JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

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 and Molecular Biology, SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, USA*

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

Edited by D. Case

Fusion of one protein domain with another is a common event in both evolution and 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 (Ub) into the bacterial ribonuclease barnase (Bn), using peptide linkers from zero to 10 amino acids each. ΔG_X values are determined by measuring the extent to which Co^{2+} binding to an engineered site on the Ub domain destabilizes the Bn 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 fewer 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 Bn domain, which relieves conformational strain. The simulations suggest that an effective route to mechanical unfolding begins with disruption of the hydrophobic core of Bn near the Ub insertion site.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: mutually exclusive; barnase; ubiquitin; coupled; molecular dynamics

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, bifunctional 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-terminal 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,

Corresponding authors. E-mail addresses: ltchong@pitt.edu; lohs@upstate.edu.

Abbreviations used: Bn, barnase; Ub, ubiquitin; BU, barnase–ubiquitin fusion protein; WT, wild type; GdnHCl, guanidine hydrochloride; LD, Langevin dynamics; RMSF, RMS fluctuation; GB, generalized Born; MM-GBSA, molecular mechanics generalized Born surface area.

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.[1]

We previously characterized the mutually exclusive folding mechanism by inserting ubiquitin (Ub) into Bn to create the barnase–ubiquitin (BU) fusion protein.[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 (Fig. 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 Fig. 1b. The antagonistic interaction is parameterized by a coupling free-energy term ΔG_X.[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
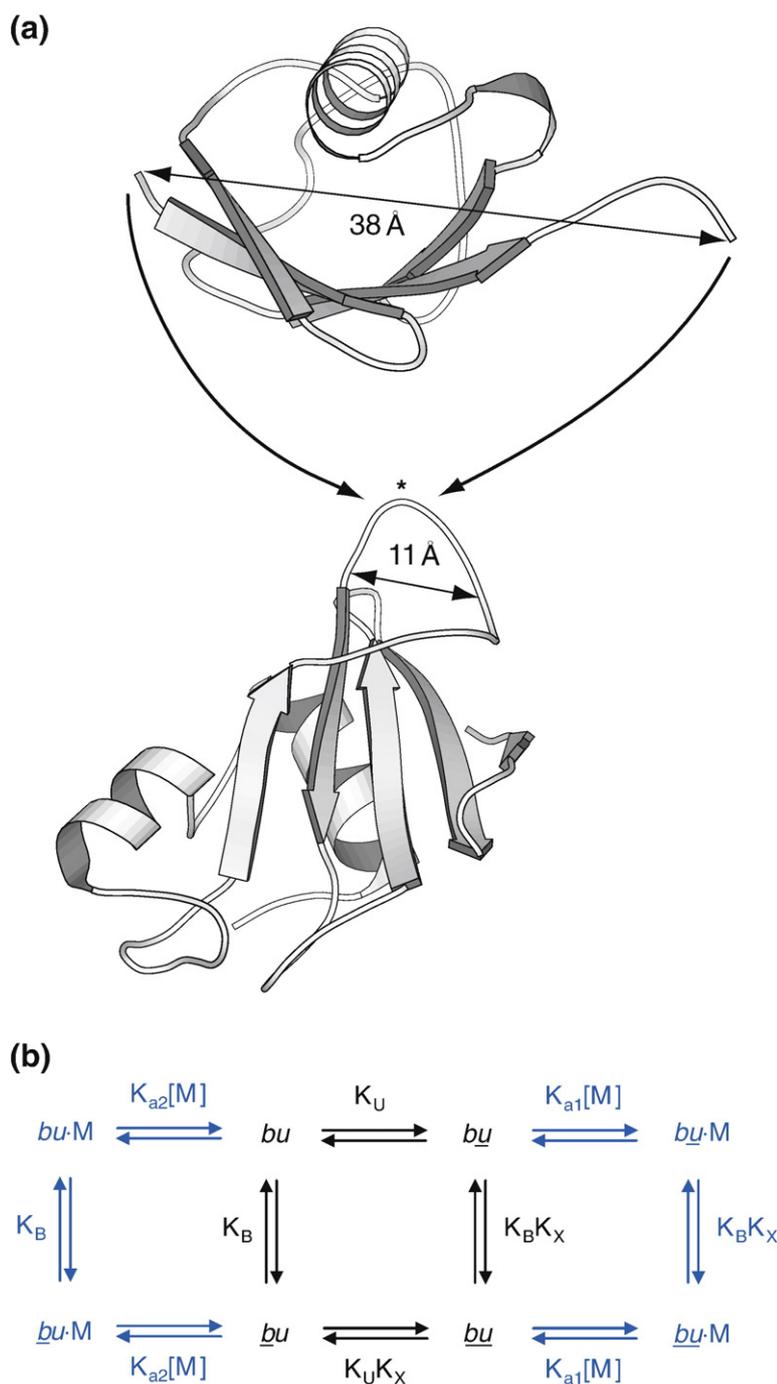
(a)

(b)

Fig. 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 (Pro64-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: $\Delta G_U = \Delta G_U^{H_2O} - m_U[\text{GdnHCl}]$ and $\Delta G_B = \Delta G_B^{H_2O} - m_B[\text{GdnHCl}]$, where $\Delta G_U^{H_2O}$ and $\Delta G_B^{H_2O}$ 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.[42] K_X is the equilibrium constant for coupling of the Bn and Ub domains, where $\Delta G_X = -RT \cdot \ln K_X$.

stability of Bn (ΔG_B) or of Ub (ΔG_U), 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 ($0 < \Delta G_X < \Delta G_B$, ΔG_U), and strong ($\Delta G_X > \Delta G_B$, ΔG_U). 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.

In an earlier study, we estimated the thermodynamic parameters of the mechanism by introducing destabilizing mutations into the Ub domain.[3] The stability of the Bn domain was inversely proportional to the stability of the Ub domain, as specified by the coupled equilibria in Fig. 1b. Fitting the experimental data to the model yielded $\Delta G_B{}^{H_2O} = 7.5$ kcal/mol, $\Delta G_U{}^{H_2O} = 5.2$ kcal/mol, and $\Delta G_X \sim 4$ kcal/mol 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 10 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 $\Delta G_B{}^{H_2O}$ and $\Delta G_U{}^{H_2O}$ when the linkers are fewer than approximately two amino acids each. To measure ΔG_X, we introduce a bi-His metal-binding site into the Ub domain via the K6 H mutation.[4,5] 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.[3] Metal binding is the preferred method to stabilize Ub, as the mutations known to increase stability do so by optimizing surface electrostatics,[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 Fig. 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.[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,[8–10] 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.[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.[11–14] 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.

Results

Thermodynamic characterization of linker length variants

GdnHCl and temperature-induced denaturation curves for BU variants are shown in Fig. 2. CD ellipticity at 230 nm reveals two GdnHCl-induced transitions (Fig. 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 (Fig. 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.

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})
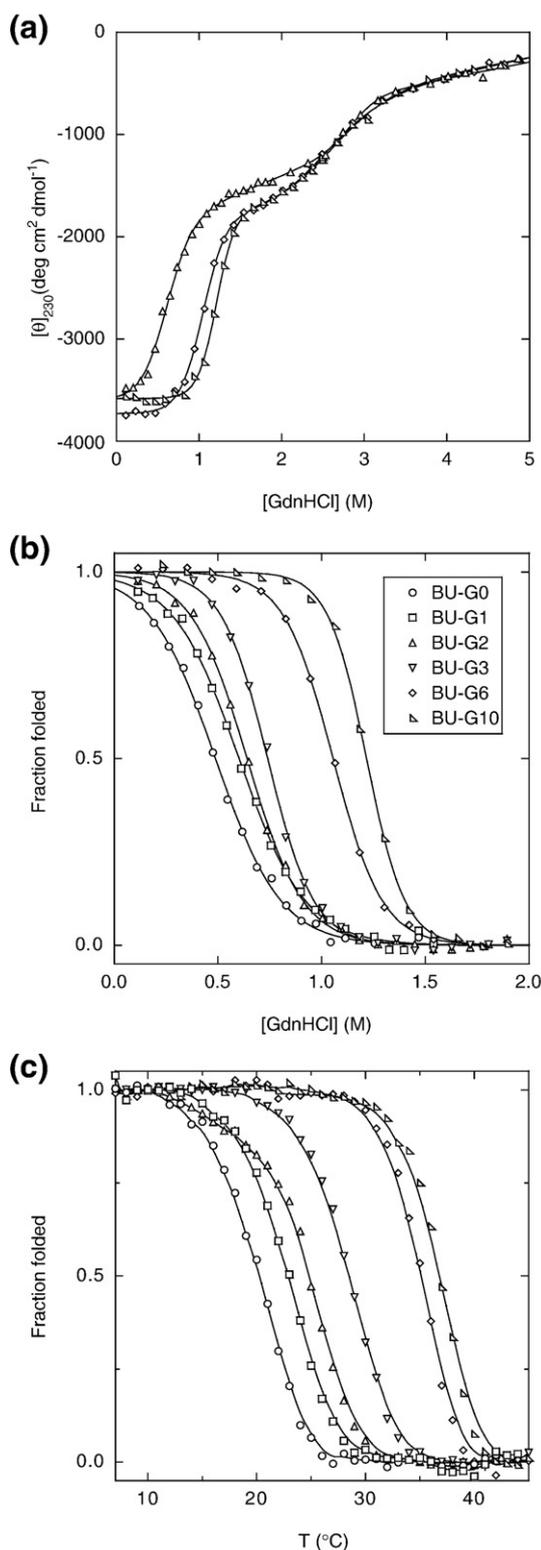
increases from 336 to 355 nm upon Bn unfolding.[2] Thus, F_{max} reports primarily on Bn conformation. Data for all variants are fit adequately by the two-state linear extrapolation equation (Fig. 2b). The resulting parameters are listed in Table 1. BU-G6 is less stable than BU-G10 as judged by both ΔG^{H_2O} 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.

We monitored thermal denaturation by CD to further characterize Bn domain stability (Fig. 2c). At pH 7.5, free Bn exhibits a melting temperature (T_m) of 51.5 °C (Table 1) and free Ub does not denature below 100 °C.[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.

Table 1. Thermodynamic parameters for unfolding of free Bn, free Ub, and the Bn domain of BU variants

Variant	Co^{2+} (mM)	ΔG^{H_2O} (kcal/mol)	m (kcal mol^{-} M^{-})	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]

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.

Fig. 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 (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 (b).

Co²⁺ 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 $RT \ln[(1+K_{a1})/(1+K_{a2})]$ when binding to both native and unfolded Ub is saturated, where K_{a1} and K_{a2} are the association constants for the respective forms (Fig. 1b). Krantz *et al.* report K_{a1} and K_{a2} values in the range of $\sim 10^4$ and $\sim 10^3$ M^{-1}, respectively, for bi-His Ub variants.[4] Consistent with those figures, we observe that $\Delta\Delta G^{H_2O}$ of K6H Ub reaches a maximum value of 1.4 kcal/mol at Co²⁺ concentrations greater than ~ 1 mM (Fig. 3a and Table 1). By contrast, 10 mM Co²⁺ 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 Co²⁺ (Table 1).

Having established that metal binds to K6H Ub and not to Bn, we next repeated the GdnHCl denaturation experiments of Fig. 2b in the presence of 10 mM Co²⁺. The model predicts that Co²⁺ 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 3b shows that Co²⁺ 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 Co²⁺. The extent of Co²⁺-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, Co²⁺ has little effect on unfolding of BU-G1 and appears to increase the C_m of BU-G0 slightly (Fig. 3c).
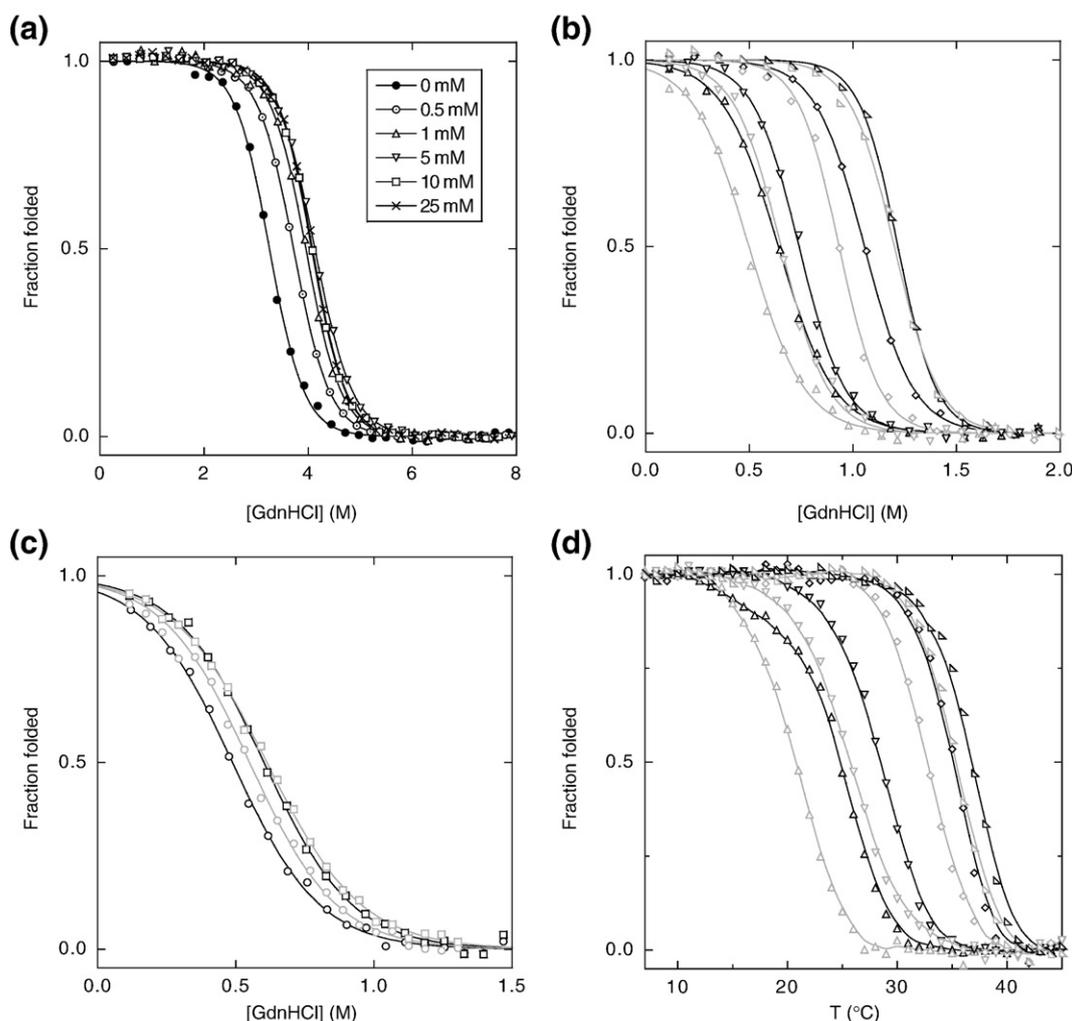

Fig. 3. Equilibrium denaturation curves of free Ub and BU variants in the presence of Co²⁺. Symbols in (b), (c), and (d) are the same as those in Fig. 2b. (a) GdnHCl-induced unfolding of free K6H Ub monitored by CD ellipticity at 225 nm. Co²⁺ concentrations are indicated in the inset. (b) GdnHCl-induced denaturation of BU-G10, BU-G6, BU-G3, and BU-G2 in the presence (gray) and absence (black) of 10 mM Co²⁺, monitored by Trp fluorescence. (c) GdnHCl-induced denaturation of BU-G1 and BU-G0 in the presence (gray) and absence (black) of 10 mM Co²⁺. (d) Thermal denaturation curves of BU variants in the presence (gray) and absence (black) of 10 mM Co²⁺. Lines in (a), (b), and (c) are best fits to the linear extrapolation equation.

Thermal denaturation curves (Fig. 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 $\geq 80\%$ reversible in all cases.

Structural characterization by CD

CD spectra of free Bn, free Ub, and BU variants are shown in Fig. 4a. The CD spectrum of free Bn is characterized by unusually low spectral intensities
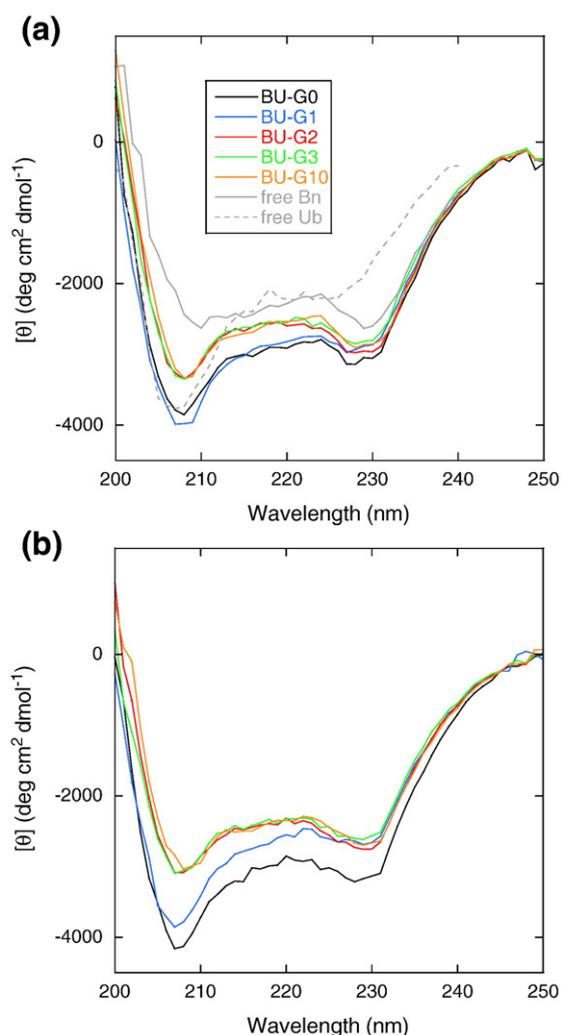

Fig. 4. Structural characterization of BU variants by CD in the (a) absence and (b) presence of 1 mM Co^{2+}. Protein concentrations are ~ 5 μM.

and an atypical minimum at 231 nm.[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. Krantz *et al.* reported that metal binding does not alter the far-UV CD spectrum of bi-His Ub variants.[4] Figure 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.

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. Representative data sets, showing quality of the fits, are available as electronic supplementary material (Fig. S3). As expected, BU-G10 sediments predominantly as a 21.8-kDa monomer (theoretical molecular mass, 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 molecular mass, but K_d decreases by nearly 10^3-fold to 6.0 μM. K_d further decreases (1.3 μM) in BU-G1, and the apparent molecular mass of the monomeric species increases to 35.5 kDa. 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 species of large molecular mass (data not shown). Thus, at the protein concentrations employed in fluorescence and CD experiments (1–2 μM), 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

Table 2. Apparent monomer molecular weights and dimer–monomer dissociation constants of BU variants, obtained from sedimentation equilibrium experiments

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

N.D., not determined due to poor fit to dimer–monomer model.

length is consistent with the anomalous Co^{2+} binding results observed for BU-G0 and BU-G1.

Computer simulations

The goal of the simulations is to characterize the mechanism of mechanical unfolding in conformationally strained BU variants. There are no existing experimental structures of BU from which starting models can be generated, particularly for strained conformations. 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 two domains were linked and orientations between them were explored exhaustively by MODELLER to identify the most favorable orientations based on the free-energy function (see Materials and Methods). In order to accommodate the insertion, Bn must stretch and/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-terminal to C-terminal distance in free Ub compresses to 35, 37, 35, 31, 27, and 25 Å for BU-G10, BU-G6, BU-G3, BU-G2, BU-G1, and BU-G0, respectively. In contrast, the C^{α}–C^{α} distance between the ends of the Bn surface loop (Pro64-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),[17,18] 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^{α} RMSDs 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 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 (Fig. 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 suggests 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).[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 (Fig. 5b and Table 3). 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 3). Comparing panel c with panel b in Fig. 5 reveals the extent of unfolding of BU-G2, the most strained monomeric variant, relative to the Bn_{cut} and Ub_{cut} controls. $Core_3$ exhibits large RMSDs, partially unfolding while the remaining hydrophobic cores remain intact (Table 3).

In addition to RMSD, it is informative to calculate RMS 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, 1.9 ± 0.6, and 1.7 ± 0.9 Å, respectively, averaged over the entire 50 ns (Fig. S1). $Core_2$ appears to be more flexible (RMSF$=3.2\pm1.0$ Å). Table 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.

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 6 shows snapshots of BU-G2 taken at
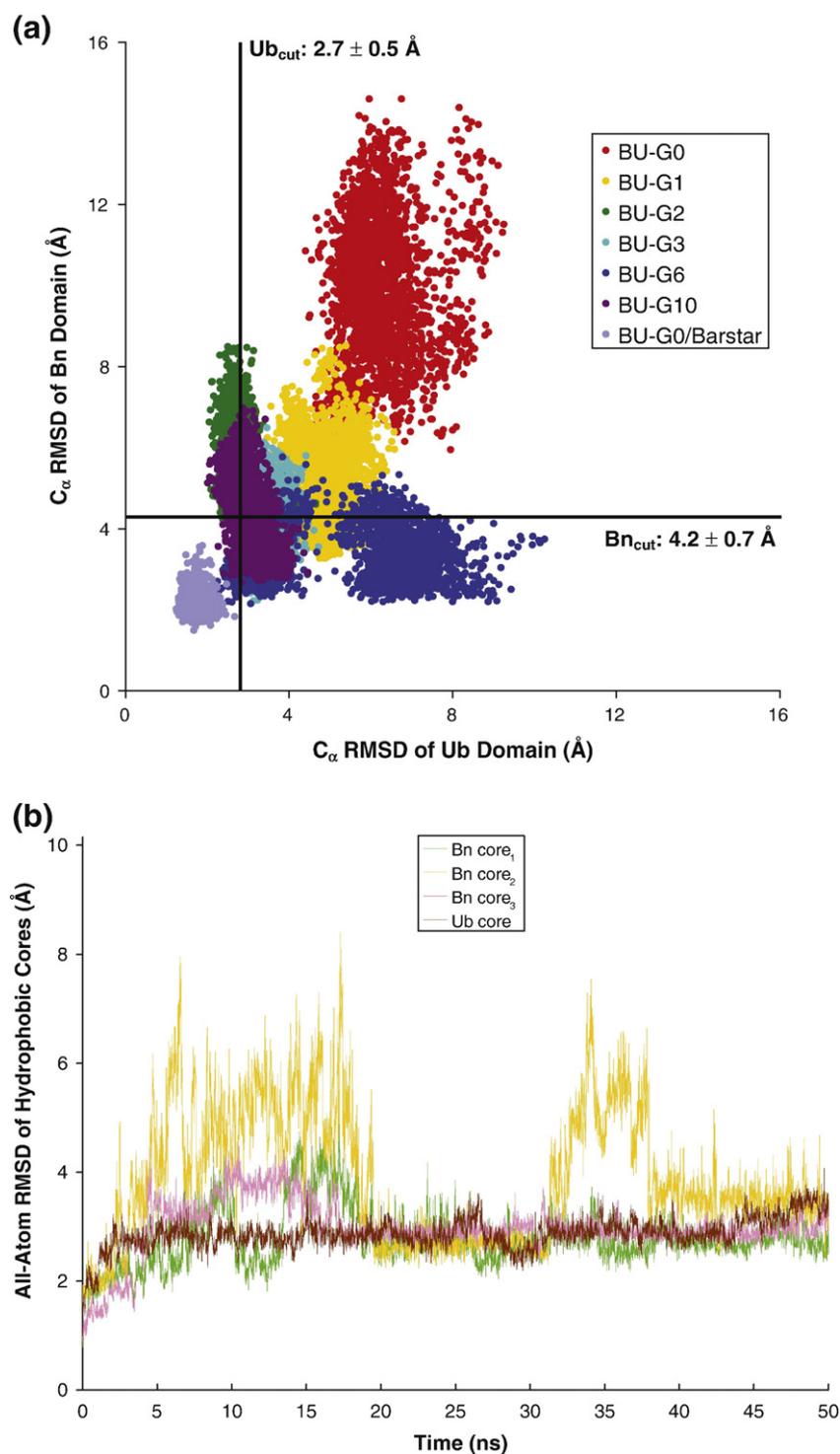
(a)

Ub_{cut}: 2.7 ± 0.5 Å

Bn_{cut}: 4.2 ± 0.7 Å

Legend:
- BU-G0
- BU-G1
- BU-G2
- BU-G3
- BU-G6
- BU-G10
- BU-G0/Barstar

C_α RMSD of Bn Domain (Å)

C_α RMSD of Ub Domain (Å)

(b)

All-Atom RMSD of Hydrophobic Cores (Å)

Legend:
- Bn $core_1$
- Bn $core_2$
- Bn $core_3$
- Ub core

Time (ns)

Fig. 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), which were left unrestrained during the generation of starting models (see Materials and 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.
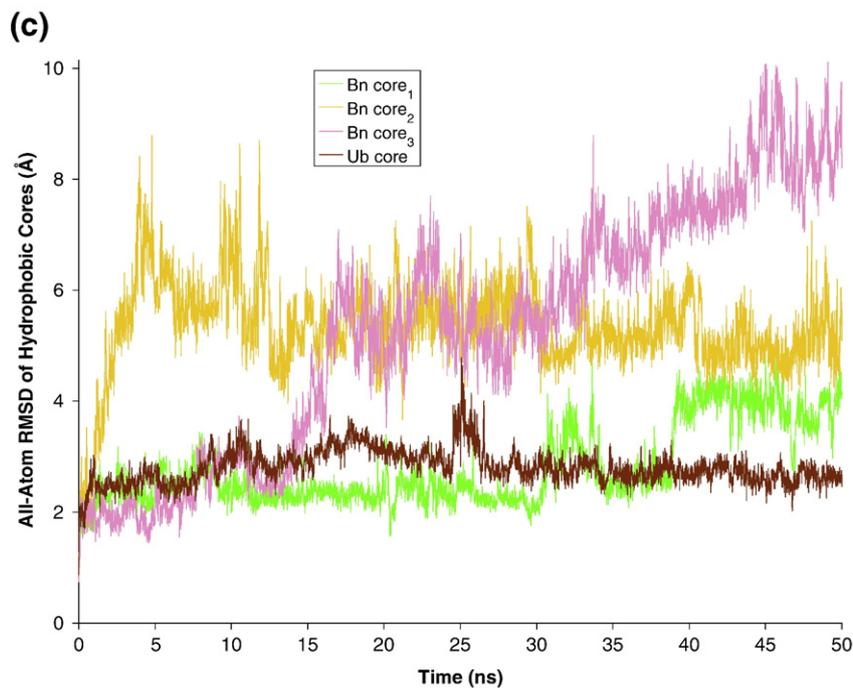
(c)

Fig. 5 (*legend on previous page*)

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 (Fig. 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).

Discussion

Thermodynamic model for antagonistic coupling

Our strategy for testing the mechanism of Fig. 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 $\Delta G_B^{H_2O}$ and $\Delta G_U^{H_2O}$ and mutually exclusive folding will be attained. ΔG_X is estimated from ΔG^{H_2O}, 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.

Table 3. Average C^α RMSDs of hydrophobic cores of Bn and Ub domains obtained from LD simulations, relative to the respective starting structures

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)

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$_{cut}$ or Ub$_{cut}$. N.A., not applicable.
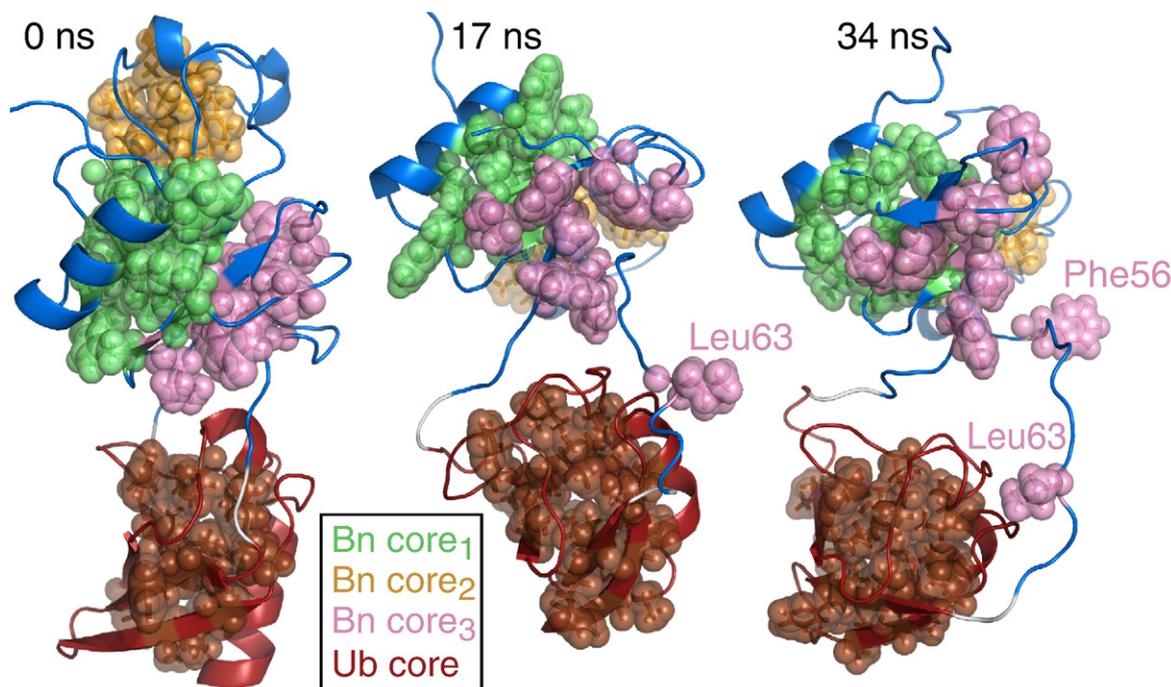
Fig. 6. Snapshots of BU-G2 taken at indicated times of simulation, illustrating the sequence of mechanically disrupted states.

Zero coupling

When $\Delta G_X = 0$, the Bn unfolding transition is $\underline{bu} \leftrightarrows \underline{bu}$ (Fig. 1). The observed m value is theoretically identical to that of free Bn. The observed ΔG^{H_2O} 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^{H_2O} and m values decrease to $(\overline{\Delta G_B}^{H_2O} - \Delta G_U^{H_2O})$ 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 $RT \ln[(1 + K_{a1})/(1 + K_{a2})]$ 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 $\underline{bu} \leftrightarrows \underline{bu}$ 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 (Fig. 2b). It is therefore appropriate to compare ΔG^{H_2O} 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.

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 1). Comparison of GdnHCl denaturation curves for BU-G10, BU-G6, BU-G2 and BU-G2 (Fig. 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 or 4.7 kcal mol^{-1} M^{-1} (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^{-1} M^{-1}. These findings agree with those of our earlier study, in which coupling was found to be in the high-intermediate range $(\Delta G_X \sim 4 \text{ kcal/mol})$ 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 (Fig. 3b). At a concentration of 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 $\Delta\Delta G^{H_2O}$ value of -0.9 kcal/mol is 64% of the theoretical value for strong coupling. Thermal denaturation curves follow a similar profile. At 10 mM concentration, 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 Fig. 1.

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 (Fig. 3c), 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). 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 nonswapped forms can be modulated. Kuhlman *et al.* induced protein L to form a swapped dimer by mutating three residues in a β-turn.[12] These mutations selectively destabilize the monomer by forcing turn residues to adopt unfavorable backbone dihedral angles. Byeon *et al.* induced dimerization of GB1 by positioning a bulky Phe residue in its hydrophobic core.[20] Steric repulsions were relieved by domain swapping. LeFevre and Cordes took the opposite approach and converted dimeric λ Cro protein to monomer by selectively stabilizing the latter.[21] Patel *et al.* proposed that cadherins use domain swapping to create homophilic binding interfaces that are highly specific, yet of low affinity.[22] 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 the 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.[23]

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).[24] 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,[24] 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 that 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.[1] WT Bn forms a domain-swapped trimer at high protein concentrations and moderately destabilizing conditions.[25] Numerous pairs of Bn fragments, generated by cleaving the protein at various surface loops, can bind and refold to form native-like complexes.[24] Thus, domain swapping is the explanation that best accounts for all of our present observations.

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[26] and opening/closing motions of the flaps in HIV protease.[27,28] In such cases, reducing viscosity accelerates reaction rates but does not significantly change the structures that are populated.[29] 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[30–34] and ion pair interactions.[33,35–38] 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.[30,33,39] 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 (Fig. 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.

Conclusions

We have demonstrated that the extent of inter-domain 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 fewer 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.

Materials and Methods

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 *Escherichia 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.[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- to 100-fold) into ice-cold buffer and allowed to equilibrate for >12 h at 4 °C.

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) and 0.1 M NaCl containing various amounts of GdnHCl. Those aliquots were prepared previously by mixing different ratios of buffer and 6 M GdnHCl with 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 data collection. Final protein concentration was 1–2 μM. GdnHCl concentrations were measured at the end of each experiment by index of refraction.[40] 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 an Aviv Model 202 spectropolarimeter. Data were fit to the linear extrapolation equation.[41]

CD wavelength scans were recorded at 10 °C using 5 μM 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.

Analytical ultracentrifugation experiments

Lyophilized protein was dissolved in 6 M urea and 10 mM Tris (pH 7.5) at concentrations of ~700, ~350 and ~175 μM. 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†. Monomer molecular mass and K_d were obtained from a global fit of the data from three different protein concentrations to a self-associating monomer–dimer model.

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 [Protein Data Bank (PDB) code 1A2P][42] and Ub (PDB code 1UBQ),[43] using the MODELLER 9v1 software package.[44,45] 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.[46] 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^α atoms of Bn residues 71 and 79 and Ub residue 4) and torsion between the domains (C^α

---

† http://www.analyticalultracentrifugation.com

atoms of Bn residues 17 and 74 and Ub residues 1 and 8) in 10° increments from 90° to 180° and from 10° to 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, molecular mechanics generalized Born surface area (MM-GBSA) free-energy calculations[47,48] were performed on the minimized model using the AMBER ff99SB force field,[49] a variation of the GB implicit solvent model by Onufriev *et al.* (igb = 5 in AMBER 9),[50] 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.[51] 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)[52] onto those of Bn in the BU-G0 model.

Computer simulations

Simulations starting from models of the strained BU variants were performed with the AMBER 9.0 software package.[46] 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 time scale, 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 328 K, using the solvent dielectric constant of water at that temperature ($\varepsilon = 68.3$).[53] To enable a 2-fs time step, bonds to hydrogen were constrained to their equilibrium values with the SHAKE algorithm.[54] 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, 1.0, and 0.1 kcal mol$^{-1}\cdot$Å^{-2}, respectively. After equilibration, fully unrestrained production simulations were carried out for 50 ns at 328 K. Each 50-ns simulation required approximately 1 month on a dual 2.66-GHz quad-core processor server, using all eight cores in parallel.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.10.090

References

1. Ha, J.-H., Butler, J. S., Mitrea, D. M. & Loh, S. N. (2006). Modular enzyme design: regulation by mutually exclusive protein folding. *J. Mol. Biol.* **357**, 1058–1062.
2. Radley, T. L., Markowska, A. I., Bettinger, B. T., Ha, J.-H. & Loh, S. N. (2003). Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol.* **332**, 529–536.
3. Cutler, T. & Loh, S. N. (2007). Thermodynamic analysis of an antagonistic folding–unfolding equilibrium between two protein domains. *J. Mol. Biol.* **371**, 308–316.
4. Krantz, B., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using ψ-analysis. *J. Mol. Biol.* **337**, 463–475.
5. Sosnick, T. R., Dothager, R. S. & Krantz, B. (2004). Differences in the folding transition state of ubiquitin indicated by ϕ and ψ analyses. *Proc. Natl Acad. Sci. USA*, **101**, 17377–17382.
6. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein via optimization of charge–charge interactions on a protein surface. *Biochemistry*, **38**, 16419–16423.
7. Teichmann, S. A., Chothia, C. & Gerstein, M. (1999). Advances in structural genomics. *Curr. Opin. Struct. Biol.* **9**, 390–399.
8. Holm, L. & Sander, C. (1994). The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res.* **22**, 3600–3609.
9. Jones, S., Stewart, M., Michie, A., Swindells, M. B., Orengo, C. & Thornton, J. M. (1998). Domain assignment for protein structures using a consensus approach: characterization and analysis. *Protein Sci.* **7**, 233–242.
10. Aroul-Selvam, R., Hubbard, T. & Sasidharan, R. (2004). Domain insertions in protein structures. *J. Mol. Biol.* **338**, 633–641.
11. Liu, Y. & Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285–1299.
12. Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y. & Baker, D. (2001). Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc. Natl Acad. Sci. USA*, **98**, 10687–10691.
13. Newcomer, M. E. (2002). Protein folding and three-dimensional domain swapping: a strained relationship? *Curr. Opin. Struct. Biol.* **12**, 48–53.
14. Patel, S. D., Chen, C. P., Bahna, F., Honig, B. & Shapiro, L. (2003). Cadherin-mediated cell–cell adhesion: sticking together as a family. *Curr. Opin. Struct. Biol.* **13**, 690–698.
15. Lazar, G. A., Desjarlais, J. R. & Handel, T. M. (1997). De novo design of the hydrophobic core of ubiquitin. *Protein Sci.* **6**, 1167–1178.
16. Vuilleumier, S., Sancho, J., Loewenthal, R. & Fersht, A. R. (1993). Circular dichroism studies of barnase and its mutants: characterization of the contribution of aromatic side chains. *Biochemistry*, **32**, 10303–10313.

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

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

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

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

21. LeFevre, K. R. & Cordes, M. H. J. (2003). Retro-evolution of λ Cro toward a stable monomer. *Proc. Natl Acad. Sci. USA*, **100**, 2345–2350.

22. Patel, S., Ciatto, C., Chen, C. P., Bahna, F., Rajebhosale, M., Arkus, N. *et al.* (2006). Type II cadherin ectodomain structures: implications for classical cadherin specificity. *Cell*, **124**, 1255–1268.

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

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

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

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

27. Hornak, V., Okur, A., Rizzo, R. C. & Simmerling, C. (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.

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

29. Walser, R. & van Gunsteren, W. F. (2001). Viscosity dependence of protein dynamics. *Proteins: Struct. Funct. Genet.* **42**, 414–421.

30. Nymeyer, H. & Garcia, A. E. (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.

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

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

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

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

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

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

37. Felts, A. K., Harano, Y., Gallicchio, E. & Levy, R. M. (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.

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

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

40. Pace, C. N. & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach* (Creighton, T. E., ed.), chapt. 12, pp. 229–320, Oxford University Press, New York.

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

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

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

44. Eswar, N., Marti-Renom, M. A., Webb, B., Madhusudhan, M. S., Eramian, D., Shen, M. *et al.* (2007). Modeling Structure From Sequence, Unit 5.6 Comparative Protein Structure Modeling Using Modeller. In *Current Protocols in Bioinformatics*, pp. 5.6.1–5.6.30, John Wiley & Sons, Inc.

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

46. Case, D. A., Darden, T. A., Cheatham, I. T. E., Simmerling, C. L., Wang, J., Duke, R. E. *et al.* (2006). University of California, San Francisco. http:// ambermd.org/.

47. Kollman, P. A., Massova, I., Reyes, C. M., Kuhn, B., Huo, S., Chong, L. T. *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897.

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

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

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

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

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

53. Archer, D. G. & Wang, P. (1990). The dielectric constant of water and Debye–Huckel limiting low slopes. *J. Phys. Chem. Ref. Data*, **19**, 371–411.

54. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (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.