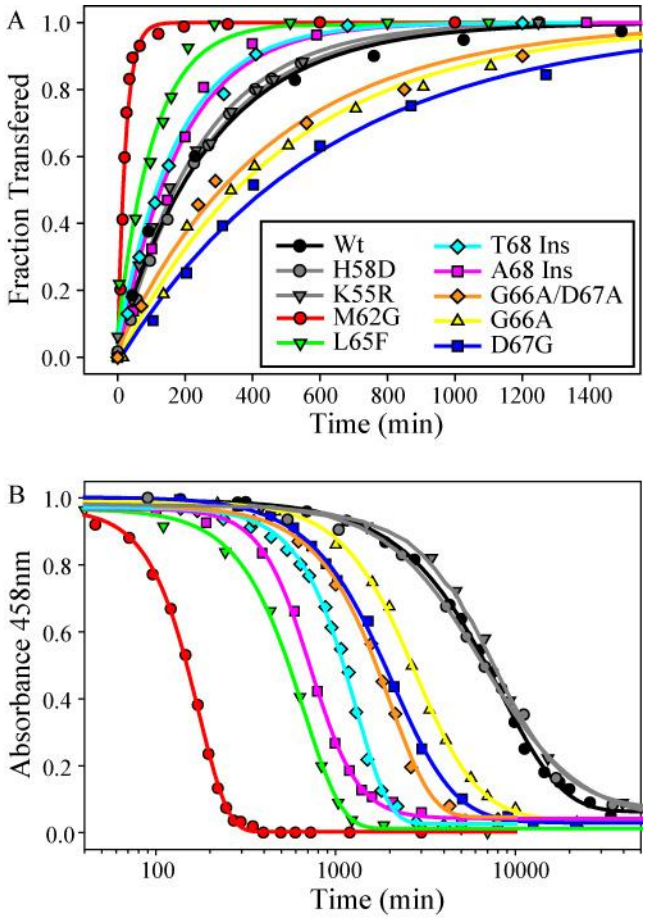

Figure 5.2: The L2 region strongly influences [2Fe-2S] cluster transfer function of mitoNEET as well as cluster stability.
(A) Transfer of the [2Fe-2S] cluster from mitoNEET and L2 mutants to apo-ferredoxin. Cluster transfer experiments were performed aerobically at 35°C at pH 8.0 using 100 μM mitoNEET or L2 mutants and 100 μM apo-Ferredoxin in 50 mM Bis-tris and 100 mM NaCl. (B) Stability of mitoNEET's [2Fe-2S] cluster compared to L2 mutants. Cluster stability measurements were performed aerobically at 35 °C. Loss of the [2Fe-2S] cluster was measured over time as a decrease in absorbance at 458 nm. Studies were performed using 40 μM mitoNEET or L2 mutants in 100 mM Bis-tris and 100 mM NaCl, pH 7.0. Wild type is shown in black and controls are shown in grey.

In addition to cluster transfer, the [2Fe-2S] cluster in mitoNEET may be important for binding to other proteins and small molecule effectors. Besides inhibiting cluster transfer, TZDs are known to stabilize mitoNEET's cluster against release (26). As mitoNEET's cluster stability is dramatically affected by drug binding (26), it is also important to examine the effect that L2 mutants have on cluster stability. We assayed our loop mutants for cluster stability by monitoring loss of mitoNEET's 458 nm absorbance peak over time as described previously. We found that all L2 mutations caused lower stability in the [2Fe-2S] cluster (Figure 5.2B). However, cluster transfer rates do not correlate with cluster stability, indicating that allosteric effects can influence interactions important for cluster transfer and protein stability, although these two effects are independent of one another. Figure 5.2 also includes data for the K55R and the H58D mutations, which show no change in either cluster transfer rates or cluster stability from Wt. We have summarized our findings on the influence of L2 residues for redox potential, cluster transfer, and cluster stability in Table 1.

The Structural Integrity of the Protein is Maintained Upon L2 Mutation

In order to determine whether the L2 mutation effects on cluster properties were the result of conformational rearrangements in the protein, we initiated structural analyses and subsequently obtained crystal structures for the G66A/D67A, A68 Insert, M62G, and D67G mutant proteins and compared them to the WT structure (26-28) (Figure 5.3). Overall, the entire structural fold remains intact upon mutation in L2. In all cases the cluster binding domain of the mutant protein is superimposable with that of the WT protein (Figure 5.3A, B). Only small local changes are observed in the L2 region as a result of the necessary accommodation upon insertion of a residue between residues 67

Table 5.1: Cluster Properties of L2 Mutants

Mutant	$E_{M,7}$ (mV)	Decay half-time (min)	Transfer half-time at 100 µM (min)
WT	+26 ± 3	5000±500	185±15
M62G	+14 ± 3	160±30	14±1
L65F	-2 ± 3	500±80	77±8
G66A	+1 ± 4	3000±300	300±60
D67G	+12 ± 2	2400±100	378±40
G66A/D67A	-18 ± 3	1700±200	223±44
A68 Insert	+43 ± 7	720±50	147±26
T68 Insert	-5 ± 3	1300±100	153±12

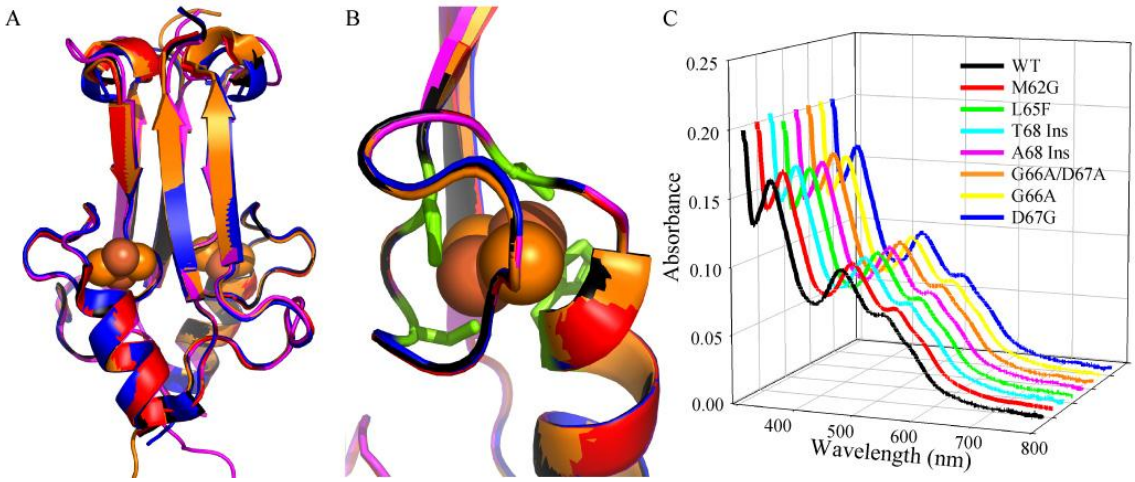

Figure 5.3: Structural analysis of L2 mutants show no conformational changes
(A) Crystal structures of the G66A/D67A (pdb id: 4F2C), A68 Insert (pdb id: 4EZF), M62G (pdb id: 4F28), and D67G (pdb id: 4F1E) mutants are overlaid with WT mitoNEET. (B) An overlay of the cluster binding domains and the cluster coordinating ligands show that the cluster coordinating pockets of all L2 mutants are superimposable with that of WT mitoNEET. (C) The UV-visible spectra of all 7 L2 mutants are identical to that observed for WT. The absorption spectra reports on the environment near the [2Fe-2S] cluster.

and 68 to make the A68 insertion (Figure 5.4A). The D67G mutation crystallized with a completely different set of crystal packing interactions suggesting that the lack of conformational changes is not a result of crystal packing forces. For example, in the D67G structure a packing interaction exists between Lys68 and Asp64, however this packing interaction is not observed in the WT structure. Therefore, mitoNEET is well folded and robust to mutation. We also measured the optical spectrum of the L2 mutants to probe for possible cluster binding changes in the proteins that were not crystallized. We found that all seven L2 mutants have absorption spectra identical to that observed for WT (Figure 5.3C). This is important as we have found that replacement of the cluster-coordinating His87 ligand led to noticeable changes in the visible region (19, 62)(Figure 5.4B) and also that mutations that affected the pKa of His87, namely Lys55, led to small but measureable optical shifts of ~3 nm near the signature 458 nm absorption peak (34). The absence of any spectral shifts taken together with crystallographic analyses indicates that mutations do not alter the overall fold or the cluster binding environment.

All-Atom Simulations Demonstrate Communication Between L2 and the Cluster Coordinating Histidine.

To probe possible mechanisms of cluster property regulation, we investigated the fluctuations of the WT protein in the native basin with SBM simulations. These models are based on energy landscape theory (39) and accurately describe the native state ensemble fluctuations of many systems (49, 103-107). We performed simulations with an all-atom SBM with interactions present in the crystal structure as stabilizing, and with one energetic basin describing the folded state. To examine possible mechanisms whereby distal regions exert control over the [2Fe-2S] cluster, we examined the motions

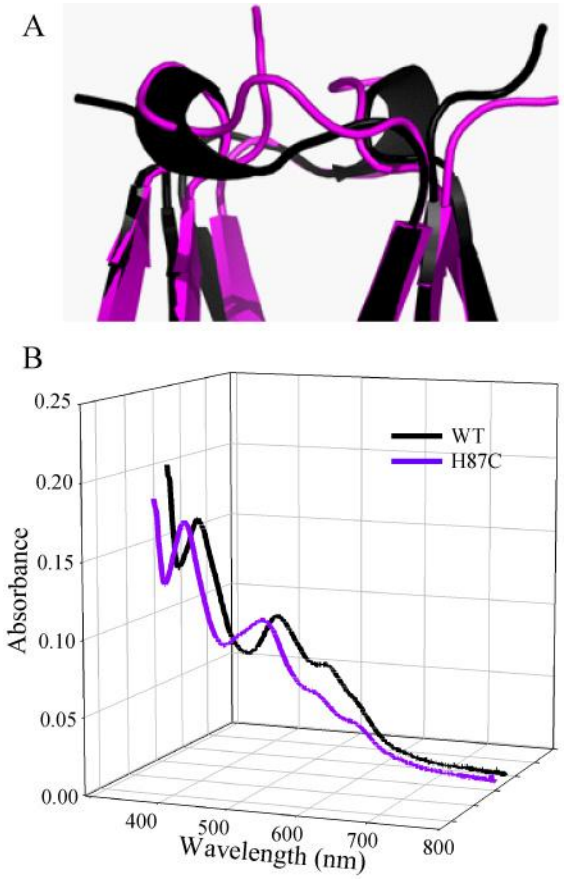

Figure 5.4: Structural heterogeneity in related mutants
(A) An overlay of L2 in the A68 Insertion mutant (pink) with wild type(black). (B) The UV-visible spectra of WT protein (black) compared to the shifted spectra of the H87C mutation (purple).

within the native basin and plotted correlated motions in a covariance matrix (Figure 5.5). The covariance matrix is overlaid with a contact map of the protein in black squares for clarity. Regions of correlated motions are plotted in red, circled, and mapped on to the protein to designate intra (Figure 5.5A,C) and inter-protomer (Figure 5.5B,D) covariance.

The distribution of correlated motions on the protein shows that the L2 region is motionally coupled to regions on the opposite end of the protein. Aside from residues that are in direct contact with each other, regions towards the center of the protein show little correlated motion. In order to further characterize these motions, we use principle component analysis to extract the large scale collective motions from the trajectory. The collective motions of the Cα atoms for the first eigenvector are plotted in Figure 5.6 with 10 superimposed frames from a movie of the first eigenvector. The initial starting frame is colored white, with each subsequent frame moving from white to red for protomer A, and white to blue for protomer B. From the top view, the protein exhibits a torsional motion with $L2_A$ and $L2_B$ twisting together at the top of the beta cap domain in the foreground, and the cluster binding domain moving in the opposite direction in the background. (Figure 5.6A) The side view of the protein (Figure 5.6B) shows that the twisting of the beta cap domain results in scissoring in the cluster binding domain, where $\alpha1_A$ and $L1_B$ move together opposite of $\alpha1_B$ and $L1_A$. Interestingly, the [2Fe-2S] clusters appear to be a pivot point for these motions, so that the entire protein rearranges around the cluster binding region. Much of the cluster coordinating pocket, including the coordinating cysteines, exhibit little motion relative to the rest of the protein, however the Cα atom for the coordinating histidine exhibits a much larger range of motion. (Figure

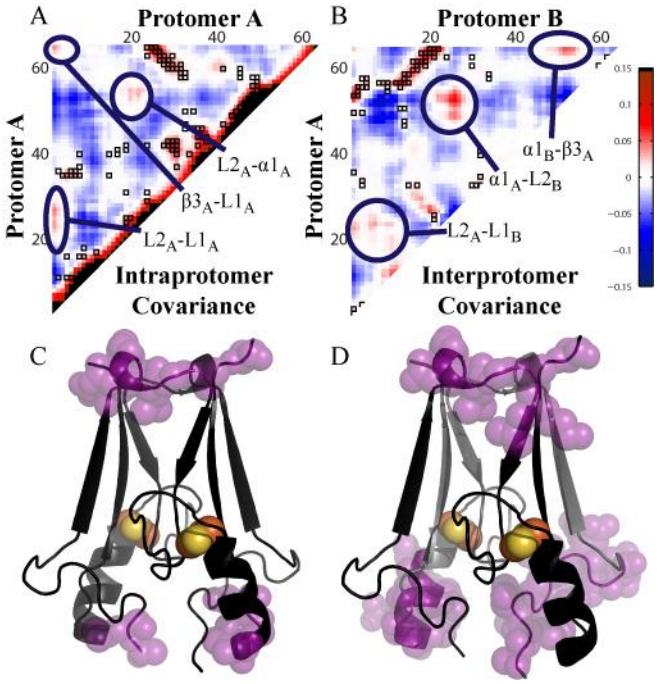

Figure 5.5: Covariance matrices show correlated motions between L2 and the cluster binding domain

Spatial cross correlation matrix showing correlated motions between Cα atoms within a single protomer (A) and between the two protomers (B). Regions of positive covariance are circled and are shown as purple spheres on the protein in C and D. Panel C shows intraprotomer covariance between $L2_A$ and $L1_A$, $L2_A$ and $\beta3_A$, and $L1_A$ with $\alpha1_A$, and panel D shows interprotomer covariance between $L1_B$, and $L2_A$, $L2_A$ and $\alpha1_B$ and $\beta3_A$ with $\alpha1_B$.

5.6C) Because the coordinating histidine is positioned at the bottom of α1, it scissors with α1 away from the cluster coordinating pocket in response to twisting in L2.

DISCUSSION

In conventional systems, the properties of metalloproteins are controlled by local ligand environment. Complex biological systems include an additional control involving the use of strain to exert long-range coupling (50, 64). A simple analogy for this type of control is a clothes pin, where a strained spring is necessary for regulating clasp and release. Pinching a clothes pin at one end rearranges everything around a central pivot point and induces the largest motions at the opposite end of the structure. In mitoNEET, twisting or "pinching" of L2 triggers scissoring at the bottom of the cluster binding domain, so that $\alpha1_A$ and $L1_B$ scissor against $\alpha1_B$ and $L1_A$. The [2Fe-2S] cluster is positioned at the pivot point of these motions, and much of the cluster binding pocket, including the cluster coordinating cysteines exhibit little motion. However, twisting of L2 causes the entire protein to rearrange around the metal center, and results in the coordinating histidine swinging away from the metal center (Figure 5.6). Mutations that alter the flexibility, or expand or collapse L2 may alter the degree to which the coordinating histidine moves. This rearrangement of the protein backbone (Figure 5.6A,B), especially at the coordinating histidine (Figure 5.6C), may account for the measured redox shifts (Figure 5.1) and more importantly the order of magnitude increase in the rate of cluster transfer to an apo-acceptor protein (Figure 5.2).

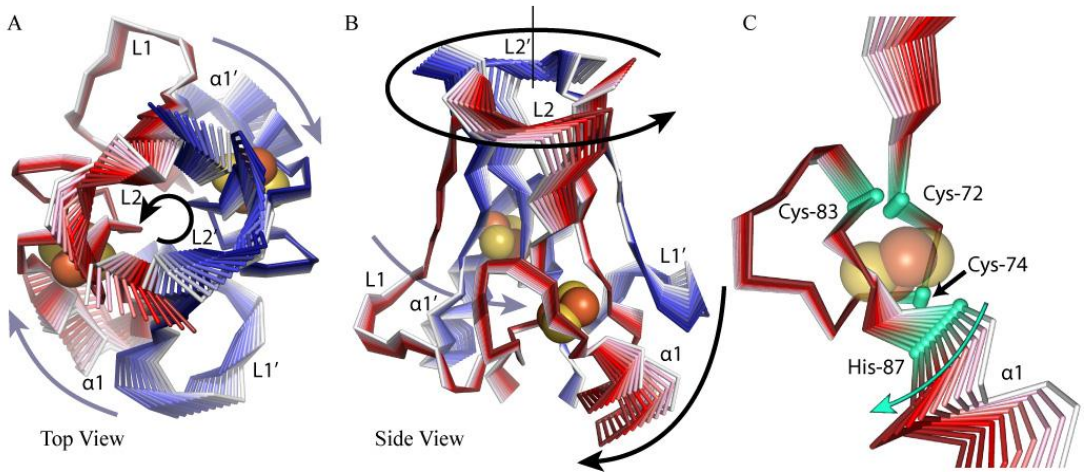

Figure 5.6: Twisting in L2 controls swinging of His87
(A)10 Frames from the collective motions of the Cα atoms for the first eigenvector are superimposed. The initial starting point is colored white, with each subsequent frame moving from white to red for protomer A, and white to blue for protomer B. From the top view, the protein exhibits a torsional motion with L2 and L2' twisting together at the top of the beta cap domain in the foreground, and the cluster binding domain moving in the opposite direction in the background. B) The side view of the protein shows that the twisting of the beta cap domain results in scissoring of α1 and L1' with α1' and L1 in the cluster binding domain, with the [2Fe-2S] cluster acting as a hinge point. C) The backbone of the cluster coordinating pocket and the coordinating cysteines shows little movement, however the coordinating histidine swings open in response to twisting in L2.

A protein going through such hinging motions may move through energetically less favorable states. It has been proposed that partially unfolded states may compete energetically with some of these higher energy states, relieving strain (49). This local unfolding, or cracking (49), is important for the activity for many proteins, and indeed there are cases where the addition of denaturant can speed up enzymatic activity (108-110). In mitoNEET, cracking may be an essential part of the cluster transfer mechanism. An apo-acceptor protein could pinch L2, which would result in displacement of residues in α1 and L1 in the cluster binding domain (Figure 5.6). This displacement could create strain on the protein, causing it to "crack" open and increase accessibility of the metal center which may lead to decreased cluster stability or enhanced cluster transfer. However, it is important to note that cluster stability and transfer efficiency are not necessarily linked(33). In the present studies all L2 mutations increase cluster decay rates. However, not all mutations speed up cluster transfer. Mutations to residues 62, 65, and 68 all increase cluster transfer (Figure 5.2A), however, mutations to residues 66 and 67 slow down cluster transfer despite being destabilizing, suggesting different mechanisms governing cluster stability vs. transfer. One hypothesis for this is that all mutations increase cracking in L1 and α1 and thus increase decay rate; however, residues 66 and 67 may be important for the binding interaction with apo-acceptor proteins. The largely negative charge on the surface of the beta-cap may be a docking site for a mitoNEET protein partner, and judging by the major functional effects that perturbation of the L2 region indeed have on its cluster properties, we postulate that this is likely the case.

CONCLUSIONS

We use a combination of experimental techniques and structure-based simulations to test the hypothesis that a distal loop in mitoNEET acts as regulatory control region for the [2Fe-2S] cluster (100). Mutagenesis to L2 results in significant destabilization of the [2Fe-2S] cluster, and can either abrogate or accelerate cluster transfer by factor of up to 15 fold. Additionally, these perturbations to L2 20Å from the [2Fe-2S] cluster result in significant redox potential, thus demonstrating that the L2 region is able to distally regulate functional properties of the protein found in the cluster binding domain. As the most dramatic mutations do not involve charged amino acids, the observed shifts in redox potential over the 20Å distance cannot be due to electrostatic effects. Crystal structures obtained for several mutants show that the cluster binding domains of the mutant proteins are superimposable with that of the WT protein, indicating that these effects are also not the result of a long range conformational change. Using structure-based simulations we propose a mechanism of communication in which twisting in L2 triggers scissoring in the cluster binding domain, and results in displacement of the coordinating histidine. This long range allosteric control of the metal center suggests that many other MPs may in fact require full use of their uniquely evolved scaffolds to perform complex biological tasks. Taken together, this work provides a new foundation for investigating how metal centers in metalloproteins are influenced by the global motions and expands our understanding beyond the control of simple electron transfer by distal mutations (111). This approach is critical for designing new therapeutics for targeting this class of [2Fe-2S] proteins as well as de novo design of new metalloproteins (112, 113).

SPECIAL METHODS

Protein expression and purification—Over-expression and purification of the soluble fragment of mitoNEET (aa33-108) and L2 mutants was performed as outlined previously (33). Purification of apo-ferredoxin was performed as outlined previously (114).

Cluster transfer kinetics, potentiometric redox titrations, and cluster stability measurements

All UV-Visible absorption spectra were measured from the near UV to the near IR (300 – 800 nm) on a Cary50 spectrophotometer (Varian Inc. Palo Alto CA) equipped with a temperature-controlled cell.

Cluster transfer experiments were performed similarly to previous reports (33). Cluster transfer experiments were performed aerobically at 35 °C at pH 8.0 using 100 μM mitoNEET or L2 mutants and 100 μM apo-ferredoxin in 50 mM Bis-tris 100 mM NaCl. The samples were covered with mineral oil (Hampton Research) to prevent losses due to evaporation. Transfer rates were obtained by following the (423 nm) / (458 nm) ratio corresponding to loss of the mitoNEET 458 nm peak and emergence of the holo-ferredoxin 423 nm peak with time as described previously (33). Data were fit to a single exponential rise and initial transfer rates were determined by taking the slope of the exponential fit in the first 15 minutes.

Optical potentiometric redox titrations were performed as outlined previously (34). Briefly, experiments were performed anaerobically at 25 °C under an argon atmosphere using 50 μM mitoNEET or L2 mutants in 100 mM Bis-tris 100 mM NaCl in the presence of mediators to facilitate efficient electron transfer between the protein and

Ag/AgCl reference electrode (Microelectrodes Inc. Bedford NH). Sodium dithionite (Sigma-Aldrich) was titrated in via syringe to reduce the [2Fe-2S] clusters. After mitoNEET was fully reduced, the protein was re-oxidized by titrating in 10 µL of fixed aliquots of ambient oxygen. Optical scans (300 - 800 nm) were performed following each addition of dithionite or oxygen. Optical potentiometric redox titration data were fit to the Nernst equation as described previously (34).

Cluster stability measurements were performed aerobically at 35 °C. Cluster loss was measured over time as a decrease in absorbance at 458 nm. Studies were performed using varying concentrations of mitoNEET and mutants in 100 mM Bis-tris 100 mM NaCl at pH 7.0.

Crystallization, X-Ray Data Collection, and Refinement

Crystallizations from purified preparation of the mitoNEET G66A/D67A, A68 Insert, M62G, and D67G mutants were performed using conditions to those previously described (115). Crystals that were cooled to cryogenic temperatures were screened for x-ray diffraction at SSRL (116) using the Stanford Automated Mounter (117) operated by Blu-Ice (118). X-Ray diffraction data to resolutions of 1.19 Å, 1.6Å, and 2.4 Å for the A68 Insert mutant, the M62G, and D67G respectively were collected at 100K and a wavelength of 0.9795 Å at SSRL BL 9-2 with the data recorded on a Rayonix MX325 CCD imaging plate detector. For the G66A/D67A mutant, data were collected at 100K and a wavelength of 0.97964 Å at SSRL BL 7-1 with the data recorded on a ADSC Quantum 315R CCD detector. Indexing, integration, and scaling were performed using MOSFLM (119) and SCALA (120) for the G66A/D67A and M672 mutants; and XDS/XSCALE (121) for the A68 Insert and D67G mutants. Diffraction data from the

G66A/D67A, A68 Insert, and M62G mutants were indexed in an orthorhombic space group $P2_12_12_1$ with cell dimensions similar to that previously reported for the wild type (115). The D67G mutant was crystallized in a new orthorhombic crystal form, space group $P2_12_12_1$ with unit cell dimensions of a=56.4 Å b=110.9 Å c=207.3 Å. Molecular replacement and crystallographic refinement was carried out using the previously-determined coordinates of mitoNEET (PDB ID 2qh7) (115) as a starting model. Molecular replacement was performed using PHASER (122). Refinement was implemented with BUSTER (123) for the M62G and D67G mutants; and PHENIX (124) for the G66A/D67A and A68 Insert mutants. Interactive model building and electron density display between rounds of refinement was accomplished using COOT (125). To aid in the modeling of multiple conformations, the program qFit that models disecrete heterogeneities in electron density maps was used (126). Data collection and refinement statistics are shown in Table S2.

The atomic coordinates of the L2 mutants have been deposited in the Protein Data Bank. The A68 Insert was deposited under PDB ID code 4EZF, the M62G was deposited under PDB ID code 4F28, the G66A/D67A mutant was deposited under PDB ID 4f2c, and the D67G mutant was deposited under PDB ID code 4F1E.

All-Atom Simulations

The *S*tructure-based *MO*dels in *G*romacs (SMOG) Web tool (http://smog.ucsd.edu)(57) was used to generate an all-atom structure-based force field from the crystal structure of mitoNEET stored in Protein Data Bank ID 2QH7 (26). The shadow algorithm was used to identify native contacts (59). Simulations were performed using version 4.0.5 of the *GRO*ningen *MA*chine for *C*hemical *S*imulations (GROMACS)

software package (56). The integrator used was stochastic dynamics. The time step τ was 0.0005. Each protomer was temperature coupled separately.

ACKNOWLEDGEMENTS

This work was supported in part by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (NSF, Grant PHY-0822283) and by NSF MCB-1214457, and by National Institutes of Health Grant GM-54038. R.N. acknowledges the support of the Israel Science Foundation - ISF 863/09, E.L.B. was also supported by a San Diego Fellowship. J.A.Z. was supported by the Heme and Blood Proteins Training Grant (GN 5T32DK007233-34).

Chapter 5, in part, is a reprint as it appears in Proceedings of the National Academy of Sciences, 2012. Baxter EL, Zuris JA, Wang C, Vo PLT, Axelrod HL, Cohen AE, Nechushtai R, Onuchic JN, Jennings PA. 2012. The dissertation author was the primary investigator and author of this paper.

Table 5.2: x-ray data collection and refinement statistics

Mutant	A68 Insert	D66G	M62G	G66A/D67A
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795
Resolution (Å)	1.19	2.40	1.55	1.35
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell Dimensions (Å)	a=39.63 b=55.36 c=58.92	a=56.40 b=110.86 c=207.29	a=41.33 b=56.60 c=59.11	a=45.63 b=51.96 c=59.02
Total Reflections	392,189	316,697	290,174	284,487
Unique Reflections	41,819	51,596	20,449	31,045
Multiplicity	9.4(5.9)[1]	6.1(4.2)[1]	14.2(8.3)[1]	9.2(4.5)[1]
Completeness (%)	98.8(98.8)[1]	99.5(99.5)[1]	99.1(99.1)[1]	98.3(85.1)[1]
R_{merge}	5.0(75.1)[1]	6.9(71.7)[1]	4.7(88.6)[1]	6.9(65.7)[1]
R_{meas}	5.3(82.4)[1]	7.5(81.9)[1]	4.8(94.5)[1]	7.7(80.6)[1]
Mean I/σ(I)	19.8(2.3)[1]	15.1(2.0)[1]	24.5(2.0)[1]	16.8(1.8)[1]
Refinement				
Resolution (Å)	1.19	2.40	1.55	1.35
Reflections	41,740	51,524	20,401	30,975
R_{work} (%)	16.5	19.3	21.6	15.3
R_{free} (%)	16.7	22.7	24.9	18.1
rms bond length deviation (Å)	0.009	0.010	0.010	0.008
Rms bond angle deviation (degrees)	1.67	1.14	1.07	1.53
no. atoms	1301	9244	1094	1308
no. solvent	133	89	84	185
mean B-factor ($Å^2$)	19.3	74.0	50.0	17.4

[1]Value in parentheses corresponds to the value for the highest resolution shell

Chapter 6

Conclusions and Future Directions

The goal of this dissertation was to use signals of frustration in folding as an assay for regions of functional importance in a newly discovered diabetes drug target with a unique fold. To this end, we used Structure-Based Models to characterize the folding and functional landscapes of mitoNEET. Through careful analysis of the folding mechanism of the protein, we were able to identify a loop distal from the [2Fe-2S] cluster which creates a constraint in the folding of the protein. We predicted that although this site slows folding, it was specifically conserved, or even selected for, because it conferred a functional advantage. The metal center in mitoNEET makes mitoNEET an advantageous protein to study because the properties linked to it's function are easily monitored, allowing us to test this hypothesis experimentally.

We were able to demonstrate that point mutations to this loop resulted in changes to the properties of the [2Fe-2S] cluster from 20Å away. In traditional systems, properties of metal centers are controlled by the small fraction of amino acids directly surrounding the metal center. The sensitivity of mitoNEET's metal centers to perturbations in a distal loop suggests a structure fine tuned to facilitate allostery. It may be that partner proteins interact with mitoNEET through loop 2, and information is transmitted to the metal centers upon binding. As this diabetes drug target is better characterized, it will be interesting to see whether any binding partners and drugs interact with Loop 2, and what effect they have on the cluster properties.

Nature is often efficient and elegant in its design, and there are many examples of proteins which perform more than one function. Likewise, loop 2 plays more than one role in mitoNEET's structure. In chapter 4, we saw that not only is a strand unswapped conformation of mitoNEET accessible, but also that loop 2 regulates the conformational

balance. Strikingly, loop 2 functions as a conformational switch which can be activated by cluster insertion, suggesting that information between Loop 2 and the cluster is transmitted both ways. This switching mechanism in mitoNEET may be a part of a failsafe mechanism to prevent the release of toxic free iron, or part of mitoNEET's cluster insertion mechanism.

While this paper was under review, a study published 4 new structures from the CDGSH domain family of proteins. All four structures exhibited variations in the number of strands in the beta cap domain, as well as permutations in the arrangement of the strands. One of these structures exhibits a strand unswapped conformation, in which each structural half is composed of a single protomer, instead of being composed of two protomers as is seen in mitoNEET. The beta cap domain in the strand unswapped structure differs from both the native strand swapped structure of mitoNEET as well as the unswapped structure; the arrangement of strands is different, and in fact the beta cap domain is composed of 8 strands instead of the 6 strands seen in mitoNEET. It does however confirm that strand unswapping is possible within this family of proteins, and supports the possibility that mitoNEET may access its own strand unswapped conformation.

Permutations in the beta cap topology may act to differentiate the function of different family members. We saw that point mutations in the beta cap domain of mitoNEET are able to alter cluster properties, therefore it is possible that structural rearrangements of this domain result in metal centers with a wider range of stabilities and redox potentials. Changes in beta cap topology may also differentiate between potential binding partners, as well as alter how it interacts with binding partners. In Chapter 1 we

introduced the idea that mutations which constrain folding may be selected for if they confer a functional advantage to the organism. The wide variations in the beta strand topologies within the CDGSH family of proteins may be an example of this phenomenon, and we look forward to seeing a more extensive characterization of this family of proteins.

References

1. Control CfD & Prevention (2011) National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. (Department of Health and Human Services, Centers for Disease Control and Prevention Atlanta).

2. Reaven GM (1997) Banting lecture 1988: Role of insulin resistance in human disease. *Nutrition* 13(1):64.

3. Cheal KL, *et al.* (2004) Relationship to Insulin Resistance of the Adult Treatment Panel III Diagnostic Criteria for Identification of the Metabolic Syndrome. *Diabetes* 53(5):1195-1200.

4. Lehmann JM, *et al.* (1995) An Antidiabetic Thiazolidinedione Is a High Affinity Ligand for Peroxisome Proliferator-activated Receptor γ (PPARγ). *Journal of Biological Chemistry* 270(22):12953-12956.

5. Olefsky JM & Saltiel AR (2000) PPARγ and the Treatment of Insulin Resistance. *Trends in Endocrinology & Metabolism* 11(9):362-368.

6. Chinetti G, Fruchart JC, & Staels B (2000) Peroxisome proliferator-activated receptors (PPARs): nuclear receptors at the crossroads between lipid metabolism and inflammation. *Inflammation Research* 49(10):497-505.

7. Nawrocki AR, *et al.* (2006) Mice Lacking Adiponectin Show Decreased Hepatic Insulin Sensitivity and Reduced Responsiveness to Peroxisome Proliferator-activated Receptor γ Agonists. *Journal of Biological Chemistry* 281(5):2654-2660.

8. Guan Y, *et al.* (2005) Thiazolidinediones expand body fluid volume through PPAR[gamma] stimulation of ENaC-mediated renal salt absorption. *Nat Med* 11(8):861-866.

9. Zhang H, *et al.* (2005) Collecting duct-specific deletion of peroxisome proliferator-activated receptor γ blocks thiazolidinedione-induced fluid retention. *Proceedings of the National Academy of Sciences of the United States of America* 102(26):9406-9411.

10. He W, *et al.* (2003) Adipose-specific peroxisome proliferator-activated receptor γ knockout causes insulin resistance in fat and liver but not in muscle. *Proceedings of the National Academy of Sciences* 100(26):15712-15717.

11. Jones JR, *et al.* (2005) Deletion of PPARγ in adipose tissues of mice protects against high fat diet-induced obesity and insulin resistance. *Proceedings of the National Academy of Sciences* 102(17):6207-6212.

12. Gavrilova O, *et al.* (2003) Liver Peroxisome Proliferator-activated Receptor γ Contributes to Hepatic Steatosis, Triglyceride Clearance, and Regulation of Body Fat Mass. *Journal of Biological Chemistry* 278(36):34268-34276.

13. LeBrasseur NK, *et al.* (2006) Thiazolidinediones can rapidly activate AMP-activated protein kinase in mammalian tissues. *American Journal of Physiology - Endocrinology And Metabolism* 291(1):E175-E181.

14. Balkrishnan R, *et al.* (2007) Comparisons of rosiglitazone versus pioglitazone monotherapy introduction and associated health care utilization in Medicaid-enrolled patients with type 2 diabetes mellitus. *Clinical therapeutics* 29(6 Pt 1):1306-1315.

15. Beysen C, *et al.* (2008) A pilot study of the effects of pioglitazone and rosiglitazone on de novo lipogenesis in type 2 diabetes. *Journal of Lipid Research* 49(12):2657-2663.

16. Vries CSd & Russell-Jones DL (2009) Rosiglitazone or pioglitazone in type 2 diabetes? *BMJ* 339.

17. Chen Z, *et al.* (2012) Insulin Resistance and Metabolic Derangements in Obese Mice Are Ameliorated by a Novel Peroxisome Proliferator-activated Receptor γ-sparing Thiazolidinedione. *Journal of Biological Chemistry* 287(28):23537-23548.

18. Colca JR, *et al.* (2004) Identification of a novel mitochondrial protein ("mitoNEET") cross-linked specifically by a thiazolidinedione photoprobe. *Am J Physiol Endocrinol Metab* 286(2):E252-260.

19. Wiley SE, Murphy AN, Ross SA, van der Geer P, & Dixon JE (2007) MitoNEET is an iron-containing outer mitochondrial membrane protein that regulates oxidative capacity. *Proc Natl Acad Sci U S A* 104(13):5318-5323.

20. Wiley SE, *et al.* (2007) The outer mitochondrial membrane protein mitoNEET contains a novel redox-active 2Fe-2S cluster. *Journal of Biological Chemistry* 282(33):23745-23749.

21. Kusminski CM, *et al.* (2012) MitoNEET-driven alterations in adipocyte mitochondrial activity reveal a crucial adaptive process that preserves insulin sensitivity in obesity. *Nat Med* 18(10):1539-1549.

22. Chen YF, *et al.* (2009) Cisd2 deficiency drives premature aging and causes mitochondria-mediated defects in mice. *Genes Dev* 23(10):1183-1194.

23. Amr S, *et al.* (2007) A homozygous mutation in a novel zinc-finger protein, ERIS, is responsible for Wolfram syndrome 2. *Am J Hum Genet* 81(4):673-683.

24. Barrett TG & Bundey SE (1997) Wolfram (DIDMOAD) syndrome. *Journal of Medical Genetics* 34(10):838-841.

25. Barrett TG, Bundey SE, Fielder AR, & Good PA (1997) Optic atrophy in Wolfram (DIDMOAD) syndrome. *Eye* 11(6):882-888.

26. Paddock ML, *et al.* (2007) MitoNEET is a uniquely folded 2Fe 2S outer mitochondrial membrane protein stabilized by pioglitazone. *Proc Natl Acad Sci U S A* 104(36):14342-14347.

27. Hou X, *et al.* (2007) Crystallographic studies of human MitoNEET. *J Biol Chem* 282(46):33242-33246.

28. Lin J, Zhou T, Ye K, & Wang J (2007) Crystal structure of human mitoNEET reveals distinct groups of iron sulfur proteins. *Proc Natl Acad Sci U S A* 104(37):14640-14645.

29. Lin J, Zhang L, Lai S, & Ye K (2011) Structure and Molecular Evolution of CDGSH Iron-Sulfur Domains. *Plos One* 6(9):e24790.

30. Conlan AR, *et al.* (2009) Crystal structure of Miner1: The redox-active 2Fe-2S protein causative in Wolfram Syndrome 2. *J Mol Biol* 392(1):143-153.

31. Conlan AR, *et al.* (2009) The novel 2Fe-2S outer mitochondrial protein mitoNEET displays conformational flexibility in its N-terminal cytoplasmic tethering domain. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* 65:654-659.

32. Zuris JA, *et al.* (2012) NADPH Inhibits [2Fe-2S] Cluster Transfer from the Diabetes Drug Target MitoNEET to an Apo-acceptor Protein. *J Biol Chem*.

33. Zuris JA, *et al.* (2011) Facile transfer of [2Fe-2S] clusters from the diabetes drug target mitoNEET to an apo-acceptor protein. *Proc Natl Acad Sci U S A* 108(32):13047-13052.

34. Zuris JA, *et al.* (2010) Engineering the redox potential over a wide range within a new class of FeS proteins. *J Am Chem Soc* 132(38):13120-13122.

35. Bak DW, Zuris JA, Paddock ML, Jennings PA, & Elliott SJ (2009) Redox characterization of the FeS protein MitoNEET and impact of thiazolidinedione drug binding. *Biochemistry* 48(43):10193-10195.

36. Dicus MM, *et al.* (2010) Binding of Histidine in the (Cys)(3)(His)(1)-Coordinated 2Fe-2S Cluster of Human mitoNEET. *Journal of the American Chemical Society* 132(6):2037-2049.

37. Tirrell TF, *et al.* (2009) Resonance Raman studies of the (His)(Cys)3 2Fe-2S cluster of MitoNEET: comparison to the (Cys)4 mutant and implications of the effects of pH on the labile metal center. *Biochemistry* 48(22):4747-4752.

38. Zwanzig R, Szabo A, & Bagchi B (1992) Levinthal's paradox. *Proceedings of the National Academy of Sciences* 89(1):20-22.

39. Bryngelson JD, Onuchic JN, Socci ND, & Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21(3):167-195.

40. Bryngelson JD & Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences* 84(21):7524-7528.

41. Leopold PE, Montal M, & Onuchic JN (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* 89(18):8721-8725.

42. Clementi C, Nymeyer H, & Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298(5):937-953.

43. Finke JM & Onuchic JN (2005) Equilibrium and Kinetic Folding Pathways of a TIM Barrel with a Funneled Energy Landscape. *Biophysical journal* 89(1):488-505.

44. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, & Shakhnovich EI (2002) Molecular Dynamics Simulation of the SH3 Domain Aggregation Suggests a Generic Amyloidogenesis Mechanism. *Journal of Molecular Biology* 324(4):851-857.

45. Levy Y, Wolynes PG, & Onuchic JN (2004) Protein topology determines binding mechanism. *Proceedings of the National Academy of Sciences of the United States of America* 101(2):511-516.

46. Yang S, *et al.* (2004) Domain swapping is a consequence of minimal frustration. *Proceedings of the National Academy of Sciences of the United States of America* 101(38):13786-13791.

47. Capraro DT, Roy M, Onuchic JN, & Jennings PA (2008) Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? *Proc Natl Acad Sci U S A* 105(39):14844-14848.

48. Whitford PC, Gosavi S, & Onuchic JN (2008) Conformational transitions in adenylate kinase. Allosteric communication reduces misligation. *J Biol Chem* 283(4):2042-2048.

49. Whitford PC, Miyashita O, Levy Y, & Onuchic JN (2007) Conformational transitions of adenylate kinase: switching by cracking. *J Mol Biol* 366(5):1661-1671.

50. Gosavi S, Whitford PC, Jennings PA, & Onuchic JN (2008) Extracting function from a beta-trefoil folding motif. *Proc Natl Acad Sci U S A* 105(30):10384-10389.

51. Mills JE, *et al.* (2007) A Novel Disulfide Bond in the SH2 Domain of the C-terminal Src Kinase Controls Catalytic Activity. *Journal of Molecular Biology* 365(5):1460-1468.

52. Ferreiro DU, Hegler JA, Komives EA, & Wolynes PG (2007) Localizing frustration in native proteins and protein assemblies. *Proc Natl Acad Sci U S A* 104(50):19819-19824.

53. Jager M, *et al.* (2006) Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci U S A* 103(28):10648-10653.

54. Karanicolas J & Brooks CL, 3rd (2004) Integrating folding kinetics and protein function: biphasic kinetics and dual binding specificity in a WW domain. *Proc Natl Acad Sci U S A* 101(10):3432-3437.

55. Kalbitzer HR, Spoerner M, Ganser P, Hozsa C, & Kremer W (2009) Fundamental link between folding states and functional states of proteins. *J Am Chem Soc* 131(46):16714-16719.

56. Hess B, Kutzner C, van der Spoel D, & Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput* 4(3):435-447.

57. Noel JK, Whitford PC, Sanbonmatsu KY, & Onuchic JN (2010) SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic acids research* 38(Web Server issue):W657-661.

58. Sobolev V, Sorokine A, Prilusky J, Abola EE, & Edelman M (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics* 15(4):327-332.

59. Noel JK, Whitford PC, & Onuchic JN (2012) The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J Phys Chem B*.

60. He Z-Y, Chitnis PR, & Nechushtai R (1998) Molecular Cloning of the petF Gene Encoding Ferredoxin 1 of the Thermophilic Cyanobacterium Mastigocladus laminosus (Accession No. AF030001) (PGR 98-027). *Plant Physiol* 116:867.

61. Geldenhuys WJ, Funk MO, Awale PS, Lin L, & Carroll RT (2011) A novel binding assay identifies high affinity ligands to the rosiglitazone binding site of mitoNEET. *Bioorg Med Chem Lett* 21(18):5498-5501.

62. Conlan AR, *et al.* (2011) Mutation of the His ligand in mitoNEET stabilizes the 2Fe-2S cluster despite conformational heterogeneity in the ligand environment. *Acta Crystallogr D Biol Crystallogr* 67(Pt 6):516-523.

63. Lindberg MO, Haglund E, Hubner IA, Shakhnovich EI, & Oliveberg M (2006) Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proceedings of the National Academy of Sciences of the United States of America* 103(11):4083-4088.

64. Chavez LL, Gosavi S, Jennings PA, & Onuchic JN (2006) Multiple routes lead to the native state in the energy landscape of the β-trefoil family. *Proceedings of the National Academy of Sciences* 103(27):10254-10258.

65. Shimada J & Shakhnovich EI (2002) The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proceedings of the National Academy of Sciences* 99(17):11175-11180.

66. Wright CF, Lindorff-Larsen K, Randles LG, & Clarke J (2003) Parallel protein-unfolding pathways revealed and mapped. *Nat Struct Mol Biol* 10(8):658-662.

67. Lindberg M, Tangrot J, & Oliveberg M (2002) Complete change of the protein folding transition state upon circular permutation. *Nat Struct Mol Biol* 9(11):818-822.

68. Levy Y, Caflisch A, Onuchic JN, & Wolynes PG (2004) The Folding and Dimerization of HIV-1 Protease: Evidence for a Stable Monomer from Simulations. *Journal of Molecular Biology* 340(1):67-79.

69. Levy Y, Cho SS, Onuchic JN, & Wolynes PG (2005) A Survey of Flexible Protein Binding Mechanisms and their Transition States Using Native Topology Based Energy Landscapes. *Journal of Molecular Biology* 346(4):1121-1145.

70. Gambin Y, *et al.* (2009) Direct single-molecule observation of a protein living in two opposed native structures. *Proceedings of the National Academy of Sciences* 106(25):10153-10158.

71. Schug A, Whitford PC, Levy Y, & Onuchic JN (2007) Mutations as trapdoors to two competing native conformations of the Rop-dimer. *Proceedings of the National Academy of Sciences* 104(45):17674-17679.

72. Schreiber G, Buckle AM, & Fersht AR (1994) Stability and function: two constraints in the evolution of barstar and other proteins. *Structure* 2(10):945-951.

73. Zhang X-j, Baase WA, Shoichet BK, Wilson KP, & Matthews BW (1995) Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Engineering* 8(10):1017-1022.

74. Gosavi S, Chavez LL, Jennings PA, & Onuchic JN (2006) Topological frustration and the folding of interleukin-1 beta. *Journal of Molecular Biology* 357(3):986-996.

75. Shannon P, *et al.* (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* 13(11):2498-2504.

76. Kamada T & Kawai S (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters* 31(1):7-15.

77. Baxter EL, Jennings PA, & Onuchic JN (2011) Interdomain communication revealed in the diabetes drug target mitoNEET. *Proc Natl Acad Sci U S A* 108(13):5266-5271.

78. Okazaki K-i, Koga N, Takada S, Onuchic JN, & Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proceedings of the National Academy of Sciences* 103(32):11844-11849.

79. Dima RI & Thirumalai D (2004) Probing the instabilities in the dynamics of helical fragments from mouse PrPC. *Proceedings of the National Academy of Sciences of the United States of America* 101(43):15335-15340.

80. Cho SS, Levy Y, Onuchic JN, & Wolynes PG (2005) Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerization and aggregation. *Physical Biology* 2(2):S44.

81. Levy Y, Cho SS, Shen T, Onuchic JN, & Wolynes PG (2005) Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery. *Proceedings of the National Academy of Sciences of the United States of America* 102(7):2373-2378.

82. Whitford PC, Onuchic JN, & Wolynes PG (2008) Energy landscape along an enzymatic reaction trajectory: Hinges or cracks? *HFSP Journal* 2(2):61-64.

83. Miyashita O, Wolynes PG, & Onuchic JN (2005) Simple Energy Landscape Model for the Kinetics of Functional Transitions in Proteins. *The Journal of Physical Chemistry B* 109(5):1959-1969.

84. Geldenhuys WJ, Funk MO, Barnes KF, & Carroll RT (2010) Structure-based design of a thiazolidinedione which targets the mitochondrial protein mitoNEET. *Bioorganic & Medicinal Chemistry Letters* 20(3):819-823.

85. Gruebele M (2005) Downhill protein folding: evolution meets physics. *Comptes Rendus Biologies* 328(8):701-712.

86. Shoichet BK, Baase WA, Kuroki R, & Matthews BW (1995) A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences* 92(2):452-456.

87. Friel CT, Alastair Smith D, Vendruscolo M, Gsponer J, & Radford SE (2009) The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat Struct Mol Biol* 16(3):318-324.

88. Nordlund A, *et al.* (2009) Functional features cause misfolding of the ALS-provoking enzyme SOD1. *Proceedings of the National Academy of Sciences* 106(24):9667-9672.

89. Beinert H (2000) Iron-sulfur proteins: ancient structures, still full of surprises. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry* 5(1):2-15.

90. Waldron KJ, Rutherford JC, Ford D, & Robinson NJ (2009) Metalloproteins and metal sensing. *Nature* 460(7257):823-830.

91. Rees DC & Howard JB (2003) The interface between the biological and inorganic worlds: iron-sulfur metalloclusters. *Science* 300(5621):929-931.

92. Gray HB & Winkler JR (1996) Electron transfer in proteins. *Annual review of biochemistry* 65:537-561.

93. Bertini I (2007) *Biological inorganic chemistry : structure and reactivity* (University Science Books, Sausalito, Calif.) pp xxv, 739 p.

94. Napier I, Ponka P, & Richardson DR (2005) Iron trafficking in the mitochondrion: novel pathways revealed by disease. *Blood* 105(5):1867-1874.

95. Johnson DC, Dean DR, Smith AD, & Johnson MK (2005) Structure, function, and formation of biological iron-sulfur clusters. *Annual review of biochemistry* 74:247-281.

96. Lill R & Muhlenhoff U (2008) Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annual review of biochemistry* 77:669-700.

97. Bak DW, Zuris JA, Paddock ML, Jennings PA, & Elliott SJ (2009) Redox Characterization of the FeS Protein MitoNEET and Impact of Thiazolidinedione Drug Binding. *Biochemistry* 48(43):10193-10195.

98. Hernandez C, Genesca J, Ignasi Esteban J, Garcia L, & Simo R (2000) [Relationship between iron stores and diabetes mellitus in patients infected by hepatitis C virus: a case-control study]. *Medicina clinica* 115(1):21-22.

99. Jiang R, *et al.* (2004) Body iron stores in relation to risk of type 2 diabetes in apparently healthy women. *JAMA : the journal of the American Medical Association* 291(6):711-717.

100. Baxter EL, Jennings PA, & Onuchic JN (2012) Strand swapping regulates the iron-sulfur cluster in the diabetes drug target mitoNEET. *Proc Natl Acad Sci U S A* 109(6):1955-1960.

101. Gosavi S, Chavez LL, Jennings PA, & Onuchic JN (2006) Topological frustration and the folding of interleukin-1 beta. *J Mol Biol* 357(3):986-996.

102. Lammert H, Schug A, & Onuchic JN (2009) Robustness and generalization of structure-based models for protein folding and function. *Proteins: Structure, Function, and Bioinformatics* 77(4):881-891.

103. Whitford PC, *et al.* (2009) An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* 75(2):430-441.

104. Roy M, *et al.* (2005) The native energy landscape for interleukin-1beta. Modulation of the population ensemble through native-state topology. *J Mol Biol* 348(2):335-347.

105. Haliloglu T & Bahar I (1998) Coarse-grained simulations of conformational dynamics of proteins: application to apomyoglobin. *Proteins* 31(3):271-281.

106. Ratje AH, *et al.* (2010) Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* 468(7324):713-716.

107. Whitford PC, *et al.* (2010) Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA* 16(6):1196-1204.

108. Fan YX, Ju M, Zhou JM, & Tsou CL (1995) Activation of chicken liver dihydrofolate reductase in concentrated urea solutions. *Biochimica et biophysica acta* 1252(1):151-157.

109. Zhang HJ, Sheng XR, Pan XM, & Zhou JM (1997) Activation of adenylate kinase by denaturants is due to the increasing conformational flexibility at its active sites. *Biochemical and biophysical research communications* 238(2):382-386.

110. Duffy TH, Sato JK, Vitols KS, & Huennekens FM (1985) L1210 dihydrofolate reductase: activation and enhancement of methotrexate sensitivity. *Advances in enzyme regulation* 24:13-25.

111. Nechushtai R, *et al.* (2011) Allostery in the ferredoxin protein motif does not involve a conformational switch. *Proceedings of the National Academy of Sciences* 108(6):2240-2245.

112. DeGrado WF, Summa CM, Pavone V, Nastri F, & Lombardi A (1999) De novo design and structural characterization of proteins and metalloproteins. *Annual review of biochemistry* 68:779-819.

113. Lu Y, Yeung N, Sieracki N, & Marshall NM (2009) Design of functional metalloproteins. *Nature* 460(7257):855-862.

114. Fish A, Danieli T, Ohad I, Nechushtai R, & Livnah O (2005) Structural basis for the thermostability of ferredoxin from the cyanobacterium Mastigocladus laminosus. *J Mol Biol* 350(3):599-608.

115. Paddock ML, *et al.* (2007) MitoNEET is a uniquely folded 2Fe–2S outer mitochondrial membrane protein stabilized by pioglitazone. *Proceedings of the National Academy of Sciences* 104(36):14342-14347.

116. Soltis SM, *et al.* (2008) New paradigm for macromolecular crystallography experiments at SSRL: automated crystal screening and remote data collection. *Acta Crystallographica Section D* 64(12):1210-1221.

117. Cohen AE, Ellis PJ, Miller MD, Deacon AM, & Phizackerley RP (2002) An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot. *Journal of Applied Crystallography* 35(6):720-726.

118. McPhillips TM, *et al.* (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of Synchrotron Radiation* 9(6):401-406.

119. Battye TGG, Kontogiannis L, Johnson O, Powell HR, & Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D* 67(4):271-281.

120. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D* 62(1):72-82.

121. Kabsch W (2010) XDS. *Acta Crystallographica Section D* 66(2):125-132.

122. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *Journal of Applied Crystallography* 40(4):658-674.

123. Smart OS, *et al.* (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallographica Section D* 68(4):368-380.

124. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* 66(2):213-221.

125. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* 60(12 Part 1):2126-2132.

126. van den Bedem H, Dhanik A, Latombe J-C, & Deacon AM (2009) Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallographica Section D* 65(10):1107-1117.