COMPUTATIONAL STUDIES OF MUTUALLY EXCLUSIVE FOLDING IN A TWO-DOMAIN MOLECULAR SWITCH

by

Brandon Michael Mills

Submitted to the University

Honors College in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

UNIVERSITY HONORS COLLEGE

This thesis was presented

by

Brandon M. Mills

It was defended on

April 8, 2009

and approved by

Jeffrey L. Brodsky, Biological Sciences, University of Pittsburgh

Irwin D. Kuntz, Pharmaceutical Chemistry, University of California, San Francisco

Daniel M. Zuckerman, Computational Biology, University of Pittsburgh

Thesis Advisor: Lillian T. Chong, Chemistry, University of Pittsburgh

DEDICATION

This work is dedicated to the late Jean Marie Mills and Richard Alan Barber for reminding me every day why we try so very hard.

ACKNOWLEDGEMENTS

This work was completed thanks to the support and encouragement of many friends, family members and mentors. I could not ask for a better advisor than Prof. Lillian Chong, who was brave enough to entrust one of her many great project ideas to me in my junior year. I have her to thank for the great strides made in these past two years that allowed me to produce this thesis. I thank my committee members for their exceptional service: Prof. Jeffrey Brodsky, Prof. Daniel Zuckerman, and particularly Prof. Irwin Kuntz from UCSF for taking time out of his visit to serve on the committee. I consider each of them to be outstanding scientists and feel fortunate that their insights helped to shape this document. I also extend thanks to Prof. Michael Grabe and members of the Grabe group for frequent suggestions and guidance throughout the project. I would like to thank my fellow members of the Chong group for their friendship and assistance on innumerable occasions. Finally, I wish to thank my parents, Alan and Diana Mills, who have supported everything I've done since day one. The path I have chosen to follow has been taxing; I am greatly indebted to all who have been there along the way.

Chapter 2 of this thesis was originally published in *J. Mol. Biol.* in 2009.

COMPUTATIONAL STUDIES OF MUTUALLY EXCLUSIVE FOLDING IN A TWO-DOMAIN MOLECULAR SWITCH

Brandon Michael Mills, B. Phil.

University of Pittsburgh, 2009

Enzymatic proteins have their activity tightly regulated, often via conformational switching (shape-changing) events which can turn them on or off in a reversible fashion. A change in shape at one location on a protein can induce a change at another location. The Loh Group at SUNY Medical School has engineered a model system for studying such changes in molecular switches by inserting a guest protein (ubiquitin) into a host (barnase). The two protein domains undergo a thermodynamic tug-of-war that is concluded by the mechanically induced unfolding (and deactivation) of one domain. It has been experimentally shown that through changes in environmental conditions or the addition of effector molecules, the unfolded domain can refold by unfolding its competitor. Methodologies for the design of engineered switches may be used to design novel biological sensors and therapeutics.

However, it is difficult to obtain structural information for these molecular switches due to their partially unfolded nature. Therefore, we have conducted atomistic and coarse-grained simulations in order to gain structural insight into mutually exclusive folding. To our knowledge, the simulations described in this thesis document are the first at any level of structural detail to show that the folding of one protein domain can drive the unfolding of another.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 OVERVIEW

Biological systems are characterized by their ability to precisely maintain the internal conditions necessary for life. This is possible because these systems have found ways to detect and respond to stimuli rapidly and effectively. Engineers and drug developers are also interested in finding ways to detect the presence of biologically relevant molecules by producing an enzymatic or chemically detectable response. Researchers are increasingly making use of biological macromolecules such as proteins or nucleic acids as scaffolds or building blocks in order to engineer molecular switches with novel and highly selective signal-response couplings [1.1-1.3].

Many proteins undergo conformational switching in response to signals from the surrounding molecular environment. The molecular chaperone Hsp90 interconverts between an open ATP-bound state and a closed ADP-bound state [1.4]. The ATP-bound state binds unfolded or misfolded proteins; then, the hydrolysis of ATP to ADP causes Hsp90 to clamp down on the target protein and induce folding. Ion channels also behave as molecular switches, transitioning between open and closed states in response to a number of signals including membrane potential [1.5], phosphorylation [1.6], or ligand binding [1.7]. By far the most common signaling event for activating or deactivating conformational changes is the addition or removal of a phosphate group by a kinase. The human proteome consists of hundreds of kinases and phosphatases that

switch proteins on and off by the addition or removal of a phosphate group [1.8, 1.9]. Of particular interest are proteins that consist of multiple domains, or distinct units that that are able to fold into their native conformations when isolated from the rest of the protein [1.10]. Over two-thirds of human proteins are composed of multiple domains [1.11].

Two protein domains can be linked (both naturally and *via* engineering) in either an end-to-end fashion or via insertion. While end-to-end fusions have been used extensively as a protein engineering tool (for instance, to monitor gene expression), the insertion of one domain into another causes the two domains to become thermodynamically linked such that conformational changes in one domain can be transferred to the other [1.12]. This domain insertion allows the activity of a signaling domain to be linked to the conformational state of a sensor domain.

Under most of the design schema currently being utilized for engineering novel domain-inserted molecular switches, the inserted domain must have proximal amino and carboxyl termini in order to keep both domains structurally intact [1.13-1.20]. However, Stewart Loh and co-workers have formulated an interesting alternative requiring the opposite: in their paradigm, dubbed “mutually exclusive folding”, a protein with a large N-to-C terminal distance is inserted into a small-diameter surface loop on another protein [1.21]. As a result of this design criterion, the mechanical strain imposed on each domain by the other’s folded structure will result in a thermodynamic tug-of-war between the two domains. If the strain is greater than the intrinsic stability of one domain, that domain will unfold in order to relieve the strain. Since the intrinsic stability of each domain is dependent on external conditions such as temperature and the presence of effector molecules, a change in these conditions can produce a switching event, in which the unfolded domain refolds and causes the other domain to unfold.

A barnase-ubiquitin fusion protein has been engineered using the mutually exclusive folding design paradigm; the ubiquitin gene was inserted between the Lys66 and Ser67 codons of barnase (Figure 1-1) [1.21]. Barnase is a ribonuclease produced and secreted by the bacteria *Bacillus amyloliquefaciens* that is lethal to both prokaryotic and eukaryotic cells [1.22]. *Bacillus amyloliquefaciens* is able to survive by co-expressing the barnase inhibitor barstar. Ubiquitin is a key component of many protein degradation pathways. These proteins were chosen based on a number of criteria. First, there proteins have already been the basis for a large number of structural and protein folding studies. Additionally, the activity of the barnase domain can be observed *in vivo* by cell death. Most importantly, these proteins fit the primary design criterion, since ubiquitin has a much larger end-to-end distance than that of the chosen insertion loop on barnase (Figure 1-1) [1.21]. The barnase-ubiquitin fusion protein was shown, *via* circular dichroism (CD) spectroscopy, tryptophan fluorescence and *in vivo* enzymatic activity monitoring in *E. coli* to exhibit temperature-based switching from a state with barnase folded near 10 °C to one with ubiquitin folded near 30 °C [1.21]. In another fusion protein, designed by inserting the GCN4 leucine zipper into the barnase surface loop, the binding of DNA at the GCN4 domain was shown to unfold and deactivate the barnase domain [1.23].

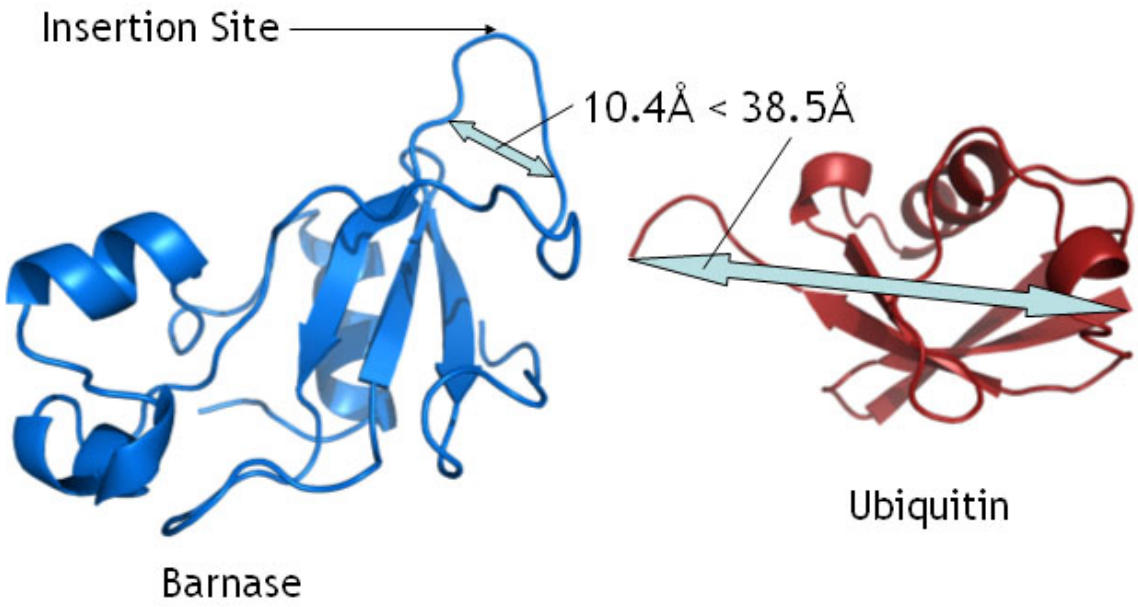

Figure 1-1. Representation of the proteins barnase and ubiquitin. The large amino-to-carboxyl end distance of ubiquitin compared to the barnase surface loop distance fit the mutually exclusive folding design criterion. Images were generated using the PyMOL program (DeLano Scientific).

1.2 SCOPE

While CD spectroscopy experiments are able to provide low-resolution structural information about molecular switches, higher-resolution detail from x-ray crystallographic or NMR spectroscopy studies would be difficult to obtain for mutually exclusive folding switches due to their partially unstructured nature. Disorder in the unfolded protein domain prevents the formation of crystals and can lead to broadening of resonance peaks in NMR spectra that are difficult to resolve. Furthermore, the high protein concentrations needed for NMR study may drive the unstructured domains to stabilize themselves through domain-swapping interactions

[1.24]. Therefore, physics-based computer simulations are a natural alternative for providing a detailed view of the mechanisms for mutually exclusive folding.

This thesis explores the folding antagonism between the protein domains in barnase-ubiquitin fusion proteins using computer simulations at two levels of detail: atomistic and residue-level coarse-grained (Figure 1-2). Gaining a mechanistic and structural familiarity with the conformational and transitional behavior of mutually exclusive folders advances an understanding of engineered and natural molecular switches. Understanding of these multidomain proteins is essential towards improving general methods of molecular switch design, with extensive therapeutic and molecular sensing applications.

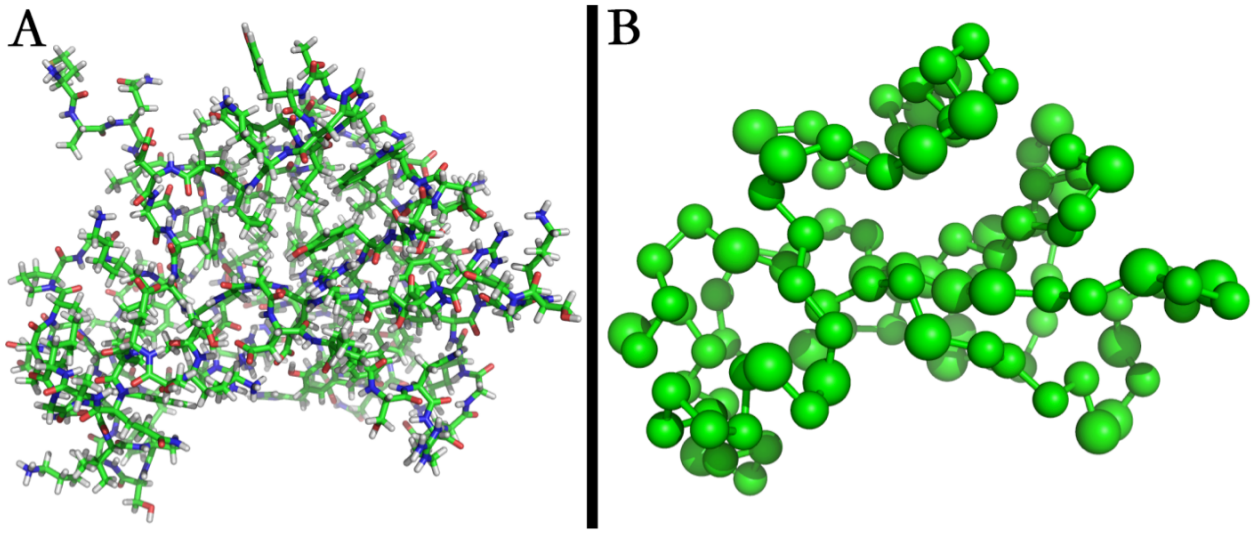

Figure 1-2. Models for the barnase protein used in simulations. Images were generated using the PyMOL program (DeLano Scientific). (A) Fully atomistic model of folded barnase with coordinates from the barnase crystal structure (PDB code 1A2P) [1.25]. Results from simulations at this level of detail are given in Chapter 2. (B) Model of folded barnase at residue-level detail. C_α-bead pseudoatoms are placed at the coordinates of the C_α atoms from the fully atomistic model and connected by pseudobonds. Results from simulations at this level of detail are given in Chapter 3.

Chapter 2 examines how the degree of unfolding of the barnase and ubiquitin domains is sensitive to changes in the length of flexible linker peptides used to join the domains. These

proteins were studied using atomistic, unforced simulations corroborated by CD spectroscopy experiments conducted by Thomas Cutler, David Lubin, and Stewart Loh at SUNY Medical School [1.24]. To our knowledge, these simulations are the first to show the mechanically induced unfolding of one protein domain by another at the atomistic level.

Chapter 3 further examines folding and unfolding in the barnase-ubiquitin fusion proteins using coarse-grained simulations that are capable of achieving better sampling and longer timescales than atomistic simulations. While mechanically induced unfolding was observed in our atomistic simulations [1.24], the protein domains had only partially unfolded after 50 ns of simulation. Our coarse-grained simulations show many complete unfolding and refolding transitions. In simulations of the barnase-ubiquitin fusion protein starting from a state where barnase is folded and ubiquitin is unfolded, refolding of the ubiquitin domain drives the unfolding of barnase. Additionally, the binding of the barstar inhibitor barnase to the fusion protein was shown to dramatically increase the folding of the barnase domain while slightly decreasing the extent of folding of the ubiquitin domain.

1.3 COMPUTATIONAL METHODOLOGY

In the ideal case, all simulations would be conducted from fully atomistic models for the proteins immersed in boxes with explicit water molecules. However, due to the timescales of protein unfolding and folding transitions (microseconds to milliseconds or beyond) and the current state of available computing resources, we must approximate the effects of solvent by using the Langevin equation of motion:

$$M\ddot{X} = -\nabla U(X) - \zeta\dot{X} + \delta R(t)$$

This equation determines the forces applied to each atom during each step of the simulation. The force on each atom comes from three terms modeling particle interactions, solvent friction, and random perturbations of the system. The interaction of particles in the simulation is determined by the force field implemented in the simulations. Our atomistic simulations employ the AMBER99ffSB force field [1.26], while the coarse-grained simulations use a Gō-type potential previously used by others [1.27, 1.28] that has been parameterized to reproduce the desired thermodynamic properties of the individual protein domains.

The second and third terms are used to implicitly model the solvent. The second term models the frictional force on each atom, where ζ is a collision frequency that models the solvent viscosity and reduces the force on each atom. The third term in this equation applies a random force to atoms, modeling random high-velocity collisions with solvent molecules by randomly perturbing the system. The Dirac delta δ ensures that the mean of these applied random forces is zero. Studies have shown that lowering solvent viscosity by reducing ζ accelerates reaction rates while not significantly changing the states populated [1.29, 1.30]. It was for this reason that we employed Langevin Dynamics in our fully atomistic simulations.

The coarse-grained simulations described in Chapter 3 used a Brownian dynamics algorithm. Brownian dynamics is an approximation of Langevin dynamics in which we assume the inertial force (the acceleration) is insignificant compared to the viscosity force. As a result, the time evolution of a Brownian dynamics system can be formulated as the following equation [1.28, 1.31]:

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \sum_j \mathbf{D}_{ij}\mathbf{F}_j \Delta t / k_B T + \mathbf{R}_i$$

where, for each atom *i*, $\mathbf{r_i}$ is the position vector on the atom at time t, **D** is the diffusion tensor, **F** is the total force acting on each other atom *j* in the simulation, $\mathbf{R}_i$ is the random displacement applied to atom *i*, k_B is the Boltzmann constant, and T is the simulation temperature in Kelvin.

The diffusion tensor is an N × N matrix of 3 × 3 matrices, with each i,j pair corresponding to a 3 × 3 matrix that describes the coupling of the components of motion for atoms i and j. In many Brownian dynamics simulations, hydrodynamic interactions are not calculated. As a result, the correlated motions of water are not modeled in these simulations, and the displacement of each atom is not affected by the forces acting on other atoms. In the calculation, all of the off-diagonal elements of the N × N supermatrix are set to zero. To include hydrodynamics, the off-diagonal elements of this matrix are determined by implementing the Rotne-Prager-Yamakawa diffusion tensor calculation [1.32, 1.33]. It has been recently shown that the inclusion of hydrodynamics can accelerate protein folding by 2-3–fold [1.34]. Folding can be further accelerated by decreasing the hydrodynamic radii, the tradeoff being that experimental translational and rotational diffusion coefficients are not accurately reproduced in this case [1.34]. We have included hydrodynamic interactions in the manner implemented in [1.34] in our coarse-grained BD simulations using a reduced value for the hydrodynamic radii.

2.0 EFFECT OF INTERDOMAIN LINKER LENGTH ON AN ANTAGONISTIC FOLDING-UNFOLDING EQUILIBRIUM BETWEEN TWO PROTEIN DOMAINS

Thomas A. Cutler[1], Brandon M. Mills[2], David J. Lubin[1], Lillian T. Chong[2] and Stewart N. Loh[1]

[1]Department of Biochemistry & Molecular Biology, SUNY Upstate Medical University, 750 East Adams Street, Syracuse NY 13210

[2]Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh PA 15260

This chapter was previously published in *J. Mol. Biol.* (2009) **386**, 854-868, and is reproduced here in its entirety. Laboratory experiments were conducted by TA Cutler, DJ Lubin and SN Loh. Computer simulations were conducted by BM Mills and LT Chong.

2.1 ABSTRACT

Fusion of one protein domain with another is a common event in both evolution and in protein engineering experiments. When insertion is at an internal site (*e.g.* a surface loop or turn), as opposed to one of the termini, conformational strain can be introduced into both domains. Strain is manifested by an antagonistic folding-unfolding equilibrium between the two domains, which we previously showed can be parameterized by a coupling free energy term (ΔG_X). The extent

of strain is predicted to depend primarily on the ratio of the N-to-C distance of the guest protein to the distance between ends of the surface loop in the host protein. Here, we test that hypothesis by inserting ubiquitin into the bacterial ribonuclease barnase, using peptide linkers from zero to ten amino acids each. ΔG_X values are determined by measuring the extent to which Co^{2+} binding to an engineered site on the ubiquitin domain destabilizes the barnase domain. All-atom, unforced Langevin dynamics simulations are employed to gain structural insight into the mechanism of mechanically induced unfolding. Experimental and computational results find that the two domains are structurally and energetically uncoupled when linkers are long, and that ΔG_X increases with decreasing linker length. When the linkers are less than two amino acids, strain is so great that one domain unfolds the other. However, the protein is able to refold as dimers and higher-order oligomers. The likely mechanism is a three-dimensional domain swap of the barnase domain, which relieves conformational strain. The simulations suggest that an effective route to mechanical unfolding begins with disruption of the hydrophobic core of barnase by the loss of the N-terminal a-helix.

2.2 INTRODUCTION

The goal of this study is to define the structural and thermodynamic mechanism by which folding of one protein domain is coupled to unfolding of another domain in a new class of engineered, bi-functional proteins. According to this design, which we call ‘mutually exclusive folding’, a guest protein is inserted into a surface loop of a host protein. If the N-to-C terminal distance of the guest is longer than the distance between ends of the surface loop of the host, a thermodynamic struggle ensues in which each protein attempts to mechanically unfold the other.

The guest exerts a stretching force on the host at the point of insertion. The host compresses the termini of the guest. If the above distance differential is large enough, the two native structures are incompatible. The host splits the guest in two, or the guest compresses and unfolds the host, depending on which protein is more intrinsically stable. The protein thus interconverts between two functional forms. This property can be exploited to generate a switching mechanism that is cooperative, reversible, and responsive to a variety of effector signals, including ligand binding and changes in temperature or pH. For example, by inserting the GCN4 DNA binding domain into the ribonuclease barnase (Bn), we created an enzyme whose activity is allosterically regulated by site-specific DNA binding [2.1].

We previously characterized the mutually exclusive folding mechanism by inserting ubiquitin (Ub) into Bn to create the barnase-ubiquitin (BU) fusion protein [2.2, 2.3]. Ub (76 amino acids) was inserted between residues 66 and 67 of Bn (110 amino acids), at the tip of a solvent-exposed loop whose ends are ~11 Å apart (Figure 2-1A). The minimal folding mechanism of BU, in which two-state folding is assumed for each domain, consists of the four states shown in black in Figure 2-1B. The antagonistic interaction is parameterized by a coupling free energy term ΔG_X [2.3]. ΔG_X is the energetic penalty imposed on folding of one domain by the native structure of the other. We hypothesize that ΔG_X will depend largely on the length of the linker peptides used to join the two proteins. If very long linkers are used, then the two domains fold and unfold independently and $\Delta G_X = 0$. As the linkers are progressively shortened, each domain begins to exert strain on the other, causing ΔG_X to increase. If ΔG_X exceeds the intrinsic stability of Bn (ΔG_{Bn}) or of Ub (ΔG_{Ub}), then folding becomes mutually exclusive. The two domains cannot exist simultaneously in their native states. Linker length is therefore expected to define three coupling regimes: zero ($\Delta G_X = 0$), intermediate ($0 < \Delta G_X <$

ΔG_{Bn}, ΔG_{Ub}), and strong ($\Delta G_X > \Delta G_{Bn}, \Delta G_{Ub}$). In the strong coupling limit, the molecule interconverts between two functional forms. The position of the conformational equilibrium is governed by the relative stabilities of the two proteins.

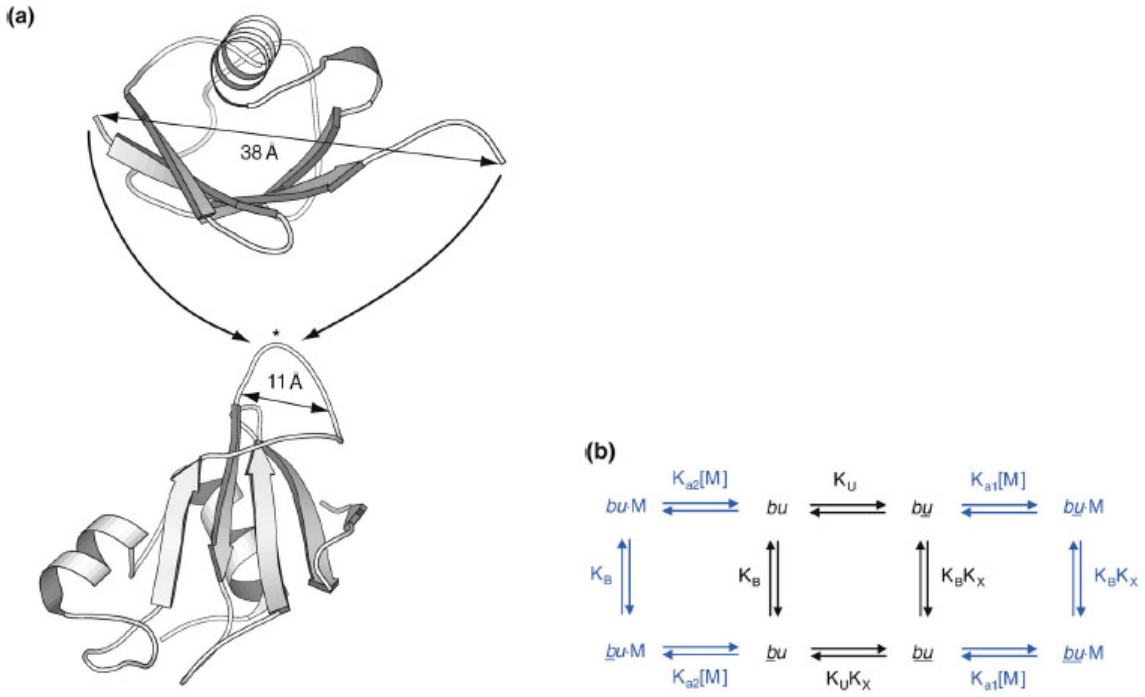

Figure 2-1. Design of BU fusion proteins and minimal folding mechanism. (A) X-ray crystal structures of Ub (top) and Bn (bottom), showing the site of insertion (asterisk). C_α- C_α distances between N- and C-termini of Ub and between the ends of the Bn surface loop (Ser57-Thr70) are indicated. (B) Folding mechanism of BU. Underlined letters and non-underlined letters denote folded and unfolded domains, respectively. Metal (M) free states are colored black and metal bound states are shown in blue. K_{a1} and K_{a2} are the association constants for metal binding to folded and unfolded Ub domains, respectively. K_U and K_B are the equilibrium constants for folding of the Ub and Bn domains when the other is unfolded, and are related to folding free energy changes by the relationships $\Delta G_U = -RT \cdot \ln K_U$ and $\Delta G_B = -RT \cdot \ln K_B$. ΔG_U and ΔG_B decrease with denaturant concentration according to the linear extrapolation equation: $\Delta G_U = \Delta G_U^{H2O} - m_U[GdnHCl]$ and $\Delta G_B = \Delta G_B^{H2O} - m_B[GdnHCl]$, where ΔG_U^{H2O} and ΔG_B^{H2O} are the values in the absence of denaturant and m_U and m_B are proportional to the difference in accessible surface between folded and unfolded states of each domain [2.42]. K_X is the equilibrium constant for coupling of the Bn and Ub domains, where $\Delta G_X = -RT \cdot \ln K_X$.

In an earlier study we estimated the thermodynamic parameters of the mechanism by introducing destabilizing mutations into the Ub domain [2.3]. The stability of the Bn domain was inversely proportional to the stability of the Ub domain, as specified by the coupled equilibria in Figure 2-1B. Fitting the experimental data to the model yielded ΔG_B^{H2O} = 7.5 kcal mol^{-1}, ΔG_U^{H2O} = 5.2 kcal mol^{-1} and ΔG_X ~ 4 kcal mol^{-1} for wild-type (WT) BU. The linkers used in that study were Gly-Thr and Gly-Ala-Ser. Thus, coupling appears to be in the intermediate regime when linkers are two and three amino acids in length. The two domains destabilize each other significantly, but not to the point where one fully unfolds the other. We hypothesized that shortening the linker peptides would intensify the conformational strain between domains and thereby increase ΔG_X.

Here, we test that hypothesis by creating a series of BU variants with linker peptides ranging in length from zero to ten Gly residues. Varying the linker length while keeping the sequences of the domains constant is an orthogonal test of the mutually exclusive folding mechanism. We predict that ΔG_X will increase with decreasing linker length, and will exceed ΔG_B^{H2O} and ΔG_U^{H2O} when the linkers are less than ~2 amino acids each. To measure ΔG_X, we introduce a bi-His metal binding site into the Ub domain via the K6H mutation [2.4, 2.5]. Zn^{2+} or Co^{2+} bind to the side chains of His6 and His68. We previously showed that ΔG_X can be determined most accurately by stabilizing the Ub domain and measuring changes in stability of the Bn domain [2.3]. Metal binding is the preferred method to stabilize Ub, as the mutations known to increase stability do so by optimizing surface electrostatics [2.6], and their effect is significantly reduced by the high ionic strength of the guanidine hydrochloride (GdnHCl) solutions employed in this study. The additional metal-bound states of K6H BU are shown in blue in Figure 2-1B.

The second objective of this study is to define some of the structural and energetic considerations that guide the evolution of multidomain proteins. Over two-thirds of human proteins are composed of more than one domain [2.7]. In most cases, they are joined in an end-to-end fashion. Approximately 25% of multidomain proteins, however, appear to have evolved by insertion of one domain into another [2.8-2.10], as we have done in this study. These proteins are subject to the same conformational strain mechanism that is characterized here. Of additional interest are proteins that dimerize or oligomerize *via* a 3D domain-swapping interaction [2.11]. In this scenario, a segment of the polypeptide chain detaches from its binding site in one molecule and docks to the same site in a second molecule. Intrinsic to this process is the concept of conformational strain. It has been proposed that strain within a monomeric protein can drive domain swapping, provided that the strain is relieved upon exchange [2.11-2.14]. Nature may modulate the extent of strain in order to adjust binding affinity while preserving the high specificity dictated by the domain-swapped interface. Our results indicate that the BU variants with long linkers are monomeric, whereas those with short linkers form dimers and higher order oligomers. This finding may help explain how domain-swapping arises during evolution, and guide future design of domain-swapped proteins.

The final goal is to understand the structural basis for how the free energy stored in the native state of one protein is used to unfold another. How is conformational strain distributed throughout the domains? How much can the native states distort without unfolding? What are the structures of mechanically disrupted states? The inability of unfolded proteins to crystallize precludes their structural analysis by x-ray methods. NMR approaches suffer from a related problem: mechanically unfolded proteins may refold as dimers or oligomers via domain-swapping interactions. Indeed, we find that the most strained BU variants form oligomers at

micromolar concentration. Therefore, structural questions regarding monomeric forms of BU variants are best addressed by computational methods. We employ atomistic, unforced Langevin dynamics (LD) simulations to characterize structural changes of the Bn and Ub domains as they exert increasing amounts of strain on each other. To our knowledge, this represents the first simulation of mechanically induced unfolding of one protein domain by another.

2.3 RESULTS

2.3.1 Thermodynamic characterization of linker length variants.

GdnHCl and temperature-induced denaturation curves for BU variants are shown in Figure 2-2. CD ellipticity at 230 nm reveals two GdnHCl-induced transitions (Figure 2-2A). The first transition corresponds to Bn unfolding. This transition is of primary interest, as it is coupled to Ub folding when $\Delta G_X > 0$. The second corresponds to unfolding of the Ub domain when the Bn domain is already unfolded. It can be seen from the primary data that shortening the linkers from 10 Gly to 2 Gly progressively destabilizes the Bn domain and has little effect on the stability of the Ub domain (Figure 2-2A). This finding is consistent with the hypothesis that decreasing linker length increases conformational strain between domains. Strain is abolished at GdnHCl concentrations above the midpoint of denaturation (C_m) of the Bn domain, as evidenced by the common C_m values of the Ub domains.

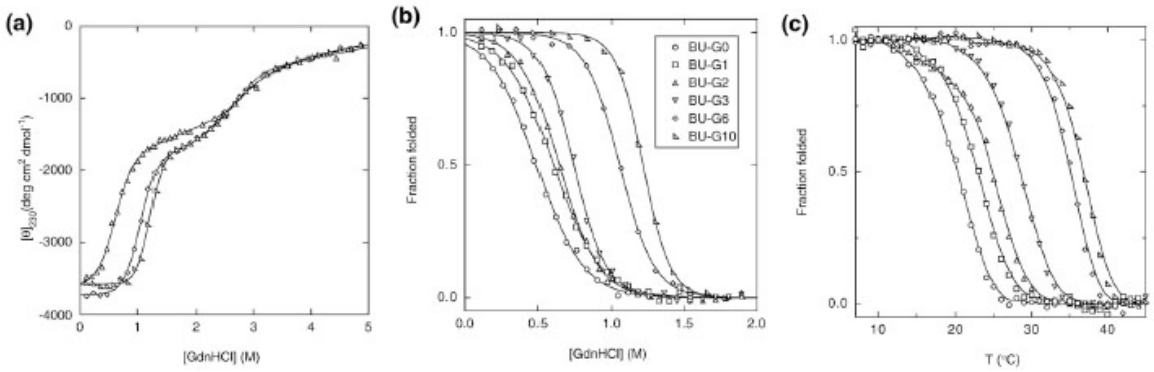

Figure 2-2. Equilibrium denaturation curves of BU variants in the absence of Co^{2+}. (A) GdnHCl-induced denaturation of BU-G10, BU-G6 and BU-G2 monitored by CD ellipticity at 230 nm. Symbols are defined in panel B. Lines are best fits to the three-state linear extrapolation equation. (B) GdnHCl-induced denaturation of BU variants monitored by Trp fluorescence, and normalized to fraction folded. Lines are best fits to the linear extrapolation equation. (C) Thermal denaturation of BU variants monitored by CD ellipticity at 230 nm and normalized to fraction folded, assuming a two-state unfolding reaction. Symbols are the same as in panel B.

To obtain thermodynamic parameters for the coupled folding-unfolding reaction, we characterized the first GdnHCl transition by Trp fluorescence. All three Trp residues are located in the Bn region of BU. The wavelength of maximum emission (F_{max}) increases from 336 nm to 355 nm upon Bn unfolding [2.2]. Thus, F_{max} reports primarily on Bn conformation. Data for all variants are fit adequately by the two-state linear extrapolation equation (Figure 2-2B). The resulting parameters are listed in Table 2-1. BU-G6 is less stable than BU-G10 as judged by both ΔG^{H2O} and C_m values. This result suggests that interdomain strain begins to be exerted with linkers as long as six amino acids each. Consistent with that interpretation, Bn stability continues to decrease as linker length further shortens.

Table 2-1. Thermodynamic parameters for unfolding of free Bn, free Ub, and the Bn domain of BU variants. N.D., not determined. [a]Errors are standard deviations of at least three experiments. [b]Not determined because of high melting temperature. [c]Not determined due to poor fit to a two-state reaction.

Variant	Co^{2+} (mM)	ΔG^{H2O} (kcal/mol)	*M* (kcal mol^{-1} M^{-1})	C_m (M)	T_m (°C)
Free Bn	0	11.5±0.5[a]	5.1±0.2	2.26±0.02	51.5
	10	12.6±1.0	5.5±0.4	2.31±0.01	51.0
Free K6H Ub	0	5.6±0.4	1.7±0.1	3.26±0.10	N.D.[b]
	10	7.0±0.1	1.7±0.02	4.07±0.02	N.D.[b]
BU-G10	0	7.8±0.3	6.4±0.3	1.23±0.02	37.2
	10	7.8±0.5	6.4±0.6	1.22±0.05	35.7
BU-G6	0	5.7±0.2	5.3±0.2	1.08±0.01	35.6
	10	6.1±0.1	6.5±0.1	0.95±0.02	33.0
BU-G3	0	3.8±0.1	5.0±0.3	0.76±0.03	28.9
	10	3.2±0.01	4.9±0.1	0.66±0.01	25.9
BU-G2	0	3.0±0.1	4.1±0.2	0.73±0.01	25.7
	10	2.1±0.04	4.2±0.1	0.50±0.01	21.1
BU-G1	0	2.3±0.3	3.6±0.3	0.64±0.07	23.1
	10	1.9±0.01	2.9±0.05	0.67±0.01	N.D.[c]
BU-G0	0	1.9±0.2	3.7±0.3	0.51±0.02	20.5
	10	2.0±0.1	3.5±0.1	0.56±0.01	N.D.[c]

We monitored thermal denaturation by circular dichroism (CD) to further characterize Bn domain stability (Figure 2-2C). At pH 7.5, free Bn exhibits a melting temperature (T_m) of 51.5 °C (Table 2-1) and free Ub does not denature below 100 °C [2.15]. In the context of the BU protein, the coupled Bn unfolding transition is thus expected to occur below 51.5 °C, whereas the T_m of Ub is predicted to remain above 100 °C. Because ΔC_p and ΔH values are not known, we did not attempt to obtain thermodynamic parameters from thermal denaturation data. Nonetheless, the relative stabilities of Bn domains can be assessed by comparing T_m values. T_m decreases progressively from BU-G10 to BU-G0, in agreement with GdnHCl denaturation results.

2.3.2 Co^{2+} binding experiments.

It is first necessary to determine whether metal binds specifically to the engineered site on Ub. It can be shown that metal-induced stabilization of the Ub domain approaches the limit of $RT \cdot \ln[(1+K_{a1})/(1+K_{a2})]$ when binding to both native and unfolded Ub is saturated, where K_{a1} and K_{a2} are the association constants for the respective forms (Figure 2-1B). Sosnick and co-workers report K_{a1} and K_{a2} values in the range of ~10^4 M^{-1} and ~10^3 M^{-1}, respectively, for Bi-His Ub variants [2.4]. Consistent with those figures, we observe that $\Delta\Delta G^{H2O}$ of K6H Ub reaches a maximum value of 1.4 kcal/mol at Co^{2+} concentrations greater than ~1 mM (Figure 2-3A and Table 2-1). By contrast, 10 mM Co^{2+} has little effect on unfolding of Bn, either in its free state or as a domain in WT BU (data not shown). Both C_m and T_m values of free Bn remain relatively unchanged following addition of Co^{2+} (Table 2-1).

Having established that metal binds to K6H Ub and not to Bn, we next repeated the GdnHCl denaturation experiments of Figure 2-2B in the presence of 10 mM Co^{2+}. The model predicts that Co^{2+} binding will not affect stability of the Bn domain if $\Delta G_X = 0$. In the strong coupling limit, the decrease in Bn stability will reach a maximum value of -1.4 kcal/mol. Figure 2-3B shows that Co^{2+} does little to destabilize the Bn domain of BU-G10. Coupling appears to progress into the intermediate regime with BU-G6, as evidenced by the significant decrease in C_m in the presence of Co^{2+}. The extent of Co^{2+}-induced destabilization increases in BU-G3 and reaches a maximum with BU-G2. Metal binding destabilizes the Bn domain of BU-G2 by 0.9 kcal/mol, or 64 % of the theoretical value for strong coupling. Curiously, Co^{2+} has little effect on unfolding of BU-G1 and appears to increase the C_m of BU-G0 slightly (Figure 2-3C).

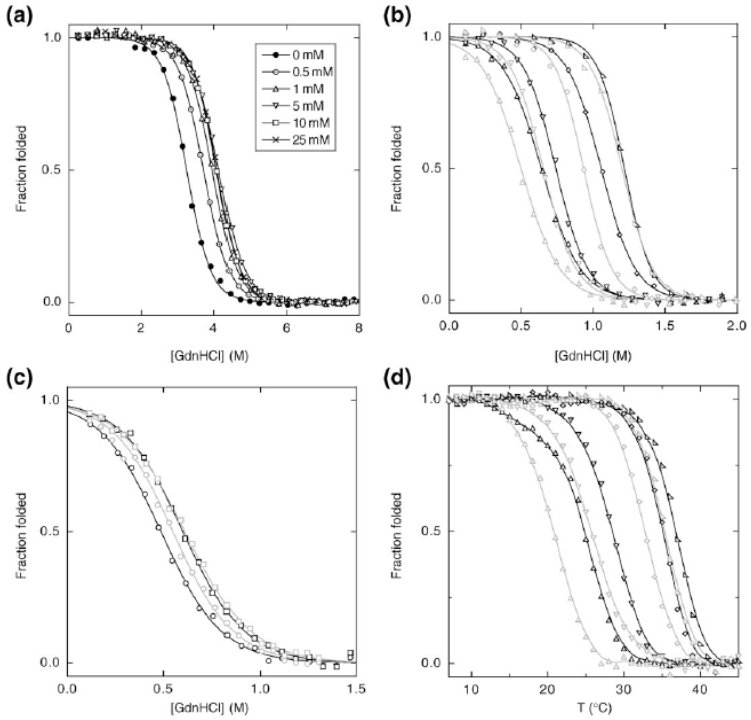

Figure 2-3. Equilibrium denaturation curves of free Ub and BU variants in the presence of Co^{2+}. Symbols are the same as those in Figure 2-2B. (A) GdnHCl-induced unfolding of free K6H Ub monitored by CD ellipticity at 225 nm. Co^{2+} concentrations are indicated in the inset. (B) GdnHCl-induced denaturation of BU-G10, BU-G6, BU-G3 and BU-G2 in the presence (grey) and absence (black) of 10 mM Co^{2+}, monitored by Trp fluorescence. (C) GdnHCl-induced denaturation of BU-G1 and BU-G0 in the presence (grey) and absence (black) of 10 mM Co^{2+}. (D) Thermal denaturation curves of BU variants in the presence (grey) and absence (black) of 10 mM Co^{2+}. Lines in panel A, panel B and panel C are best fits to the linear extrapolation equation.

Thermal denaturation curves (Figure 2-3D) reflect a similar trend. Co^{2+} slightly decreases T_m of BU-G10, suggesting that the two domains are coupled to a small extent even when linked by 10 Gly residues. Co^{2+}-induced decrease in thermal stability becomes more pronounced as linker length decreases. ΔT_m reaches a maximum with BU-G2. In agreement with the GdnHCl results, Co^{2+} does not destabilize the Bn domains of BU-G0 or BU-G1. Rather, it broadens the melting transitions so that they cannot be modeled as a two-state reaction (data not

shown). It is possible that broadening is caused by transient oligomerization, although precipitation is not observed and thermal denaturation is ≥ 80 % reversible in all cases.

2.3.3 Structural characterization by CD.

CD spectra of free Bn, free Ub and BU variants are shown in Figure 2-4A. The CD spectrum of free Bn is characterized by unusually low spectral intensities and an atypical minimum at 231 nm [2.16]. Free Ub exhibits a minimum at 208 nm. BU-G10 displays both of these minima, confirming that both domains are folded. Since ΔG_X is predicted to be greatest for BU-G0 and BU-G1, these variants might be expected to exhibit increased random coil content. However, spectra of BU-G0, BU-G1, BU-G2 and BU-G3 are similar to that of BU-G10. This result indicates that both domains remain folded in all BU variants, despite the presence of conformational strain. Sosnick and co-workers reported that metal binding does not alter the far-UV CD spectrum of bi-His Ub variants [2.4]. Figure 2-4B shows that the spectra of BU variants are similarly unchanged in the presence of Co^{2+}. Metal binding does not appear to perturb the structure of BU.

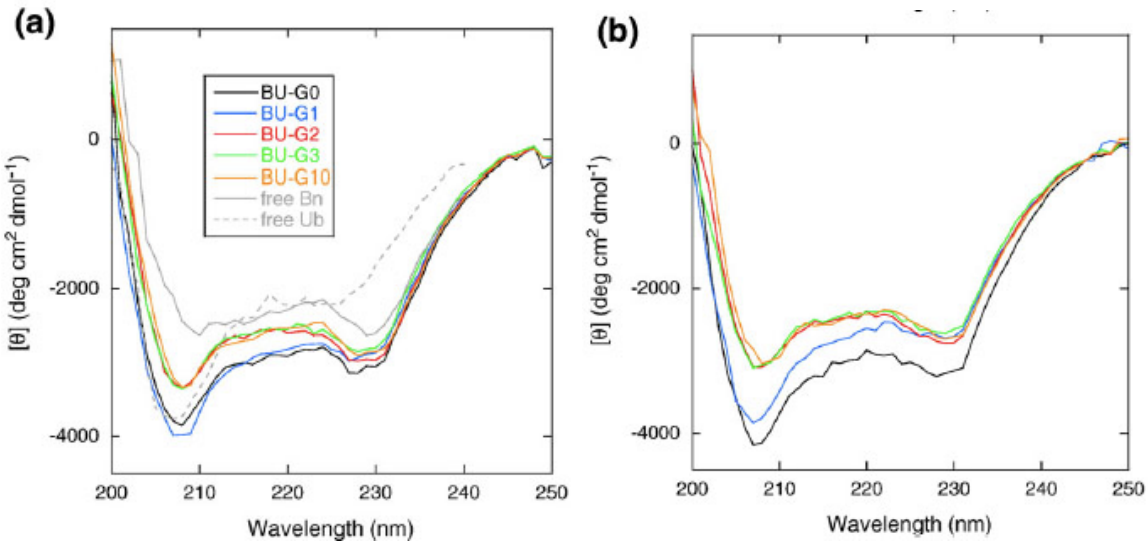

Figure 2-4. Structural characterization of BU variants by CD in the (A) absence and (B) presence of 1 mM Co^{2+}. Protein concentrations are ~5 mM.

2.3.4 Oligomerization of strained variants.

To explain the CD data and the anomalous Co^{2+} binding results obtained for BU-G0 and BU-G1, we considered the possibility that these variants may dimerize or oligomerize. Domain swapping is a logical mechanism for oligomerization. In this scenario, one domain is forced to unfold as ΔG_X progresses into the strong coupling regime. The domain will refold, however, if it can do so in a way that relieves conformational strain in the native state. That condition may be achieved by intermolecular binding and folding of the N and C-terminal fragments of the Bn domain. The simplest structure that would result is a domain-swapped dimer, although higher order oligomers are possible. Both Ub and Bn domains are expected to be folded in the dimeric state, with little if any interdomain strain. Metal binding to the Ub domain is consequently not expected to perturb stability of the Bn domain.

We tested for oligomerization by sedimentation equilibrium ultracentrifugation experiments. All data were fit to a self-associated dimer model and parameters are listed in Table 2-2. Representative data sets, showing quality of the fits, are available as electronic supplementary material. As expected, BU-G10 sediments predominantly as a 21.8 kDa monomer (theoretical MW 22.2 kDa) with a weak dimer-monomer dissociation constant (K_d) of 4.2 mM. BU-G2 also sediments at close to its expected monomeric MW, but K_d decreases by nearly 10^3-fold to 6.0 mM. K_d further decreases (1.3 mM) in BU-G1, and the apparent MW of the monomeric species increases to 35.5 kD. This result may be due to the presence of higher-order oligomers. Consistent with that interpretation, sedimentation profiles for BU-G0 could not be fit to a monomer/dimer model. BU-G0 appears to sediment as a heterogeneous mixture of large MW species (data not shown). Thus, at the protein concentrations employed in fluorescence and CD experiments (1-2 mM), BU-G0 and BU-G1 are predominantly dimeric/oligomeric, while BU-G2 and the longer linker variants are primarily monomeric. The transition from monomer to dimer/oligomer at the BU-G2/BU-G1 linker length is consistent with the anomalous Co^{2+} binding results observed for BU-G0 and BU-G1.

Table 2-2. Apparent monomer molecular weights and dimer-monomer dissociation constants of BU variants, obtained from sedimentation equilibrium experiments. N.D., not determined due to poor fit to dimer-monomer model.

Variant	Apparent molecular mass (kDa)	Theoretical molecular mass (kDa)	K_d (μM)
BU-G0	N.D.	21.1	N.D.
BU-G1	35.5	21.2	1.3
BU-G2	21.6	21.3	6.0
BU-G10	21.8	22.2	4200

2.3.5 Computer simulations.

The goal of the simulations is to characterize the mechanism of mechanical unfolding in conformationally strained BU variants (primarily BU-G0 and BU-G1). There are no existing experimental structures from which starting models can be generated. Indeed, it is unlikely that experimental methods can provide high-resolution structures of BU-G0 or BU-G1 in their strained, monomeric states. Crystallization and NMR conditions would strongly favor formation of oligomers in which strain is relieved. We therefore built starting models from available X-ray structures of free Bn and free Ub. The orientations between domains were explored exhaustively by MODELLER to identify the most favorable orientations based on the free energy function (see Methods). In order to accommodate the insertion, Bn must stretch or Ub must compress. Neither outcome was biased in initial model building; Bn and Ub atoms were restrained to their positions in the respective crystal structures using identical energy functions. MODELLER chose the latter solution. The ~38 Å N-to-C distance in free Ub compresses to 35 Å, 37 Å, 35 Å, 31 Å, 27 Å and 25 Å for BU-G10, BU-G6, BU-G3, BU-G2, BU-G1 and BU-G0, respectively. In contrast, the C_a-C_a distance between the ends of the Bn surface loop (Ser57-Thr70) remains close to that observed in the isolated Bn crystal structure (~11 Å).

To characterize the structures of mechanically-disrupted states, the starting model of each BU variant was subjected to 50 ns of LD simulations at 328 K. Simulations were also performed on isolated Bn (Bn_{cut}) and isolated Ub (Ub_{cut}) proteins as controls. Bn_{cut} and Ub_{cut} were generated by "cutting" them out of the starting model of BU-G0 and capping the new N- and C-termini with acetyl and N-methyl groups, respectively. Finally, we performed a 50 ns simulation of BU-G0 with the inhibitor barstar bound to Bn. Since barstar binds to barnase extremely tightly

($K_d = 10^{-14}$ M [2.17, 2.18], the prediction is that barstar binding will stabilize the Bn domain and drive unfolding of the Ub domain.

To monitor structural changes during each 50 ns simulation, we determined C_a root-mean square deviations (RMSD) of each domain. RMSD is the deviation between one structure in the simulation relative to the energy-minimized starting structure, and reveals the extent of unfolding at a given time. Both Bn_{cut} and Ub_{cut} remain folded as demonstrated by small and relatively constant values of RMSD (4.2 ± 0.7 Å and 2.7 ± 0.5 Å, respectively, over the last 25 ns). The degree of unfolding of one or both domains of BU variants, as monitored by RMSD, increases as linker length decreases (Figure 2-5A). Starting with BU-G10 and progressing to BU-G2, moderately elevated RMSDs (relative to Bn_{cut} and Ub_{cut} controls) are observed in either the Bn domain or the Ub domain, but not in both. BU-G6 appears to sample two substates, one before 35 ns, in which both domains are folded, and the other, after 35 ns, in which the Ub domain exhibits significant deviations. BU-G2 undergoes moderately elevated deviations in the Bn domain; the Ub domain remains folded. These data are consistent with strain being present but small enough in magnitude that it can be dissipated within one domain.

In contrast, RMSDs of both domains are significantly increased in BU-G1 (5.2 ± 1.0 Å in Bn, 4.9 ± 0.5 Å in Ub) and BU-G0 (9.9 ± 1.7 Å in Bn, 6.2 ± 0.8 Å in Ub). This result is consistent with the experimental data, and suggest that strain in these two variants is so great that it cannot be contained within a single domain. When barstar is bound to BU-G0, both domains remain folded. Although no Ub unfolding is evident, the fact that the deviations of Bn are lower than those of Bn_{cut} indicates that barstar binding stabilizes Bn significantly and prevents Bn from unfolding.

To characterize structural changes in greater detail, we monitored RMSDs of the hydrophobic cores of Bn and Ub over the entire 50 ns of simulation. Ub contains a single core (Ile3, Phe4, Val5, Ile13, Leu15, Val17, Ile23, Val26, Ile30, Ile36, Leu43, Phe45, Leu50, Leu56, Ile61, Leu67, Leu71) whereas Bn contains three: $core_1$ (Phe7, Val10, Ala11, Leu14, Leu20, Tyr24, Ala74, Ile76, Ile88, Tyr90, Trp94, Ile96, Ile109), $core_2$ (Ile25, Ala30, Leu33, Trp35, Leu42, Val45, Ile51), and $core_3$ (Phe56, Leu63, Trp71, Leu89, Leu95, Tyr97, Tyr103, Phe106) [2.19]. All of the hydrophobic cores in Bn_{cut} and Ub_{cut} remain close to the starting structures except for Bn $core_2$, which shows slightly elevated RMSDs (Figure 2-5B; Table 2-2). Barstar binding reduces RMSDs of all cores to 3.0 Å or less. RMSDs of $core_1$, $core_2$ and the Ub core generally increase as linker length is shortened from 10 Gly to zero Gly (Table 2-2). Comparing panel C with panel B in Figure 2-5 reveals the extent of unfolding of BU-G2, the most strained monomeric variant, relative to the Bn_{cut} and Ub_{cut} controls. Core3 exhibits large RMSDs, partially unfolding while the remaining hydrophobic cores remain intact (Table 2-3).

In addition to RMSD, it is informative to calculate root-mean square fluctuation (RMSF), which is the RMSD between one structure in the simulation relative to the average structure of the simulation. The intensities of the fluctuations provide information on the degree of flexibility in the structure. The core of Ub_{cut} as well as $core_1$ and $core_3$ of Bn_{cut} are relatively rigid with all-atom RMSFs of 2.0 ± 0.3 Å , 1.9 ± 0.6 Å and 1.7 ± 0.9 Å, respectively, averaged over the entire 50 ns. $Core_2$ appears to be more flexible (RMSF = 3.2 ± 1.0 Å). Table 2-3 summarizes the percentage of time that the hydrophobic cores spend in large fluctuations during the simulations. Fluctuations are defined as large if the all-atom RMSF is greater than three standard deviations above the average RMSF of the corresponding cores of Bn_{cut} and Ub_{cut}. In general, the extent of large fluctuations increases as the linkers are shortened. This trend is

particularly pronounced in BU-G0, in which $core_1$ and the Ub core undergo large fluctuations 97 % and 70 % of the time, respectively.

Table 2-3. Average C_α RMSDs of hydrophobic cores of Bn and Ub domains obtained from LD simulations, relative to the respective starting structures. RMSD values are averaged from 5000 conformations sampled every 10 ps of the 50-ns simulation. Numbers in parentheses are the percentages of time that the hydrophobic cores spend undergoing large fluctuations during the simulations. Fluctuations are defined as large if the all-atom RMSF is greater than three standard deviations above the average RMSF of the corresponding core of Bn_{cut} and Ub_{cut}. N.A., not applicable.

Variant	Bn $core_1$ RMSD (Å)	Bn $core_2$ RMSD (Å)	Bn $core_3$ RMSD (Å)	Ub core RMSD (Å)
Bn_{cut}	2.8±0.5 (0)	3.9±1.2 (0)	3.0±0.5 (0)	N.A.
Ub_{cut}	N.A.	N.A.	N.A.	2.9±0.3 (0)
BU-G0/barstar	2.1±0.3 (0)	2.0±0.2 (0)	3.0±0.4 (0)	2.6±0.2 (0)
BU-G10	2.2±0.3 (0)	2.5±1.7 (3)	4.4±1.7 (4)	2.8±0.3 (0)
BU-G6	2.6±0.2 (0)	2.5±0.9 (0)	3.6±1.1 (6)	3.9±2.0 (36)
BU-G3	2.4±0.2 (0)	3.5±1.9 (1)	2.4±0.3 (0)	3.2±0.3 (0)
BU-G2	2.8±0.7 (0)	5.3±0.8 (1)	5.3±2.2 (22)	2.8±0.3 (0)
BU-G1	2.5±0.3 (0)	6.1±1.7 (5)	4.0±1.1 (0)	4.6±1.3 (70)
BU-G0	9.8±4.9 (97)	5.0±1.2 (0)	3.7±1.0 (0)	5.7±1.5 (70)

We conducted a detailed analysis of the unfolding mechanism for BU-G2, the most strained variant that is not found experimentally to be dimeric or oligomeric. Figure 2-6 shows snapshots of BU-G2 taken at various times of simulation. Unfolding of BU-G2 involves the Ub domain pulling Bn residues on the N-terminal side of the insertion. These Bn residues include Leu63 and Phe56, two of the eight residues belonging to $core_3$. The first sign of unfolding occurs at 17 ns when the Ub domain pulls Bn residue Leu63 away from the center of $core_3$. This dramatic disruption of the $core_3$ is reflected by a sharp increase in its RMSD from 2.6 Å at 14 ns to 7.1 Å at 17 ns (Figure 2-5C). At 34 ns, the Ub domain continues to pull on the Bn domain, exposing Phe56 to solution. At the end of 50 ns of simulation, the $core_3$ region of Bn continues to climb in RMSD, indicating that longer simulation times may reveal further unfolding of the Bn

domain. On the whole, unfolding appears to be localized to regions of Bn near the Ub insertion site. Other regions of both protein domains are relatively undisturbed. $Core_1$ experiences slight fluctuations (RMSF=2.1±0.5 Å) that are intrinsic to that core as evident from the Bn_{cut} simulation (RMSF=1.9±0.6 Å). $Core_2$ experiences moderate fluctuations (RMSF=3.3±1.1 Å), although considerable fluctuations are also observed in this core during the Bn_{cut} simulation (RMSF=3.2±1.0 Å). The hydrophobic core of Ub also exhibits slight fluctuations (RMSF=1.8±0.3 Å) that are comparable to those observed in the Ub_{cut} simulation (RMSF=2.0±0.3 Å).

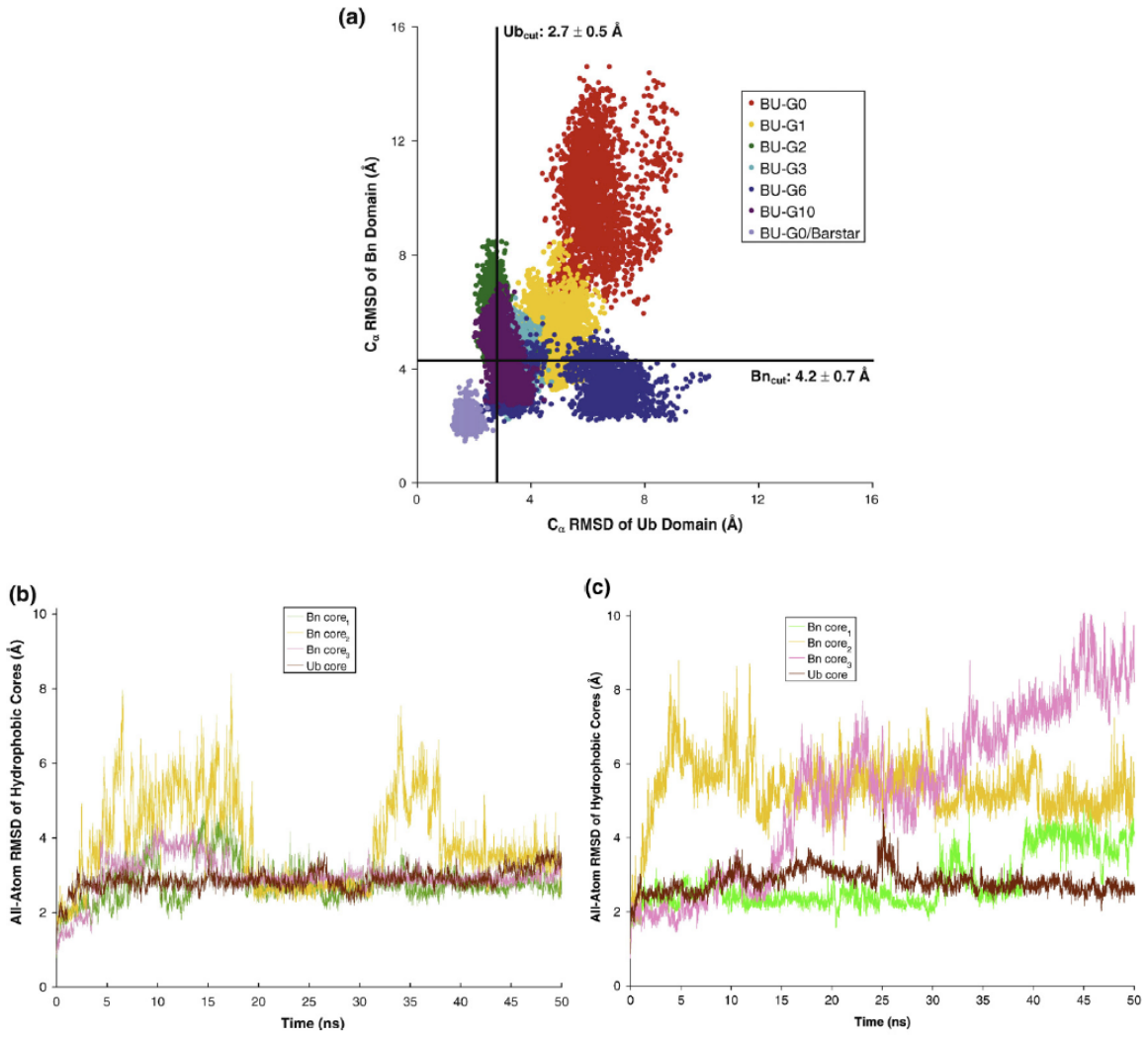

Figure 2-5. RMSDs of structures obtained from LD simulations, relative to energy-minimized starting models. (A) C_α RMSDs of Bn and Ub domains of BU variants and the BU-G0/barstar complex. For clarity, only data points corresponding to the last 25 ns of the 50 ns simulation are shown (2500 conformations sampled every 10 ps). RMSDs include all amino acids except for those of the Bn surface loop (residues 65-69) that were left unrestrained during the generation of starting models (see Methods). Horizontal and vertical lines indicate C_α RMSDs of Bn_{cut} and Ub_{cut}, respectively, averaged over the last 25 ns of the 50 ns simulations. (B) All-atom RMSDs of the hydrophobic cores of Bn_{cut} and Ub_{cut} as a function of time. (C) All-atom RMSDs of the hydrophobic cores in BU-G2 as a function of time.

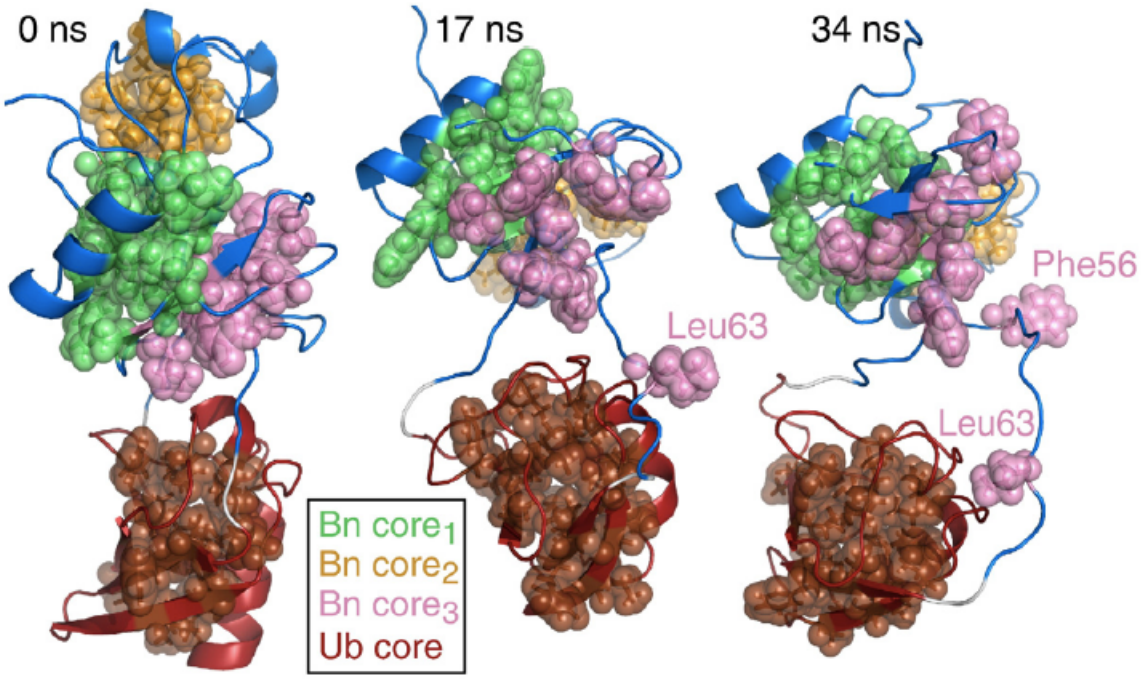

Figure 2-6. Snapshots of BU-G2 taken at indicated times of simulation, illustrating the sequence of mechanically disrupted states.

2.4 DISCUSSION

2.4.1 Thermodynamic model for antagonistic coupling.

Our strategy for testing the mechanism of Figure 2-1 is to vary the length of the linkers and measure the coupling free energy ΔG_X. The hypothesis is that shortening the linkers will increase interdomain strain, and when the length decreases below a critical threshold, ΔG_X will exceed ΔG_B^{H2O} and ΔG_U^{H2O} and mutually exclusive folding will be attained. ΔG_X is estimated from ΔG^{H2O}, *m*, and C_m values of the unfolding transitions of the Bn domain, in the absence and

presence of Co^{2+}. This approach is illustrated below using the three coupling regimes as examples.

Zero coupling. When $\Delta G_X = 0$, the Bn unfolding transition is $\underline{bu} \leftrightarrows b\underline{u}$ (Figure 2-1). The observed *m*-value is theoretically identical to that of free Bn. The observed ΔG^{H2O} value corresponds to the stability of Bn with Gly residues inserted into the surface loop. Metal binding to the Ub domain will not affect thermodynamic parameters of Bn unfolding.

Strong coupling. Here, Bn unfolding is linked to Ub folding and the Bn transition becomes $\underline{bu} \leftrightarrows b\underline{u}$. The apparent ΔG^{H2O} and *m*-values decrease to ($\Delta G_B{}^{H2O} - \Delta G_U{}^{H2O}$) and ($m_B - m_U$), respectively. Metal binding to Ub now has a pronounced effect on Bn stability. Stabilization of the Ub domain approaches the limit of $RT \cdot \ln[(1+K_{a1})/(1+K_{a2})]$ when metal binding to both native and unfolded states is saturated. Co^{2+} binding to Ub will thus destabilize Bn by 1.4 kcal/mol in the strong coupling condition.

Intermediate coupling. As ΔG_X increases from zero, the Bn unfolding transition gradually shifts from $\underline{bu} \leftrightarrows b\underline{u}$ to $\underline{bu} \leftrightarrows b\underline{u}$. Both transitions occur, although they are not resolved by Trp fluorescence and the experimental data are fit adequately by the two-state model (Figure 2-2B). It is therefore appropriate to compare ΔG^{H2O} and *m*-values only when coupling is in the zero or strong regimes. Nevertheless, increasing the value of ΔG_X clearly shifts the Bn denaturation curves to lower GdnHCl concentrations. ΔG_X can consequently be estimated from the extent to which Co^{2+} binding shifts Bn C_m values.

2.4.2 Effect of linker length on coupling.

Peptide linkers of 10 Gly each effectively decouples folding of the Bn and Ub domains. The Bn domain of BU-G10 exhibits an *m*-value similar to that of free Bn, and Co^{2+} binding does not

appreciably reduce its stability (Table 2-1). Comparison of GdnHCl denaturation curves for BU-G10, BU-G6, BU-G2 and BU-G2 (Figure 2-2B) reveals that both C_m and *m*-values decrease with linker length. Significant interdomain strain is present in BU-G6 and it intensifies as the linkers shorten. The decrease in *m*-value suggests that ΔG_X approaches the strong coupling limit in BU-G2. The theoretical *m*-value for strong coupling ($m_B - m_U$) is 3.4 kcal/mol/M or 4.7 kcal/mol/M (depending on whether m_B is obtained from free Bn or from the Bn domain of BU-G10). The *m*-value observed for BU-G2 is 4.1 kcal/mol/M. These findings agree with those of our earlier study, in which coupling was found to be in the high-intermediate range (ΔG_X ~4 kcal/mol) for a variant with linkers of two and three amino acids.

As ΔG_X increases, the free energy of metal binding to the Ub domain is transduced into destabilization of the Bn domain (Figure 2-3B). 10 mM Co^{2+} shifts C_m by –0.01 M (BU-G10), -0.13 M (BU-G6), –0.10 M (BU-G3) and –0.23 M (BU-G2). Destabilization is greatest for BU-G2, where the $\Delta\Delta G^{H2O}$ value of -0.9 kcal/mol is 64 % of the theoretical value for strong coupling. Thermal denaturation curves follow a similar profile. 10 mM Co^{2+} decreases T_m by -1.5 °C (BU-G10), –2.6 °C (BU-G6), –3.0 °C (BU-G3) and –4.6 °C (BU-G2). In summary, results obtained for BU-G10 to BU-G2 provide evidence for the mechanism of Figure 2-1.

2.4.3 Strong coupling induces oligomerization.

The above data imply that the Bn and Ub domains are strained in BU-G2, and that further shortening the linkers will cause one of the domains to unfold the other. Puzzlingly, Co^{2+} binding does not destabilize the Bn domains of BU-G1 and BU-G0 (Figure 2-3B), which would appear to indicate that coupling is close to zero in these variants. An explanation is provided by the observation that K_d of dimerization is strongly dependent on linker length (Table 2-2). BU-

G0 and BU-G1 form dimers and oligomers under experimental conditions whereas BU-G2 and the longer linker variants do not. These results together suggest that BU-G1 and BU-G0 are in fact under severe strain—strong enough so that only one domain can fold within a single molecule—but strain is relieved by intermolecular folding.

Domain swapping provides a straightforward mechanism for intermolecular folding with ample precedence. In particular, conformational strain may be a general mechanism by which the equilibrium between swapped and non-swapped forms can be modulated. Baker and colleagues induced protein L to form a swapped dimer by mutating three residues in a β-turn [2.12]. These mutations selectively destabilize the monomer by forcing turn residues to adopt unfavorable backbone dihedral angles. Gronenborn et al. induced dimerization of GB1 by positioning a bulky Phe residue in its hydrophobic core [2.21]. Steric repulsions were relieved by domain swapping. Cordes and co-workers took the opposite approach and converted dimeric l Cro protein to monomer by selectively stabilizing the latter [2.22]. Honig, Shapiro and colleagues proposed that cadherins use domain swapping to create homophilic binding interfaces that are highly specific, yet of low affinity [2.23]. Binding affinity is decreased because the docking sites for the swapped segment are similar in both the monomer and the dimer. The resulting competition between intra- and intermolecular folding lowers the free energy of complex formation. Binding affinity can in principle be tuned by modulating stability of the monomeric form. Consistent with this view, a recent X-ray structure of type I cadherin finds that the hinge region adopts a strained conformation in the monomer [2.24].

Mutually exclusive folding is a novel mechanism for introducing extreme conformational strain into a protein. Stress is so great in the monomer that one domain must unfold. It can only refold by unfolding the other domain, which is thermodynamically uphill, or by exchanging

segments with another molecule to form dimers or oligomers. With respect to the latter process, it is not clear what structural features determine whether two complementary fragments can associate to reconstitute the native fold. Bn, however, is known to be particularly amenable to fragment refolding. Native-like complexes form upon pairwise mixing of peptides consisting of residues (1-36, 37-110), (1-56, 57-110), (1-68, 69-110) and (1-79, 80-110) [2.25]. The (1-68, 69-110) peptides correspond most closely to the Ub insertion point used in the present study. These fragments produced the most stable complex of the four tested [2.25], which may explain the propensity for our BU construct to form domain-swapped dimers. In this situation, amino acids in the Bn surface loop could act as hinges which allow the C-terminal 67-110 fragment of Bn to dock to and refold with the N-terminal 1-66 portion of a second Bn molecule. Ub can remain folded during this process. If a reciprocal event occurs with the remaining unpaired fragments of Bn, the resulting structure is a closed, symmetrical dimer in which all domains are folded and interdomain strain may be minimal or absent.

The domain swapping hypothesis remains to be tested by structural experiments; nevertheless, it is supported by several previous Bn studies. We found evidence for a domain-swapped dimer in our earlier study of the barnase-GCN4 fusion protein [2.1]. Wild-type Bn forms a domain-swapped trimer at high protein concentrations and moderately destabilizing conditions [2.26]. Numerous pairs of Bn fragments, generated by cleaving the protein at various surface loops, can bind and refold to form native-like complexes [2.25]. Thus, domain swapping is the explanation that best accounts for all of our present observations.

2.4.4 Structures of mechanically disrupted states.

In order to observe unfolding within a 50 ns LD trajectory, which requires ~1 month for a single BU variant, it was necessary to reduce solvent viscosity and employ a generalized Born (GB) implicit solvation model. Low solvent viscosity has been employed in simulations of protein folding [2.27] and opening/closing motions of the flaps in HIV protease [2.28, 2.29]. In such cases, reducing viscosity accelerates reaction rates but does not significantly change the structures that are populated [2.30]. Implicit solvation is required in order to enable simulation at low viscosity. It is important to recognize that GB models lack some features of explicit water models, and this difference can lead to artifacts such as the tendency to overstabilize α-helices [2.31-2.35] and ion pair interactions [2.34, 2.36-2.39]. It was also necessary to increase the simulation temperature to 55 °C to facilitate unfolding within 50 ns. No unfolding was detected at 40 °C for even the most strained variant (BU-G0). Since T_m of free Bn is 51.5 °C, it might be expected that the Bn domains of all variants would denature in the 55 °C simulations. However, the temperature scales of GB models have been reported to be elevated by 50-100 °C [2.31, 2.34, 2.40]. In fact, Bn_{cut} (and Ub_{cut}) remain folded during the entire 50 ns simulation at 55 °C. The unfolding that we observe in each domain is therefore caused by the presence of the other. This unfolding is likely due at least in part to mechanical strain.

Our simulation of the most strained monomeric variant, BU-G2, provides insight into a potential route to unfolding by mechanical disruption. Conformational strain causes the Bn domain to sample partially unfolded conformations (Figure 2-6). Because Ub is more thermostable than Bn, it is expected that the Ub domain will unfold the Bn domain at the temperature of the simulations. The BU-G2 simulation suggests that the Bn residues on the N-

terminal side of the Ub insertion constitute a weak point in the Bn structure. Unfolding of the Bn domain observed in this simulation is consistent with the mutually exclusive folding hypothesis, which holds that the guest Ub domain exerts a stretching force on the host Bn domain.

2.5 CONCLUSIONS

We have demonstrated that the extent of interdomain coupling between folding and unfolding of two linked proteins is inversely proportional to the length of the peptides used to join them. When the linkers are less than two amino acids each, ΔG_X exceeds the stability of the individual domains and one is forced to unfold. The partially unfolded protein appears to undergo a domain-swap which relieves conformational strain and allows it to refold as a dimer or oligomer. Our study provides guidance for future designs of molecular switches based on the mutually exclusive folding mechanism. It also suggests that domain insertion may be an effective means for creating protein binding interfaces via domain swapping. The affinity of such an interaction can in principle be adjusted by modulating the extent to which one domain destabilizes the other in the monomeric protein (*i.e.* ΔG_X). ΔG_X can be controlled by varying the lengths of the linker peptides.

2.6 MATERIALS AND METHODS

2.6.1 Nomenclature, construction and purification of BU variants.

BU variants were constructed by inserting the human K6H Ub gene between the codons for Lys66 and Ser67 of WT Bn. Symmetrical linkers of 0, 1, 2, 3, 6 and 10 Gly residues were used to join the two proteins. BU-G10 refers to Bn and K6H Ub joined by two linkers of 10 Gly each, and so on. All sequences were verified by DNA sequencing. Proteins were co-expressed in *E. coli* with the inhibitor barstar to prevent cell toxicity. BU was purified in the presence of 9 M urea to dissociate the Bn-barstar complex and was subsequently dialyzed to remove urea [2.3]. Due to variations in refolding conditions that inevitably occur during dialysis, the purified proteins exhibited different ratios of monomer to oligomer. To ensure uniformity of samples used in experiments, we prepared all samples by dissolving the lyophilized protein in 6 M GdnHCl. This procedure disrupts any oligomers that may have formed during purification. Proteins were then refolded by rapid dilution (typically 50-100 fold) into ice-cold buffer and allowed to equilibrate for >12 h at 4 °C.

2.6.2 CD and fluorescence experiments.

All GdnHCl denaturation experiments were carried out at 10 °C. Samples were prepared by a 50-fold dilution of the protein in 6 M GdnHCl into aliquots of ice-cold 10 mM Tris (pH 7.5), 0.1 M NaCl containing various amounts of GdnHCl. Those aliquots were prepared previously by mixing different ratios of buffer and 6 M GdnHCl using a Hamilton 500 diluter. For Co^{2+} binding experiments, 10 mM $CoCl_2$ was added to each stock solution prior to mixing. Samples

were then incubated at 4 °C for 12-15 h and transferred to a 10 °C bath for at least 2 h prior to collecting data. Final protein concentration was 1-2 mM. GdnHCl concentrations were measured at the end of each experiment by index of refraction [2.41]. Trp fluorescence data were recorded on a Horiba Jobin Yvon Fluoromax-3 fluorometer. The Datamax software package (Horiba Jobin Yvon) was used to calculate the wavelength of maximum fluorescence emission (F_{max}). CD data were collected on a Aviv Model 202 spectropolarimeter. Data were fit to the linear extrapolation equation [2.42].

CD wavelength scans were recorded at 10 °C using 5 mM protein in a 2 mm path length cuvette. To reduce absorbance arising from residual denaturant, samples were prepared by dissolving protein in 6 M ultrapure urea instead of 6 M GdnHCl, and were refolded by 150-fold dilution into 2 mM Tris (pH 7.5).

Thermal denaturation experiments were performed using a heating rate of 1 °C/m and monitoring CD ellipticity at 230 nm. Data were fit to a two-state unfolding mechanism, with linear corrections applied to the baseline slopes. Reversibility was >80 % for all variants.

2.6.3 Analytical ultracentrifugation experiments.

Lyophilized protein was dissolved in 6 M urea, 10 mM Tris (pH 7.5) at concentrations of ~700 mM, ~350 mM and ~175 mM. Proteins were then refolded by 70-fold dilution into ice-cold 2 mM Tris (pH 7.5). Sedimentation equilibrium experiments were performed at 10 °C using a Beckman XL-A protein characterization system equipped with an eight-cell AN50-Ti rotor. Absorbance was monitored at 225 nm using a six-channel Epon 12 mm centerpiece. Data were collected at 30,000, 35,000 and 40,000 rpm with 12 h equilibration time followed by 1 h acquisition time. Data were processed and analyzed using the SEDFIT/SEDPHAT software

package (http://www.analyticalultracentrifugation.com). Monomer molecular weight and K_d were obtained from a global fit of the data from three different protein concentrations to a self-associating monomer-dimer model.

2.6.4 Structural models of strained BU variants.

Heavy-atom models of each strained BU variant, in which both domains are folded, were generated by satisfaction of spatial restraints taken from the crystal structures of Bn (PDB code 1A2P) [2.43] and Ub (PDB code 1UBQ) [2.44], using the MODELLER 9v1 software package [2.45, 2.46]. To prevent the generation of "knots" in the structures, residues 65-69 of the Bn surface loop were left unrestrained in the MODELLER calculation. All other atoms in both domains were restrained using the same energy function. Hydrogen atoms were added using the LEAP module in AMBER 9 [2.47]. To determine the optimal relative orientation between the domains in each model, we first generated a total of 360 orientations by varying a defined angle (C_a atoms of Bn residues 71 and 79, and Ub residue 4) and torsion between the domains (C_a atoms of Bn residues 17 and 74, and Ub residues 1 and 8) in 10° increments from 90-180° and 10-360°, respectively, and including the angle and torsion for each orientation as additional restraints in the MODELLER calculation. Ten models were randomly generated for each orientation. The model with the lowest value of the MODELLER objective function was then protonated and subjected to energy minimization. To rank the orientations, MM-GBSA free energy calculations [2.48, 2.49] were performed on the minimized model using the AMBER ff99SB force field [2.50], a variation of the GB implicit solvent model by Onufriev et al. (igb = 5 in AMBER 9) [2.51], and no cutoff for the evaluation of nonbonded interactions. Results from these calculations revealed the same major minimum for each BU variant. The orientation near

the center of the well (150° angle/50° torsion) was selected as an optimal orientation. Heavy atoms for the K6H mutation were positioned using the SCAP side-chain prediction program in the Jackal 1.5 software package [2.52]. To generate the starting model of the BU-G0/barstar complex, barstar was "docked" into the binding site of Bn in the energy-minimized BU-G0 model by superimposing all C_α atoms of Bn, with the exception of residues 65-69 in the surface loop, from the crystal structure of the Bn-barstar complex (PDB code 1BRS) [2.53] onto those of Bn in BU-G0 model.

2.6.5 Computer simulations.

Simulations starting from models of the strained BU variants were performed using the AMBER 9.0 software package [2.47]. The force field and implicit solvent model employed were the same as those specified for the MM-GBSA calculations. To facilitate dynamic events so that mechanically induced unfolding occurs on a computationally feasible timescale, we performed LD simulations with a reduced solvent viscosity (collision frequency of 1 ps^{-1}). These events were further accelerated by simulating at a moderately elevated temperature of 328K, using the solvent dielectric constant of water at that temperature (ε = 68.3) [2.54]. To enable a 2 fs time step, bonds to hydrogen were constrained to their equilibrium values using the SHAKE algorithm [2.55]. All nonbonded interactions were evaluated at each time step. The system was initially subjected to energy minimization followed by three stages of equilibration each lasting 50 ps. During the first stage, the energy-minimized system was gradually heated from 0 K to the target temperature of 328 K. Positional restraints were applied to all atoms for the first three stages of equilibration, with force constants of 2.0 kcal/mol·$Å^2$, 1.0 kcal/mol·$Å^2$, and 0.1

kcal/mol·Å^2, respectively. After equilibration, fully unrestrained production simulations were carried out for 50 ns at 328 K. Each 50 ns simulation required approximately one month on a dual 2.66 GHz quad-core processor server, using all eight cores in parallel.

2.7 ACKNOWLEDGEMENTS

We thank M. Cosgrove for use of his analytical ultracentrifuge and B. Honig for discussions regarding domain swapping. This work was supported by NIH grant GM069755 to S. N. L.

3.0 COARSE-GRAINED SIMULATIONS OF PROTEIN UNFOLDING DRIVEN BY THE FOLDING OF ANOTHER PROTEIN IN BARNASE-UBIQUITIN FUSION PROTEINS

3.1 ABSTRACT

The barnase-ubiquitin fusion protein has previously been shown to act as a molecular switch in CD spectroscopy experiments [3.1], with corroboration from atomistic computer simulations [3.2]. To probe the folding-unfolding equilibrium that exists in the barnase-ubiquitin molecular switch on longer timescales, we employed coarse-grained Brownian dynamics (BD) simulations of barnase-ubiquitin fusion proteins. These simulations, when performed at the melting temperature of barnase (51.5 °C), show that the folding of ubiquitin greatly shifts the barnase folding-unfolding equilibrium towards the unfolded state. Artificially weakening the strength of residue-residue interactions within the ubiquitin domain results in its rapid unfolding; when the parameters are restored to their original value, ubiquitin refolds and drives the unfolding of barnase. The folding-unfolding equilibrium is sensitive to the length of interdomain linker peptides; furthermore, the binding of the barnase inhibitor, barstar, has been shown to greatly shift the barnase domain folding equilibrium towards folding while shifting the ubiquitin domain folding equilibrium towards unfolding.

3.2 INTRODUCTION

In the past decade, many molecular switch design schema have emerged and shown promise towards the development of new molecules with novel and highly specific signal-response couplings [3.3-3.7]. Of these various design methodologies, those that take advantage of domain insertion are of particular interest, for they possess two immediate advantages: they use proteins as interchangeable units, and the nature of their fusion generally results in stronger coupling between the domains than is afforded by an end-to-end linkage of the domains [3.1, 3.3, 3.4, 3.8-3.13]. While most domain insertion methods attempt to only slightly perturb the structures of the two proteins, in the mutually exclusive folding paradigm the folding of one domain drives the unfolding of the other [3.1]. In the barnase-ubiquitin and barnase-GCN4 fusion proteins, the folding-unfolding equilibrium has been shown to shift due to changes in temperature or binding to an effector molecule [3.1, 3.14, 3.15]. These results suggest that mutually exclusive folding may be a general and effective paradigm towards the design of new molecular switches.

However, mechanistic and structural information from laboratory experiments about the conformations and transitions undergone by these molecular switches is very limited due to the partially unstructured nature of the fusion proteins [3.2]. Physics-based computer simulations provide an alternate approach, giving a detailed view of the adopted conformations and transition mechanisms of mutually exclusive folders.

The goal of this study is to utilize coarse-grained simulations in order to characterize the folding-unfolding equilibrium of domains in barnase-ubiquitin fusion proteins, which have been shown in CD spectroscopy experiments and atomistic simulations to behave as molecular switches [3.2]. These switches were designed using the mutually-exclusive folding paradigm, in which a protein with a large distance between its amino and carboxyl ends is inserted into

another protein at a surface loop with a much smaller distance between the ends of the loop [3.1]. Figure 3-1 shows that this criterion is satisfied when ubiquitin is inserted between residues Lys66 and Ser67 on barnase. As a result of this design criterion, the mechanical strain imposed on each domain by the other's folded structure will result in a thermodynamic tug-of-war between the two domains. If the strain is greater than the intrinsic stability of one domain, the domain will unfold in order to relieve the strain. Since the intrinsic stability of each domain is dependent on external conditions such as temperature and the binding of effector molecules, a change in these conditions can produce a switching event, in which the unfolded domain refolds and causes the other domain to unfold.

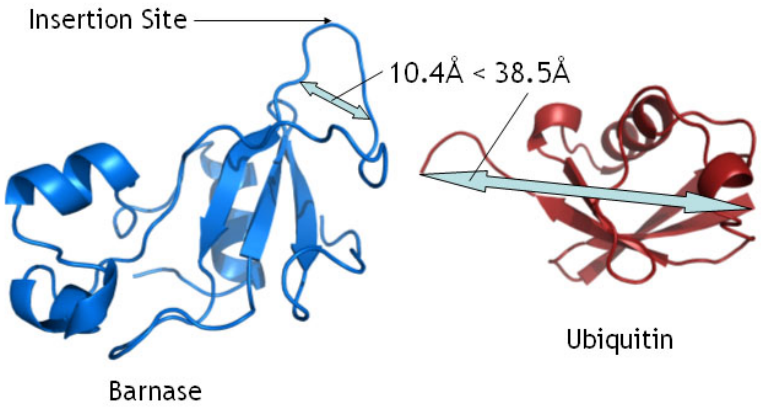

Figure 3-1. Representations of the proteins barnase and ubiquitin. The large amino to carboxyl end distance of ubiquitin compared to the barnase surface loop distance fit the mutually exclusive folding design criterion. Images were generated using the PyMOL program (DeLano Scientific).

Since molecular switches that are engineered according to the mutually exclusive folding paradigm are (by design) partially unstructured, experimental methods of structure characterization such as x-ray crystallography and NMR spectroscopy would be difficult to implement. Therefore, we previously took the approach of combining computer simulations with CD spectroscopy experiments in order to probe conformations adopted by barnase-ubiquitin molecular switches [3.2]. Atomistic simulations were conducted on a series of barnase-ubiquitin

fusion proteins in order to gauge how the interdomain linker length affected the degree of unfolding in each domain [3.2]. These simulations showed that the fusion proteins with short linker peptides became partially unfolded, while little unfolding could be observed for the variants with long linker peptides. Partial unfolding predominantly occurred in the barnase domain.

For this study we have employed coarse-grained simulations in order to more extensively probe the folding-unfolding equilibrium of barnase-ubiquitin fusion proteins. Coarse-grained simulations, due to their low computational expense, have the advantage of allowing for greater conformational sampling, the exploration of more extensive timescales, and the possibility of carrying out multiple simulations. These simulations allow for the observation of numerous complete protein unfolding and refolding events.

3.3 METHODS

3.3.1 The protein model.

The structure and energetics of the protein models are governed by a Gō-type model [3.16] similar to that described by Elcock, Clementi, Onuchic, and others [3.17-3.22]. The protein is reduced from atomistic detail to residue-level detail by using single pseudoatoms, placed at the position of the C_α atom in the full-atom model, to represent individual residues. For folded structures of barnase, ubiquitin, and barstar, coordinates for the C_α beads come from those of the C_α atoms found in the corresponding crystal structures [PDB code 1A2P, 1UBQ, and 1BRS, respectively] [3.23-3.25]. C_α-bead models for the barnase-ubiquitin fusion protein variants BU-

G2, BU-G6, and BU-G10 with both domains folded were generated from atomistic models generated as described in [3.2]. These variants are named for the number of glycine linkers that are placed between the domains on each side of inserted ubiquitin. Random coil structures for the individual proteins and BU-G2 were generated using a statistical coil model implemented by the Unfolded State Server (U. Chicago, http://godzilla.uchicago.edu/cgi-bin/unfolded.cgi [3.26]) and then these structures were used to generate C_α-bead models. Pseudoatoms are held together by spring-like pseudobonds. Interactions between two pseudoatoms are characterized as either nonbonded or bonded, and nonbonded interactions are characterized as either native or non-native.

Whether a nonbonded interaction is treated as favorable or unfavorable is dependent upon the protein's native state structure: if two residues form a contact in the native state, their interaction is treated favorably; otherwise, it is treated unfavorably. Two residues were considered to form a native contact if any of the heavy atoms on one residue were within 5.5 Å of any heavy atoms on the other residue in the native state structure. Following others [3.18-3.20], native interactions were governed by the following Lennard-Jones potential:

$$E_{ij,native} = \varepsilon_{native}\left\{5\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}}\right) - 6\left(\frac{\sigma_{ij}^{10}}{r_{ij}^{10}}\right)\right\}$$

where ε is the energy well depth, r_{ij} is the distance between pseudoatoms i and j in the simulation, and σ_{ij} is the distance between pseudoatoms i and j in the native state structure. Non-native interactions between pseudoatoms separated by more than four bonds were governed by the following repulsive potential:

$$E_{ij,non-native} = \varepsilon_{non-native}\left\{\frac{\sigma_{ij}^{12}}{r_{ij}^{12}}\right\}$$

where σ_{ij} and $\varepsilon_{non\text{-}native}$ are set to 4.0 Å and 0.60 kcal/mol, respectively.

Bonded interactions between pseudoatoms are governed by the following potential:

$$E_{bonded} = \sum_{bonds} k_{bond}\left(r - r_{eq}\right)^2 + \sum_{angles} k_{angle}\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} V_1\left[1 + \cos\left(\varphi - \varphi_1\right)\right] + V_3\left[1 + \cos\left(3\varphi - \varphi_3\right)\right]$$

where r, θ, and φ are pseudo bond lengths, angles, and dihedrals, respectively, r_{eq} and θ_{eq} are the bond lengths and angles in the native state structure, and φ_1 and φ_3 are phases of the torsional potentials. Following the work of [3.19], the force constants k_{bond} and k_{angle} were set to 100 kcal/mol/Å and 20 kcal/mol/rad, respectively.

3.3.2 Parameterization of ε_{native}, V_1 and V_3 values.

The parameters for ε_{native}, V_1 and V_3 were set to different values for barnase, ubiquitin and barstar in order to reproduce the melting temperature of each protein. The balance between bonded and nonbonded interactions that is governed by these parameters has been shown to influence the cooperativity of folding equilibria simulated with Gō models [3.27]. Therefore, following others, the ratio of ε_{native} : V_1 : V_3 was fixed to 0.60 : 0.50 : 0.25 for each individual protein or domain of the barnase-ubiquitin fusion proteins. These parameters are related to the stability of each domain, as uniformly increasing them increases the proportion of time the protein spends in the folded state. We therefore refer to ε_{native} as the "stabilization parameter" which implies an associated V_1 and V_3 according to the proportion given above. All interdomain contacts are treated as nonnative, with one exception: in simulations of the BU-G2–barstar complex, interdomain contacts were treated as in the barnase-barstar complex. All nonbonded interactions involving a linker residue are treated as nonnative.

For the barnase-ubiquitin fusion proteins BU-G2, BU-G6, and BU-G10, each domain was assigned the ε_{native}, V_1 and V_3 parameters associated with the free protein. Contacts between the two protein domains, as well as all contacts involving the glycine linkers, were treated as non-native. To allow for free rotation about bonds involving glycine linker residues, the V_1 and V_3 terms were set to zero for these dihedrals. For the complexes between barnase and barstar and between the barnase domain of BU-G2 and barstar, barnase-barstar contacts were determined to be native or non-native using the same distance criterion as for each domain's internal contacts; however, to ensure that barstar remains bound for the entire simulation, an ε_{native} of 1.2 kcal/mol was used. The dependence of barnase-basrtar binding on the value of ε_{native} will be explored in further studies. Table 3-1 summarizes the parameters used and number of native contacts for each domain or domain-domain interaction:

Table 3-1. Number of native contacts and summary of parameters used for each protein. All intradomain contacts are treated as in the free proteins.

Protein	Native Contacts	ε_{native} (kcal/mol)	V_1 (kcal/mol)	V_3 (kcal/mol)
Barnase	335	0.560	0.467	0.233
Ubiquitin	226	0.650	0.542	0.271
Barstar	272	0.625	0.521	0.260

3.3.3 Simulation protocol.

Simulations were performed using software written and kindly made available by Adrian Elcock (U. Iowa) [3.19, 3.22]. This software utilizes a Brownian dynamics (BD) algorithm developed by Ermak and McCammon [3.28]. These simulations include hydrodynamic interactions that model the correlated motions of waters: the displacement of one pseudoatom is directly affected by the forces that act on each other pseudoatom. Hydrodynamic interactions were implemented as described in [3.22], where it has been shown that the inclusion of hydrodynamic interactions

accelerates folding by a factor of 2 to 3. To further accelerate the configurational sampling, a low value (i.e. 1.5 Å) was used for the hydrodynamic radii assigned to the pseudoatoms. Simulations were carried out for 10 μs with a timestep of 50 fs. To reduce the computer time used to complete each simulation, a distance cutoff was imposed on the calculation of nonbonded interactions. These interactions were only calculated if r_{ij} was less than 10 Å for non-native contacts or (σ_{ij} + 6Å) for native contacts. The list of interactions satisfying these conditions was updated every 20 timesteps. Pseudobond lengths were constrained to their corresponding lengths in the native state structure after every timestep using the LINCS algorithm [3.29]. In order to statistically compare results for each system, five independent simulations were conducted using different random seeds for the Brownian dynamics. Each 10-μs simulation took ~4 days on a single 2.66 GHz CPU.

3.4 RESULTS

The goals of our study are (1) to evaluate the ability of Gō-type simulations to provide insight into mutually-exclusive folding, (2) to examine the effects of domain insertion and varying interdomain linker length in Gō-type simulations and compare results to experimental and atomistic simulation data, and (3) to examine the effects of ligand-binding on the folding-unfolding equilibrium of BU-G2.

3.4.1 Single-domain protein simulations.

Barnase, ubiquitin, and barstar were each individually subjected to 10 μs of BD simulations. The goals of these simulations were to test that the stabilization parameters chosen for the barnase, ubiquitin, and barstar proteins (summarized in Table 3-1 above) would reproduce experimental behavior at the melting temperature, sampling both unfolded and folded states with equal probability ($\Delta G^{\circ}_{fold} = 0$). Ten such independent simulations were conducted on each protein (five starting from the folded structure, five starting from a random coil structure). These simulations were conducted at the experimental melting temperatures for each protein: 51.5 °C for barnase at pH 7.5 [3.2], 100 °C for ubiquitin at pH 7 [3.2], and 69 °C for barstar at pH 8 [3.30]. Following others [3.18, 3.19, 3.21], the folding and unfolding of proteins and domains was monitored by calculating the fraction of native contacts (Q) every 50 ps. If Q = 1, this indicates that all contacts in the native structure are present in the observed structure (i.e., the structure is fully folded). If Q = 0, the observed structure is completely unfolded such that none of the contacts in the native structure are present. Two residues were determined to be in contact if they were within a distance of 1.2 σ_{ij} (see Methods). For each simulation, histograms of observed Q values were generated with 20 bins of width 0.05. These histograms omit the first 1 ns of simulation of data in order to ensure that they are not biased towards the folded starting structures. Thus, the histograms were generated using 195,000 data points (50 ps per data point). Figures 3-2, 3-3, and 3-4 show examples of two such histograms as well as the average of all histograms from simulations of barnase, ubiquitin, and barstar, respectively. All of the histograms for barnase, ubiquitin, and barstar show that both folded and unfolded conformations are observed. These histograms show roughly equal probabilities of observing unfolded (Q ~ 0.30) and folded

(Q ~ 0.75) states, with a lower probability of observing an intermediate state (Q ~ 0.50). This indicates the presence of energy minima at unfolded and folded states separated by an energy barrier. In order to quantify the ratio of observed folded and unfolded states, any state with Q < 0.50 was considered to be unfolded while any state with Q >= 0.50 was considered to be folded. Table 3-2 summarizes the results of these histograms.

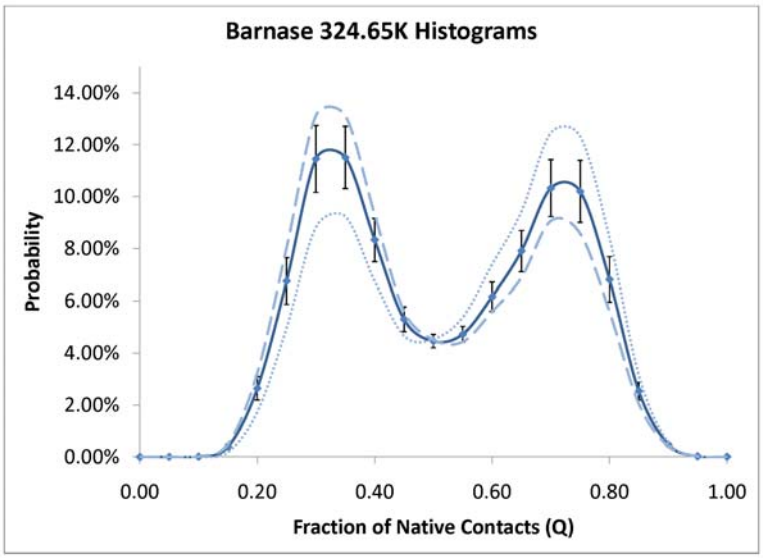

Figure 3-2. Fraction of native contacts of conformations observed in BD simulations of barnase at the barnase melting temperature (51.5 °C at pH 7.5 [3.2]). Solid line indicates the probability on each bin averaged over ten barnase simulations. Error bars show ±1 standard deviation. Dashed line corresponds to the histogram from the simulation with the largest observed probability of unfolded states. Dotted line corresponds to the histogram from the simulation with the largest observed probability of folded states.

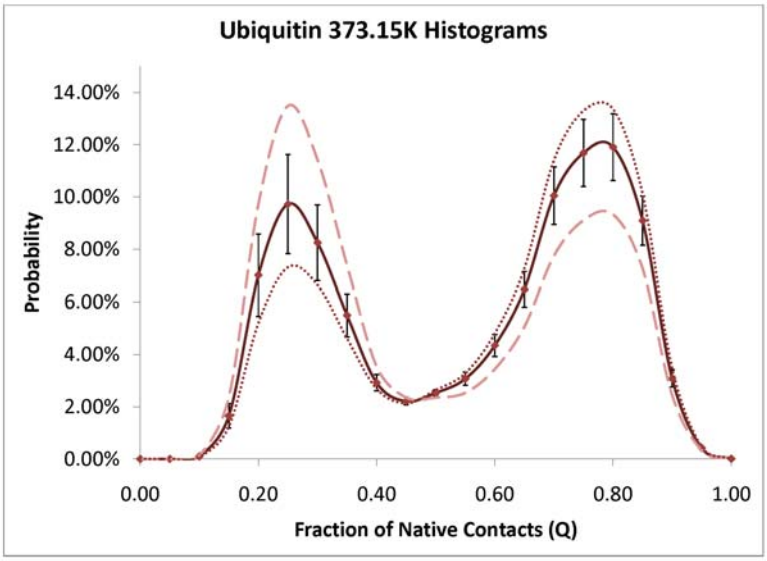

Figure 3-3. Fraction of native contacts of conformations observed in BD simulations of ubiquitin at the ubiquitin melting temperature (100.0 °C at pH 7 [3.2]). Solid line indicates the probability on each bin averaged over ten ubiquitin simulations. Error bars show ±1 standard deviation. Dashed line corresponds to the histogram from the simulation with the largest observed probability of unfolded states. Dotted line corresponds to the histogram from the simulation with the largest observed probability of folded states.

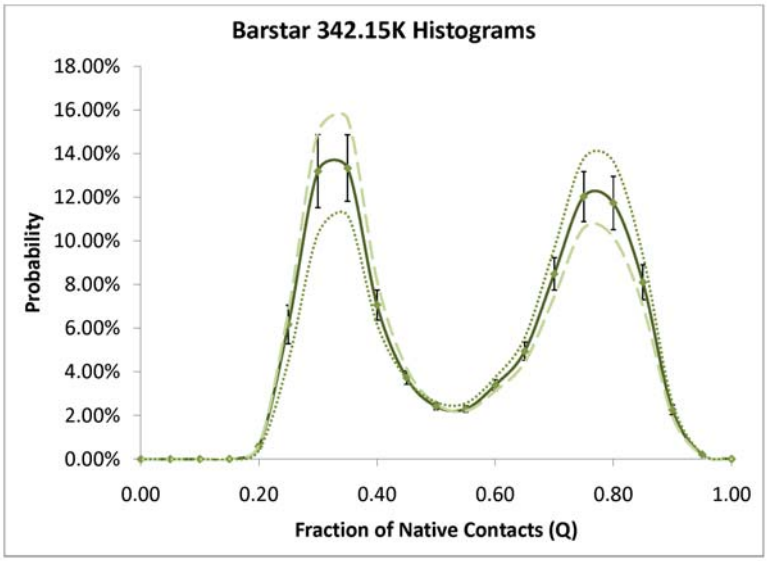

Figure 3-4. Fraction of native contacts of conformations observed in BD simulations of barstar at the barstar melting temperature (69.0 °C at pH 8 [3.30]). Solid line indicates the probability on each bin averaged over ten barstar simulations. Error bars show ±1 standard deviation. Dashed line corresponds to the histogram from the simulation with the largest observed probability of unfolded states. Dotted line corresponds to the histogram from the simulation with the largest observed probability of folded states.

Table 3-2. Summary of histogram results for single-domain protein simulations. Each protein shows nearly equal populations of folded and unfolded states. Values are averages ± one standard deviation from n=10 simulation histograms.

Protein	Temperature (°C)	Fraction Folded	ΔG°_{fold} (kcal/mol)
Barnase	51.5	49.2% ± 5.1%	-0.0 ± 0.1
Ubiquitin	100.0	60.1% ± 6.3%	-0.3 ± 0.2
Barstar	69.0	53.5% ± 4.9%	-0.1 ± 0.1

While the histograms allow us to calculate ΔG°_{fold} from the probability of observed states ($\Delta G^{\circ} = -RT \ln K$), the accuracy of this value will be poor if the simulation has not converged. One way of gauging the convergence of our simulations is to examine the probability of folding and unfolding transitions by monitoring Q versus time for simulations starting from two different structures: folded and random coil. Observing the histograms of Q reveals no statistically significant differences in the results between simulations starting from folded structures and those starting from unfolded structures. Figure 3-5 shows Q versus time for one simulation, and is representative of the behavior observed in each of the simulations of barnase, ubiquitin, and barstar: a large number of folding and unfolding transitions are observed. Table 3-2 shows a summary of the number of transitions observed during the simulations of barnase, ubiquitin, and barstar. For the purpose of quantification, a transition was determined to occur if the fraction of native contacts goes from below an unfolding threshold Q value Q_Unfold to above a folding threshold Q value Q_Fold, or vice-versa. Q_Unfold and Q_Fold are the most frequently observed values of Q associated with unfolded and folded states in the average histograms for each protein.

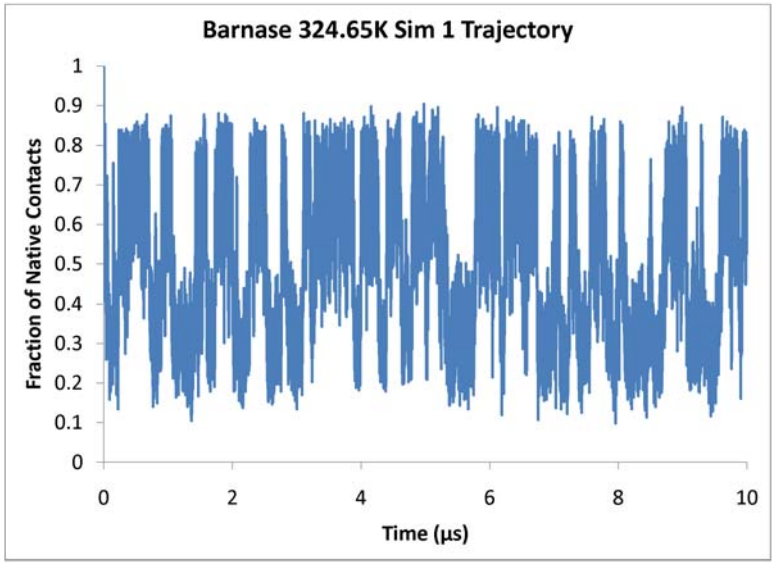

Figure 3-5. Fraction of native contacts of conformations observed in a BD simulation of barnase at the barnase melting temperature (51.5 °C) as a function of time. A large number of transitions indicates that the simulations are not sensitive to starting structure.

Table 3-3. Summary of folding and unfolding transitions observed in single-domain protein simulations at each protein's melting temperature. A folding or unfolding transition is said to occur when the observed Q goes from below Q_Unfold to above Q_Fold or vice-versa. Many transitions are observed for each simulation. Values are averages ± one standard deviation from n=10 simulation trajectories.

Protein	Temperature (°C)	Q_Unfold	Q_Fold	Average Transitions per 10 μs	Average Transition Time (ns)
Barnase	51.5	0.35	0.70	174 ± 15	58 ± 5
Ubiquitin	100.0	0.25	0.75	108 ± 11	94 ± 11
Barstar	69.0	0.35	0.80	133 ± 12	76 ± 7

3.4.2 Simulations of the BU-G2 molecular switch.

The barnase-ubiquitin molecular switch BU-G2 was subjected to 10 μs of BD simulation. This is the switch that has been experimentally determined to be the most strained BU variant that does not relieve strain by forming higher-order oligomers [3.2]. These simulations were conducted in order to see if the insertion of ubiquitin alters the folding-unfolding equilibrium observed in the

barnase simulations. Initially, ten such independent simulations were conducted: five from a state where both domains are folded, five from a state where both domains are random coils. These simulations were conducted at the barnase melting temperature (51.5 °C). Since ubiquitin's melting temperature (100 °C) is so much higher than that of barnase, it is expected that ubiquitin will fold and remain folded for the entire simulation. These simulations were analyzed by observing the fraction of native contacts on each individual domain within the protein. Figure 6 shows the fraction of native contacts observed in the barnase and ubiquitin domains, respectively, throughout the simulations, expressed as a histogram averaged over all ten simulations. While the simulations of free barnase resulted in nearly equal populations of folded and unfolded state, the vast majority of observed conformations of BU-G2 have the barnase domain in the unfolded state. As predicted, the ubiquitin domain folds and/or remains folded in every simulation. There is also far less variability in the simulations, as the ±1 standard deviation error bars on each bin are very small.

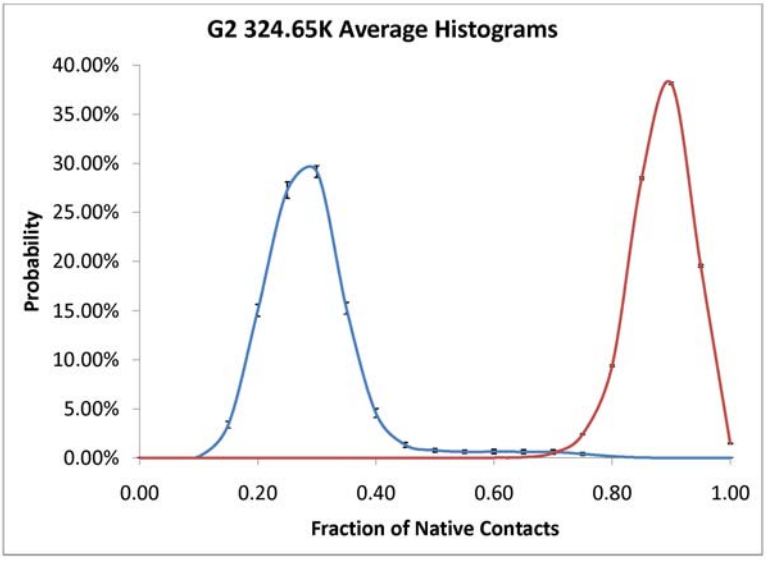

Figure 3-6. Fraction of native contacts in the barnase (blue) and ubiquitin (red) domains of BU-G2 observed in BD simulations at the barnase melting temperature (51.5 °C). While free barnase was half-folded at this temperature, the barnase domain of the barnase-ubiquitin chimera is unfolded for almost the entire simulation

duration. Histograms are an average over ten BU-G2 simulations (five from fully-folded structures, five from random coil structures). Error bars show ±1 standard deviation.

While the barnase domain remains unfolded for almost the entire 10 μs in each simulation, it should be noted that there are some observed states where barnase appears to be partially or significantly folded. As shown in Figure 3-6, the probability of observing a Q value of 0.50 to 0.75 in barnase is slightly above zero. This is made clearer by looking at the fraction of native contacts in each domain as a function of time. Figure 3-7 shows Q versus time for one representative simulation of BU-G2. This trajectory shows that it is possible for barnase to transiently fold up in the presence of folded ubiquitin. Figure 3-8 shows a series of snapshots from the representative simulation of BU-G2: the initial structure with both domains folded, the point of greatest barnase unfolding (8.16 μs), and the point of greatest barnase folding (9.81 μs).

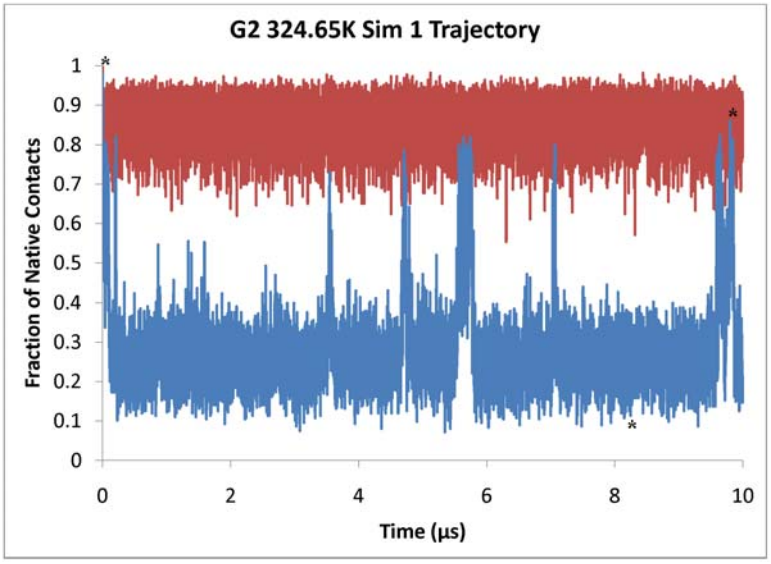

Figure 3-7. Fraction of native contacts in the barnase (blue) and ubiquitin (red) domains observed in a BD simulation of BU-G2 at the barnase melting temperature (51.5 °C) as a function of time. While barnase remains unfolded most of the time, there are observed transitions to a short-lived folded state. Asterisks indicate states with accompanying pictures in Figure 3-8.

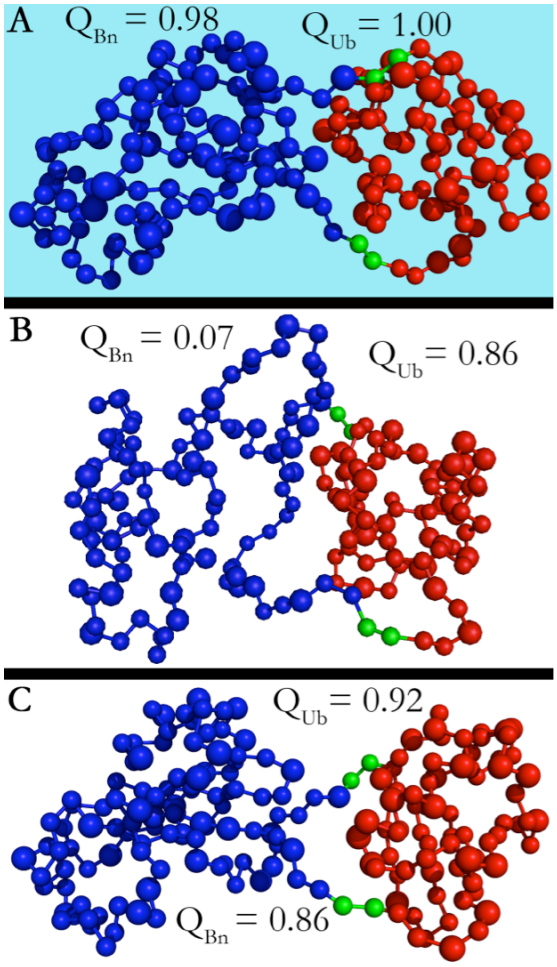

Figure 3-8. Snapshots from the representative simulation of BU-G2 plotted in Figure 3-7. The barnase domain is colored blue, ubiquitin is red, and flexible linkers are green. (A) Starting structure with both domains folded. (B) Snapshot of BU-G2 at 8.16 μs. This structure has the lowest fraction of native contacts in the barnase domain for this simulation with only 7% of contacts formed. (C) Snapshot of BU-G2 at 9.8 μs. This structure has the highest fraction of native contacts in the barnase domain with 86%. As seen in Figure 3-7, this folded state is very short-lived. Images were generated using the PyMOL program (DeLano Scientific).

To further explore the thermodynamic tug-of-war between the domains in G2, ubiquitin was unfolded by using an extremely weak value for the stabilization parameter (ε_{native} = 0.065). In this situation, ubiquitin unfolds very rapidly (within 4 ns), before barnase shows any sign of unfolding. Additional simulations of BU-G2 were conducted from the resulting state where the barnase domain is folded and the ubiquitin domain is unfolded. The fraction of native contacts

for the barnase and ubiquitin domains in this starting structure are 75.8% and 4.9%, respectively. Five simulations from this starting structure were conducted and analyzed by fraction of native contacts. Average histograms of observed fraction of native contacts in these five simulations are identical to those shown in Figure 3-6, further indicating that these simulations are insensitive to starting structure.

In the BU-G2 simulations starting from a structure where barnase is folded and ubiquitin is unfolded, ubiquitin must fold and barnase must unfold. It is interesting to consider how this occurs: does barnase need to unfold first in order for ubiquitin to fold, or is ubiquitin capable of folding while barnase is still folded? To answer this question, plots of Q versus time were generated for each trajectory. For all five independent simulations, ubiquitin folded before barnase unfolded. In all five cases, ubiquitin finished folding before barnase finished unfolding. Ubiquitin folded within 40 ns and barnase unfolded within 150 ns for each simulation. Figure 3-9 shows a plot of Q versus time for one representative simulation of BU-G2 from a state where barnase is folded and ubiquitin is unfolded. To emphasize the initial folding and unfolding transition events in the barnase and ubiquitin domains, respectively, only the first 200 ns of simulation data are shown. Figure 3-10 shows snapshots taken from this representative simulation during the first 200 ns.

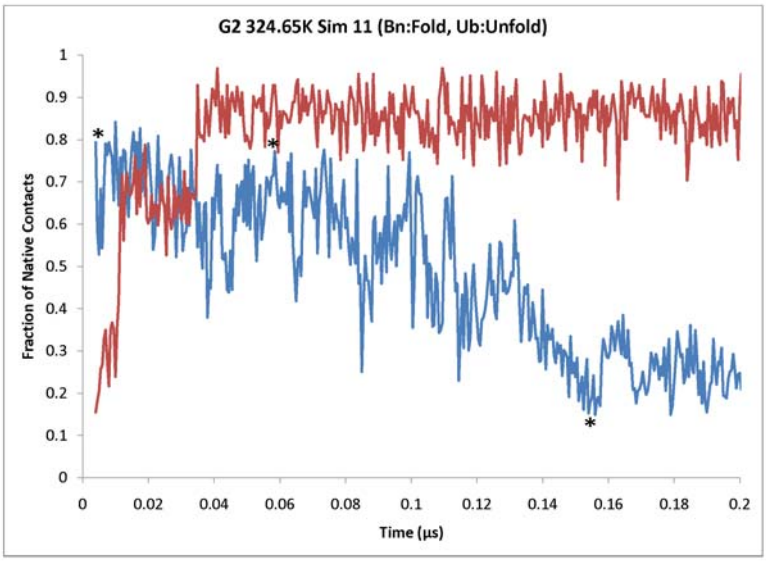

Figure 3-9. Fraction of native contacts of conformations observed in a BD simulation of BU-G2 at the barnase melting temperature (51.5 °C) as a function of time. In each simulation from this starting structure (where barnase is folded and ubiquitin is unfolded), ubiquitin folds before barnase unfolds. Asterisks indicate states with accompanying pictures in Figure 3-10.

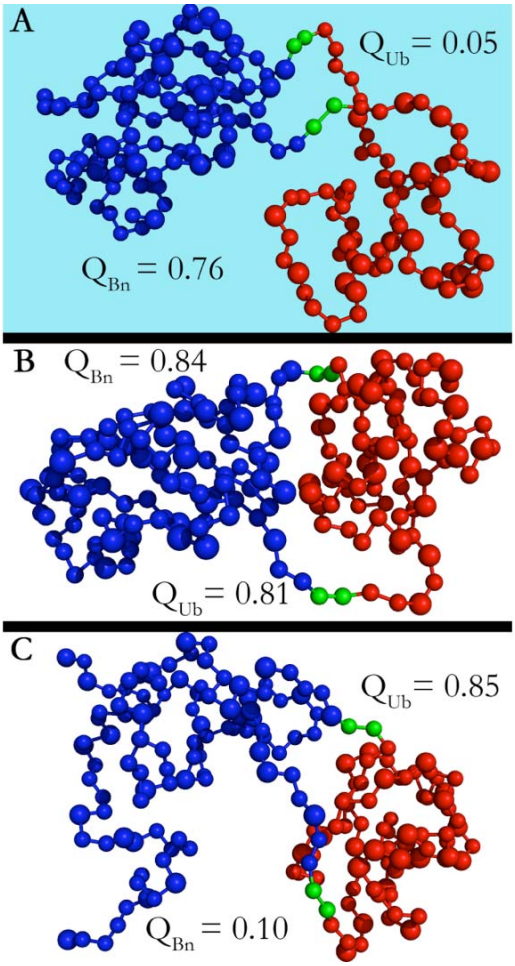

Figure 3-10. Snapshots from the simulation of BU-G2 starting from barnase in the folded state and ubiquitin in the unfolded state plotted in Figure 3-9. The barnase domain is colored blue, ubiquitin is red, and flexible linkers are green. (A) Starting structure with barnase folded and ubiquitin unfolded. (B) Snapshot of BU-G2 at 55 ns. At this point in the simulation, ubiquitin has folded with the barnase domain still intact. (C) Snapshot of BU-G2 at 150 ns. Once ubiquitin has folded, barnase begins to unfold, reaching a fraction of native contacts of 10% at 150 ns. Images were generated using the PyMOL program (DeLano Scientific).

3.4.3 BU-G6 and BU-G10 simulations.

As previously shown in Cutler et. al. [3.2], the degree of unfolding in the barnase domain is dependent on the interdomain linker length. To test the ability of our coarse-grained simulations

to model this dependence, the barnase-ubiquitin fusion proteins BU-G6 and BU-G10 were subjected to 10 μs of BD simulation. Five independent simulations were conducted on BU-G6 and BU-G10, each from a folded structure. These simulations were analyzed by generating histograms of observed fraction of native contacts in both the barnase and ubiquitin domains, as well as by plotting Q versus time for both domains in each simulation. The histograms generated from the BU-G6 and BU-G10 simulations look very similar to those generated for the BU-G2 simulations in both domains. However, a small difference can be observed in the frequency of observing folded states in the barnase domain. Figure 3-11 shows the average histograms for the fraction of native contacts in the barnase domain for BU-G2 and BU-G10, focusing on the region where barnase is partially folded. More partially folded structures are observed in BU-G10 than in BU-G2, indicating that BU-G10 is slightly less strained than BU-G2 (BU-G6 was left out of Figure 3-11 for clarity). The fraction of folded states can be quantified for the barnase-ubiquitin fusion protein variants in the same way that the single-domain protein simulations were quantified: any state with Q < 0.50 is considered to be unfolded while any state with Q >= 0.50 is considered to be folded. Using this analysis, an estimate for the free energy of folding can be determined. While $\Delta G^{\circ}_{fold} = 0$ for free barnase at 51.5 °C, the insertion of ubiquitin forces the folding equlibrium towards the unfolded state ($\Delta G^{\circ}_{fold} > 0$). The barnase domains of fusion proteins BU-G2, BU-G6, and BU-G10 are destabilized to different degrees: a variant with more flexible linkers between the domains can better accommodate insertion of ubiquitin domain, making it easier for barnase to fold. This trend is summarized in Table 4.

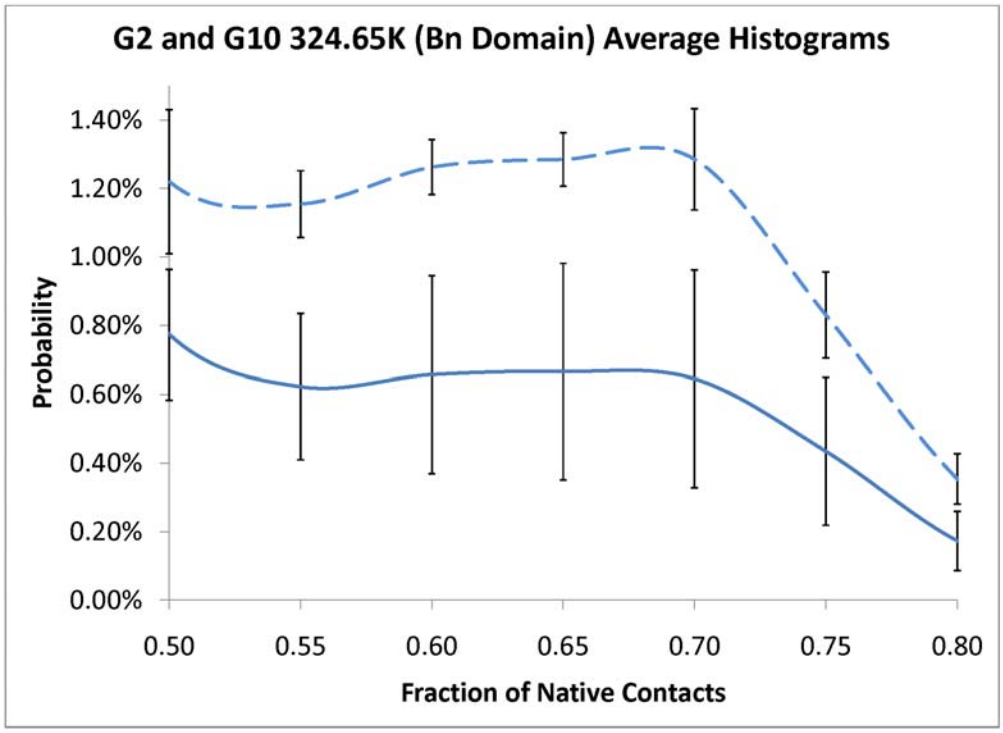

Figure 3-11. Fraction of native contacts in the barnase domains of BU-G2 (solid line) and BU-G10 (dashed line) observed in BD simulations at the barnase melting temperature (51.5 °C). More folded states are observed for BU-G10 than BU-G2, indicating that the increased flexibility of BU-G10 allows barnase to fold more easily. However, the barnase domain is unfolded in the vast majority of observed states for each simulation. The BU-G2 histogram is an average over fifteen simulations, while the BU-G10 histogram is an average over five simulations. Error bars show ±1 standard deviation.

Table 3-4. Summary of histogram results for barnase-ubiquitin fusion protein simulations. As more linkers are added between the two domains, strain is relieved and the energetic barrier to folding for barnase is reduced. Values are averages ± one standard deviation from n simulation histograms (n=15 for BU-G2, n=5 for BU-G6 and BU-G10).

Protein	Average Folding Transitions per 10 μs	Average Folded State Duration (ns)
BU-G2	14.4 ± 5.5	20.2 ± 4.2
BU-G6	18.7 ± 10.0	25.8 ± 5.8
BU-G10	26.1 ± 3.1	23.6 ± 1.6

The folded state of barnase can also be characterized based on folding and unfolding transitions. Based on the transition definition for barnase used in Table 3-3 (progression from Q

< 0.35 to Q > 0.70 or vice-versa), we can count the number of folding and unfolding transitions observed during the simulations of BU-G2, BU-G6, and BU-G10. However, since folding and unfolding transitions do not happen with similar frequency as they did in the free barnase simulation, it no longer makes sense to take averages over both folding and unfolding transitions. Therefore, the number of folding and unfolding transitions and their durations were calculated separately. The average time for an unfolding transition can be interpreted as the average duration for a folded state. Table 3-5 shows a summary of the number of folding transitions and average duration for a folded state observed during the simulations of BU-G2, BU-G6, and BU-G10.

Table 3-5. Summary of folding transitions and average folded state duration of the barnase domain in simulations of barnase-ubiquitin fusion proteins. Both measures of degree of folding in the barnase domain trend upward as more flexible linkers are added between the two domains. Values are averages ± one standard deviation from n simulation trajectories (n=15 for BU-G2, n=5 for BU-G6 and BU-G10).

Protein	Average Folding Transitions per 10 μs	Average Folded State Duration (ns)
BU-G2	14.4 ± 5.5	20.2 ± 4.2
BU-G6	18.7 ± 10.0	25.8 ± 5.8
BU-G10	26.1 ± 3.1	23.6 ± 1.6

3.4.4 G2-barstar complex simulations.

The binding of effector molecules stabilizes protein domains. In the context of a molecular switch, the binding of an effector molecule to one domain should stabilize that domain to a point where it may be more stable than the other domain. The binding of the natural barnase inhibitor barstar to the barnase domain has been experimentally shown to drive the unfolding of ubiquitin for one variant of the barnase-ubiquitin fusion protein [3.1]. Our atomistic simulations of the

BU-G2–barstar complex showed that barstar binding stabilized the barnase domain, but no unfolding of ubiquitin could be observed [3.2]. In order to test our hypothesis that barstar binding will cause ubiquitin to unfold, we conducted five independent simulations of BU-G2–barstar complex. Interdomain contacts between barnase and barstar were treated as native or nonnative depending on their distance in the crystal structure (98 barnase-barstar contacts were determined to be native), and the stabilization parameter was set to 1.2 kcal/mol in order to ensure that barstar stayed bound to the barnase domain for the entire simulation (see Methods). Data from the simulations of the BU-G2–barstar complex were compared with those from BU-G2 in the absence of barstar. Figures 3-12 and 3-13 show histograms of the fraction of native contacts of BU-G2 in the presence and absence of barstar for the barnase and ubiquitin domains, respectively. As expected, the binding of barstar has a dramatic effect on the folding equilibrium of barnase, keeping the barnase domain folded for the entire duration in all simulations. However, the effect on the ubiquitin domain is more subtle: comparison of the histograms in Figure 3-13 shows, for the BU-G2–barstar complex simulations, the shape of the ubiquitin histogram is distorted towards lower values for the fraction of native contacts. The presence of barstar has a destabilizing effect on the ubiquitin domain, but the effect is too slight to observe unfolding in the ubiquitin domain.

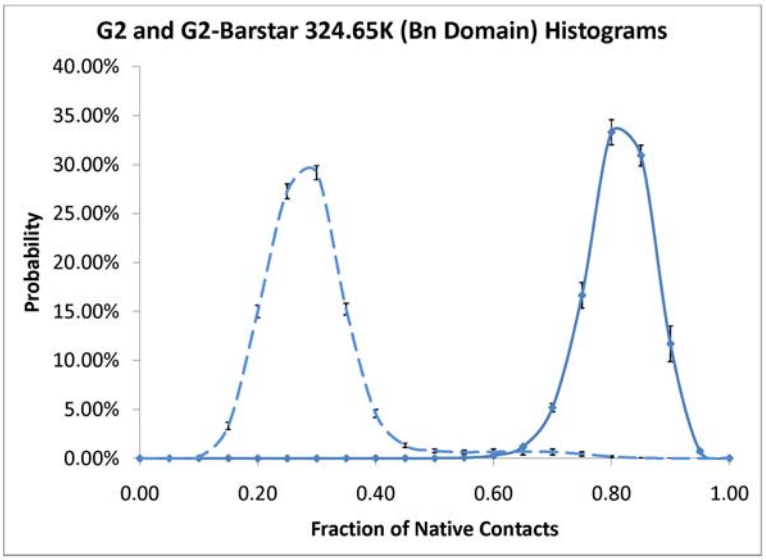

Figure 3-12. Fraction of native contacts in the barnase domain of conformations observed in BD simulations of BU-G2 at the barnase melting temperature (51.5 °C) in the presence (solid line) and absence (dashed line) of barstar. The addition of barstar dramatically shifts the barnase folding equilibrium towards the folded state. Histograms are averaged over n simulations (n=15 for lone BU-G2, n=5 for the BU-G2–barstar complex). Error bars show ±1 standard deviation.

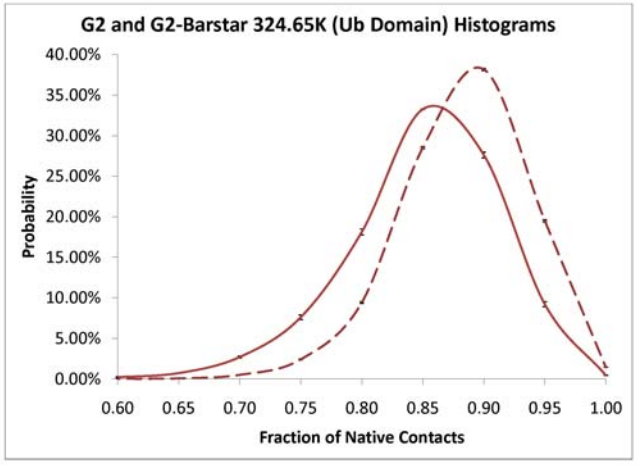

Figure 3-13. Ubiquitin domain fraction of native contacts of conformations observed in BD simulations of BU-G2 at the barnase melting temperature (51.5 °C) in the presence (solid line) and absence (dashed line) of barstar. The addition of barstar only slightly, but significantly, destabilizes the ubiquitin domain. Histograms are averaged over n simulations (n=15 for lone BU-G2, n=5 for the BU-G2–barstar complex). Error bars show ±1 standard deviation.

3.5 DISCUSSION

In this study, coarse-grained simulations using a Gō-type model we used in order to explore the thermodynamic tug-of-war between the domains of the barnase-ubiquitin fusion protein. As done by Elcock [3.19] for the simulations of cotranslational protein folding, the stabilization parameter on each of the protein domains in these simulations were set to different values in order to reproduce the thermodynamics of individual protein domains at their experimentally-determined melting temperature. This was a necessary step to reproduce the relative stabilities of the protein domains, as using the same parameter for both domains would not lead to any thermodynamic tug-of-war between the protein domains.

As with any computer simulation approach, there are limitations concerning both the model and the simulation details. There are a few notable limitations to the simulation methods implemented in this study.

Since no experimentally-determined structures are available for the barnase-ubiquitin fusion protein, our folded starting structures are based on models constructed by satisfaction of spatial restraints obtained from the crystal structures of barnase and ubiquitin in their isolated forms (See Chapter 2). One potential concern is that our results may depend on the conformation of the starting structure. While our starting structures might bias simulations towards a particular state, this seems unlikely due to fact that all data taken from simulations with a fully folded model were statistically indistinguishable from those taken from simulations starting from a random coil model of the same system.

As with any coarse-grained model that does not explicitly model the side-chains, we are unable to predict the effects of mutation in the absence of relevant thermodynamic data from laboratory experiments. In this study, a C_α model was used with the goal of exploring the design

principles using a minimalist model of the protein. To explore the effects of mutations, one can include course-grained representations of the side-chains as an addition to the C_α model.

Finally, the conformational sampling was accelerated by using a small value for the hydrodynamic radii. As a result, these simulations are not expected to accurately reproduce translational and rotational diffusion coefficients. We have chosen to do so since for the purpose of this study we are only interested in observing the relative populations of folded and unfolded states.

3.6 FUTURE DIRECTIONS

We wish to explore the temperature dependence of the folding-unfolding equilibrium in barnase-ubiquitin fusion proteins. Guanidine chloride denaturation experiments suggest that barnase may be more stable than ubiquitin at lower temperatures [3.1]. Exploring the behavior of the barnase-ubiquitin fusion protein at lower temperatures will require the utilization of histogram techniques [3.31, 3.32] to reproduce experimentally measured folding free energies at those temperatures.

Additionally, we wish to test the generality of the mutually exclusive folding mechanism by inserting ubiquitin into one of the other five insertion loops present in barnase [3.15]. Sensitivity of a binding-induced switching transition may depend on the degree of thermodynamic coupling between the domains, which in turn may differ between rigid and flexible surface loops. Designing the best molecular switch possible from two given domains *via* this mechanism may require the determination of the best possible insertion site.

3.7 CONCLUSIONS

Coarse-grained simulations of the barnase-ubiquitin fusion protein at 51.5 °C show that the folding of ubiquitin drives the unfolding of barnase. Consistent with experiments and previous atomistic simulations [3.2], decreasing the length of linker sequences between the domains decreased the folding fraction of barnase in the fusion proteins. The binding of barstar to barnase prevents the unfolding of barnase, but only slightly destabilizes the ubiquitin domain. This study has shown that Gō-type models may be effectively utilized to examine mutually exclusive folding.

4.0 CONCLUSIONS AND FUTURE DIRECTIONS

The computational studies described in this thesis have provided structural and mechanistic details of mutually exclusive folding in novel, engineered molecular switches. Atomistic and coarse-grained simulations have both proven to aid in our understanding of the thermodynamic tug-of-war between barnase and ubiquitin in barnase-ubiqutin fusion proteins.

In atomistic simulations, shortening the length of the linker peptide sequence joining the two domains increased the degree of protein unfolding in the molecular switch. Because of the level of detail of these simulations, we were able to suggest a mechanism by which the barnase domain is unfolded in the BU-G2 variant, namely that several barnase residues on the amino side of the ubiquitin insertion are pulled away from the rest of the barnase structure. Mechanisms such as this one point to interesting regions of the protein domains, and may further guide the engineering of molecular switches.

Coarse-grained simulations of the barnase-ubiquitin fusion protein extensively sampled the thermodynamic tug-of-war that exists between the two protein domains, resulting in complete protein folding and unfolding events. As a result of this extensive sampling, these simulations allowed us to observe how alterations to the system resulted in slight changes in protein stability. These simulations were able to distinguish the degree of folding in the barnase domains of BU-G2, BU-G6, and BU-G10, as well as the destabilization of ubiquitin as a result of barstar binding to the barnase domain. The ability to quickly determine a folding-unfolding

equilibrium in the presence and absence of effector molecules will become more useful as more switches are designed that make use of protein folding and unfolding.

These results suggest that computational studies are useful for probing mutually exclusive folding in other molecular switches. Because of the computational ease of coarse-grained modeling and simulation of such proteins, these simulations may be used to test design principles before engineering the molecular switch in the laboratory. As only switches of barnase-ubiquitin and barnase-GCN4 have been designed using this paradigm to date, we may examine other possible domains to be used in mutually exclusive folding molecular switches in the future. Atomistic simulations may be used on established molecular switches in order to gain insight into the fault points of individual domains in the switch. This information may be useful towards the fine-tuning of other molecular switches.

The ability to control the activity of an individual protein domain may be applied to the study of other protein switches. For instance, the targets of many protein kinases are still unknown [4.1]. Researchers may be able to apply strategies of mutually exclusive folding using protein kinases as one of the domains in order to reversibly control their activity. This may be useful as an alternative technique to radiolabeling [4.1] for the determination of the substrates of kinase proteins.

One important consideration about the simulations constructed to date is that they only include single copies of each molecule studied. CD spectroscopy and analytical ultracentrifugation experiments on the most severely strained of barnase-ubiquitin fusion protein variants, BU-G1 and BU-G0, have suggested that two of these fusion proteins may interact to form domain-swapped dimers [4.2]. It has been suggested that strain within a protein can drive it to domain-swap [4.3]. Therefore, domain-swapping is an important consideration in the future

design of molecular switches using the mutually exclusive folding paradigm. In the future, we plan to employ coarse-grained simulations of multiple copies of the barnase-ubiquitin fusion variants in order to see if these computations corroborate the experimental data and, if so, to predict potential structures of these domain-swapped dimer structures. Since the formation of a domain-swapped structure would constitute a failed switch, these simulations may be an important part of future computational assays of possible molecular switches.

We have collaborated with experimentalists in order to provide validation for our atomistic simulations by gauging the degree of unfolding in each domain with tryptophan (Trp) fluorescence and CD spectroscopy. Since barnase has three tryptophan residues, Trp fluorescence can provide structural insight into individual regions of the protein where the Trp residues are located [4.4]. Therefore, Trp fluorescence may be a useful technique for validating theoretical predictions of unfolding at the specific regions of molecular switch domains.

Atomistic and coarse-grained simulations were both able to provide interesting results for the barnase-ubiquitin fusion proteins. However, atomistic simulations are limited by their required computational effort, and the reduced detail of coarse-grained simulations limits the types of predictions that can be made. Therefore, it may be useful to utilize a simulation method of intermediate detail.

To our knowledge, the simulations described in this thesis document are the first at any level of structural detail to show that the folding of one protein domain can drive the unfolding of another. While we know that random mutations can result in the insertion of one protein domain into another in an analogous fashion to that utilized in the mutually-exclusive folding mechanism, we have yet to take our knowledge from these engineered molecular switches and

apply it to an evolutionary study of multidomain proteins. These simulations methods may be the first step towards developing the tools for such a study.

Thus, the general molecular switch design paradigm of mutually-exclusive folding raises many questions to be addressed in the future. The atomistic and coarse-grained simulations described in this thesis represent the first attempts to answer some of these questions, laying a foundation for developing computational assays for tuning the specificity of molecular switches that function through mutually exclusive folding.

BIBLIOGRAPHY

Chapter 1

1.1. Buskirk, A.R. and D.R. Liu. (2005) Creating Small-Molecule-Dependent Switches to Modulate Biological Functions. *Chem. Biol. 12*, 151-161.

1.2. Villaverde, A. (2003) Allosteric enzymes as biosensors for molecular diagnosis. *FEBS Lett. 554*, 169-172.

1.3. Dueber, J.E., et al. (2004) Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. *Curr. Opin. Struct. Biol. 14*, 690-699.

1.4. Pearl, L.H. and C. Prodromou (2006) Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem. 75,* 271-294.

1.5. Catterall, W.A. (1995) Structure and function of voltage-gated ion channels. *Annu. Rev. Biochem. 64*, 493-531.

1.6. Levitan, I.B. (1994) Modulation of ion channels by protein phosphorylation and dephosphorylation. *Annu. Rev. Biochem. 56*, 193-212.

1.7. Mascia, M.P., Trudell, J.R., and R.A. Harris. (2000) Specific binding sites for alcohols and anaesthetics on ligand-gated ion channels. *Proc. Natl. Acad. Sci. USA 97*, 9305-9310.

1.8. Johnson, S.A. and T. Hunter. (2005) Kinomics: methods for deciphering the kinome. *Nature Methods 2,* 17-25.

1.9. T. Hunter. (1995) Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell 80,* 225-236.

1.10. Lim, W.A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol. 12*, 61-68.

1.11. Teichmann, S.A., C. Chothia, and M. Gerstein. (1999) Advances in structural genomics. *Curr. Opin. Struct. Biol. 9*, 390-399.

1.12. Ostermeier, M. (2005) Engineering allosteric protein switches by domain insertion. *Protein Engineering, Design & Selection 18*, 359-364.

1.13. Guntas, G. and M. Ostermeier. (2004) Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol. 336*, 263-273.

1.14. Doi, N. and H. Yanagawa. (1999) Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett. 453,* 305-307.

1.15. Doi, N. and H. Yanagawa. (1999) Insertional gene fusion technology. *FEBS Lett. 457*, 1-4.

1.16. Baird, G.S., D.A. Zacharias, and R.Y. Tsien. (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. USA 96,* 11241-11246.

1.17. Siegel, M.S. and E.Y. Isacoff. (1997) A genetically encoded optical probe of membrane voltage. *Neuron 19,* 735-741.

1.18. Ataka, K. and V.A. Pieribone. (2002) A genetically targetable fluorescent probe of channel gating with rapid kinetics. *Biophys. J. 82,* 509-516.

1.19. Buskirk, A.R., et al. (2004) Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc. Natl. Acad. Sci. USA 101,* 10505-10510.

1.20. Skretas, G. and D.W. Wood. (2005) Regulation of protein activity with small-molecule-controlled inteins. *Protein Sci. 14,* 523-532.

1.21. Radley, T.L., et al.. (2003) Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol. 332*, 529-536.

1.22. Paddon, C.J., Vasantha, N., and R. W. Hartley. (1989) Translation and processing of *Bacillus amyloliquefaciens* extracellular RNase. *J. Bact. 171,* 1185-1187.

1.23. Ha, J.H., et al. (2006) Modular enzyme design: regulation by mutually exclusive protein folding. *J. Mol. Biol. 357,* 1058-1062.

1.24. Cutler, T.A., et al. (2009) Effect of interdomain linker length on an antagonistic folding-unfolding equilibrium between two protein domains. *J. Mol. Biol. 386*, 854-868.

1.25. Martin, C., et al. (1999) Refinement and structural analysis of barnase at 1.5 A resolution. *Acta. Crystallogr. D. Biol. Crystallogr. 55,* 386-398.

1.26. Hornak, V., et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins 65,* 712-725.

1.27. Chavez, L.L., J.N. Onuchic, and C. Clementi. (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc. 126*, 8426-8432.

1.28. Elcock, A.H. (2006) Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol. 2*, e98.

1.29. Zagrovic, B. and V. Pande. (2003) Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J. Comput. Chem. 24*, 1432-1436.

1.30. Hornak, V., et al. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA 103,* 915-920.

1.31. Ermak, D.L. and J.A. McCammon. (1978) Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys. 69*, 1352-1360.

1.32. Rotne, J. and S. Prager. (1969) Variational Treatment of Hydrodynamic Interaction in Polymers. *J. Chem. Phys. 50*, 4831-4837.

1.33. Yamakawa, H. (1970) Transport Properties of Polymer Chains in Dilute Solution: Hydrodynamic Interaction. *J. Chem. Phys. 53,* 436-443.

1.34. Frembgen-Kesner, T. and A. Elcock. (2009) Striking Effects of Hydrodynamic Interactions on the Simulated Diffusion and Folding of Proteins. *J. Chem. Theory Comput. 5,* 242-256.

Chapter 2

2.1. Ha, J.-H., Butler, J. S., Mitrea, D. M., and S. N. Loh. (2006) Modular enzyme design: regulation by mutually exclusive protein folding. *J. Mol. Biol. 357*, 1058-1062.

2.2. Radley, T. L., Markowska, A. I., Bettinger, B. T., Ha, J.-H., and S. N. Loh. (2003) Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol. 332*, 529-536.

2.3. Cutler, T., and S. N. Loh. (2007) Thermodynamic analysis of an antagonistic folding-unfolding equilibrium between two protein domains. *J. Mol. Biol. 371*, 308-316.

2.4. Krantz, B., Dothager, R. S., and T. R. Sosnick. (2004) Discerning the structure and energy of multiple transition states in protein folding using y-analysis. *J. Mol. Biol. 337*, 463-475.

2.5. Sosnick, T. R., Dothager, R. S., and B. Krantz. (2004) Differences in the folding transition state of ubiquitin indicated by Φ and ψ analyses. *Proc. Natl. Acad. Sci. USA 101*, 17377-17382.

2.6. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M., and G. I. Makhatadze. (1999) Engineering a thermostable protein via optimization of charge-charge interactions on a protein surface. *Biochemistry 38*.

2.7. Teichmann, S. A., Chothia, C., and M. Gerstein. (1999) Advances in structural genomics. *Curr. Opin. Struct. Biol. 9*, 390-399.

2.8. Holm, L., and C. Sander. (1994) The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res. 22*, 3600-3609.

2.9. Jones, S., Stewart, M., Michie, A., Swindells, M. B., Orengo, C., and J. M. Thornton. (1998) Domain assignment for protein structures using a consensus approach: characterization and analysis. *Protein Sci. 7*, 233-242.

2.10. Aroul-Selvam, R., Hubbard, T., and R. Sasidharan. (2004) Domain insertions in protein structures. *J. Mol. Biol. 338*, 633-641.

2.11. Liu, Y., and D. Eisenberg. (2002) 3D domain swapping: as domains continue to swap. *Protein Sci. 11*, 1285-1299.

2.12. Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y., and D. Baker. (2001) Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc. Natl. Acad. Sci. USA 98*, 10687-10691.

2.13. Newcomer, M. E. (2002) Protein folding and three-dimensional domain swapping: a strained relationship? *Curr. Opin. Struct. Biol. 12*, 48-53.

2.14. Patel, S. D., Chen, C. P., Bahna, F., Honig, B., and L. Shapiro. (2003) Cadherin-mediated cell-cell adhesion: sticking together as a family. *Curr. Opin. Struct. Biol. 13*, 690-698.

2.15. Lazar, G. A., Desjarlais, J. R., and T. M. Handel. (1997) De novo design of the hydrophobic core of ubiquitin. *Protein Sci. 6*, 1167-1178.

2.16. Vuilleumier, S., Sancho, J., Loewenthal, R., and A. R. Fersht. (1993) Circular dichroism studies of barnase and its mutants: characterization of the contribution of aromatic side chains. *Biochemistry 32*, 10303-10313.

2.17. Schreiber, G., and A. R. Fersht. (1993) Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry 32*, 5145-5150.

2.18. Hartley, R. W. (1993) Directed mutagenesis and barnase-barstar recognition. *Biochemistry 32*, 5978-5984.

2.19. Li, A., and V. Daggett. (1998) Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate. *J. Mol. Biol. 275*, 677-694.

2.20. Kabsch, W., and C. Sander. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577-2637.

2.21. Byeon, I.-J. L., Louis, J. M., and A. M. Gronenborn. (2003) A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *J. Mol. Biol. 333*, 141-152.

2.22. LeFevre, K. R., and M. H. J. Cordes. (2003) Retroevolution of l Cro toward a stable monomer. *Proc. Natl. Acad. Sci. USA 100*, 2345-2350.

2.23. Patel, S., Ciatto, C., Chen, C. P., Bahna, F., Rajebhosale, M., Arkus, N., Schieren, I., Jessell, T. M., Honig, B., and S. R. Price. (2006) Type II cadherin ectodomain structures: implications for classical cadherin specificity. *Cell 124*, 1255-1268.

2.24. Parisini, E., Higgins, J. M. G., Liu, J., Brenner, M. B., and J. Wang. (2007) The crystal structure of human E-cadherin domains 1 and 2, and comparision with other cadherins in the context of adhesion mechanism. *J. Mol. Biol. 373*, 401-411.

2.25. Neira, J. L., Vázquez, E., and A. R. Fersht. (2000) Stability and folding of the protein complexes of barnase. *Eur. J. Biochem. 267*, 2859-2870.

2.26. Zegers, I., Deswarte, J., and L. Wyns. (1999) Trimeric domain-swapped barnase. *Proc. Natl. Acad. Sci. USA 96*, 818-822.

2.27. Zagrovic, B., and V. Pande. (2003) Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J. Comput. Chem. 24*, 1432-1436.

2.28. Hornak, V., Okur, A., Rizzo, R. C., and C. Simmerling. (2006) HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc. 128*, 2812-2813.

2.29. Hornak, V., Okur, A., Rizzo, R. C., and C. Simmerling. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc Natl Acad Sci U S A 103*, 915-920.

2.30. Walser, R., and W. F. van Gunsteren. (2001) Viscosity dependence of protein dynamics. *Proteins: Structure, Function, and Genetics 42*, 414-421.

2.31. Nymeyer, H., and A. E. Garcia. (2003) Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. USA 100*, 13934-13939.

2.32. Zhu, J., Alexov, E., and B. H. Honig. (2005) Comparative study of generalized Born models: Born radii and peptide folding. *J. Phys. Chem. B 109*, 3008-3022.

2.33. Okur, A., Wickstrom, L., Layten, M., Geney, R., Song, K., Hornak, V., and C. Simmerling. (2006) Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *J. Chem. Theory Comput. 2*, 420-433.

2.34. Zhou, R. (2003) Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins 53*, 148-161.

2.35. Li, Z., Li, H., Devasahayam, G., Gemmill, T., Chaturvedi, V., Hanes, S. D., and P. Van Roey. (2005) The structure of the Candida albicans Ess1 prolyl isomerase reveals a well-ordered linker that restricts domain mobility. *Biochemistry 44*, 6180-6189.

2.36. Linhananta, A., Zhou, H., and Y. Zhou. (2002) The dual role of a loop with low loop contact distance in folding and domain swapping. *Protein Sci. 11*, 1695-1701.

2.37. Geney, R., Layten, M., Gomperts, R., Hornak, V., and C. Simmerling. (2006) Investigation of salt bridge stability in a generalized Born solvent model. *J. Chem. Theory Comput. 2*, 115-127.

2.38. Felts, A. K., Harano, Y., Gallicchio, E., and R. M. Levy. (2004) Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins 56*, 310-321.

2.39. Yu, Z. Y., Jacobson, M. P., Josovitz, J., Rapp, C. S., and A. R. Friesner. (2004) First-shell solvation of ion pairs: Correction of systematic errors in implicit solvent models. *J. Phys. Chem. B 108*, 6643-6654.

2.40. Pitera, J. W., and W. C. Swope. (2003) Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci. USA 100*, 7587-7592.

2.41. Pace, C. N., and J. M. Scholtz. (1997) (Creighton, T. E., Ed.) pp 299-320, Oxford University Press, New York.

2.42. Santoro, M. M., and D. W. Bolen. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha chymotrypsin using different denaturants. *Biochemistry 27*, 8063-8068.

2.43. Mauguen, Y., Hartley, R. W., Dodson, E. J., Dodson, G. G., Bricogne, G., Chothia, C., and A. Jack. (1982) Molecular structure of a new family of ribonucleases. *Nature 297*, 162-164.

2.44. Vijay-Kumar, S., Bugg, C. E., and W. J. Cook. (1987) Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol. 194*, 531-544.

2.45. Eswar, N., Marti-Renom, M. A., Webb, B., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., and A. Sali. (2000) in *Current Protocols in Bioinformatics* pp 5.6.1-5.6.30, John Wiley & Sons, Inc.

2.46. Marti-Renom, M. A., Stuart, A., Fiser, A., Sanchez, R., Melo, F., and A. Sali. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct. 29*, 291-325.

2.47. Case, D. A., Darden, T. A., Cheatham, I. T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and P. A. Kollman. (2006), University of California, San Francisco.

2.48. Kollman, P. A., Massova, I., Reyes, C. M., Kuhn, B., Huo, S., Chong, L. T., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and T. E. Cheatham. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res. 33*, 889-897.

2.49. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and D. A. Case. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc. 120*, 9401-9409.

2.50. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and C. Simmerling. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins 65*, 712-725.

2.51. Onufriev, A., Bashford, D., and D. A. Case. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct. Funct. Bioinformatics 55*, 383-394.

2.52. Xiang, Z., and B. Honig. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol. 311*, 421-430.

2.53. Buckle, A. M., Schreiber, G., and A. R. Fersht. (1994) Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. *Biochemistry 33*, 8878-8889.

2.54. Archer, D. G., and P. Wang. (1990) The dielectric constant of water and debye-huckel limiting low slopes. *J. Phys. Chem. Ref. Data 19*, 371-411.

2.55. Ryckaert, J. P., Ciccotti, G., and H. J. C. Berendsen. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327-341.

Chapter 3

3.1. Radley, T.L., et al. (2003) Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol. 332*, 529-536.

3.2. Cutler, T.A., et al. (2009) Effect of interdomain linker length on an antagonistic folding-unfolding equilibrium between two protein domains. *J. Mol. Biol. 386*, 854-868.

3.3. Ostermeier, M. (2005) Engineering allosteric protein switches by domain insertion. *Protein Engineering, Design & Selection 18*, 359-364.

3.4. Buskirk, A.R., et al. (2004) Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc. Natl. Acad. Sci. USA 101*, 10505-10510.

3.5. Dueber, J.E., et al. (2004) Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. *Curr. Opin. Struct. Biol. 14*, 2004. 690-699.

3.6. Villaverde, A. (2003) Allosteric enzymes as biosensors for molecular diagnosis. *FEBS Lett. 554*, 169-172.

3.7. Lim, W.A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol. 12*, 61-68.

3.8. Siegel, M.S. and E.Y. Isacoff. (1997) A genetically encoded optical probe of membrane voltage. *Neuron 19,* 735-741.

3.9. Doi, N. and H. Yanagawa. (1999) Insertional gene fusion technology. *FEBS Lett. 457,* 1-4.

3.10. Guntas, G. and M. Ostermeier. (2004) Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol. 336,* 263-273.

3.11. Doi, N. and H. Yanagawa. (1999) Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett. 453,* 305-307.

3.12. Baird, G.S., D.A. Zacharias, and R.Y. Tsien. (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. USA 96,* 11241-11246.

3.13. Skretas, G. and D.W. Wood. (2005) Regulation of protein activity with small-molecule-controlled inteins. *Protein Sci. 14,* 523-532.

3.14. Ha, J.H., et al. (2006) Modular enzyme design: regulation by mutually exclusive protein folding. *J. Mol. Biol. 357,* 1058-1062.

3.15. Cutler, T.A. and Loh, S. N. (2007) Thermodynamic analysis of an antagonistic folding-unfolding equilibrium between two protein domains. *J. Mol. Biol. 371,* 308-316.

3.16. Go, N. (1983) Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng. 12,* 183-210.

3.17. Das, P., et al. (2005) Characterization of the folding landscape of monomeric lactose repressor: quantitative comparison of theory and experiment. *Proc. Natl. Acad. Sci. USA 102,* 14569-14574.

3.18. Clementi, C., H. Nymeyer, and J.N. Onuchic. (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,* 937-953.

3.19. Elcock, A.H. (2006) Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol. 2,* e98.

3.20. Chavez, L.L., J.N. Onuchic, and C. Clementi. (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc. 126*, 8426-8432.

3.21. Koga, N. and S. Takada. (2001) Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol. 313*, 171-180.

3.22. Frembgen-Kesner, T. and A. Elcock. (2009) Striking Effects of Hydrodynamic Interactions on the Simulated Diffusion and Folding of Proteins. *J. Chem. Theory Comput. 5*, 242-256.

3.23. Mauguen, Y., et. al. (1982) Molecular structure of a new family of ribonucleases. *Nature 297*, 162-164.

3.24. Vijay-Kumar, S., C.E. Bugg, and W.J. Cook. (1987) Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol. 194*, 531-544.

3.25. Buckle, A.M., G. Schreiber, and A.R. Fersht. (1994) Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. *Biochemistry 33*, 8878-8889.

3.26. Jha, A.K., et al. (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA 102*, 13099-13104.

3.27. Knott, M., H.s. Kaya, and H.S. Chan. (2004) Energetics of protein thermodynamic cooperativity: contributions of local and nonlocal interactions. *Polymer 45*, 623-632.

3.28. Ermak, D.L. and J.A. McCammon. (1978) Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys. 69*, 1352-1360.

3.29. Berk Hess, H.B., Herman J. C. Berendsen, Johannes G. E. M. Fraaije. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem. 18*, 1463-1472.

3.30. Agashe, V.R. and J.B. Udgaonkar. (1995) Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry 34*, 3286-3299.

3.31. Ferrenberg, A.M. and R.H. Swendsen (1988), New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett*. 61, 2635-2638.

3.32. Ferrenberg, A.M. and R.H. Swendsen (1989), Optimized Monte Carlo data analysis. *Phys. Rev. Lett*. 63, 1195-1198.

Chapter 4

4.1 Bishop, A.C., Buzko, O. and K.M. Shokat (2001) Magic bullets for protein kinases. *Trends Cell Biol. 11*, 167-172.

4.2 Radley, T.L., et al. (2003) Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol. 332*, 529-536.

4.3 Liu, Y. and D. Eisenberg. (2002) 3D domain swapping: as domains continue to swap. *Protein Sci. 11*, 1285-1299.

4.4 Chen, Y. and M.D. Barkley. (1998) Toward understanding tryptophan fluorescence in proteins. *Biochemistry 37*, 9976-9982.