

Molecular Simulations of Mutually Exclusive Folding in a Two-Domain Protein Switch

Brandon M. Mills and Lillian T. Chong*

Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT A major challenge with testing designs of protein conformational switches is the need for experimental probes that can independently monitor their individual protein domains. One way to circumvent this issue is to use a molecular simulation approach in which each domain can be directly observed. Here we report what we believe to be the first molecular simulations of mutually exclusive folding in an engineered two-domain protein switch, providing a direct view of how folding of one protein drives unfolding of the other in a barnase-ubiquitin fusion protein. These simulations successfully capture the experimental effects of interdomain linker length and ligand binding on the extent of unfolding in the less stable domain. In addition, the effect of linker length on the potential for oligomerization, which eliminates switch activity, is in qualitative agreement with analytical ultracentrifugation experiments. We also perform what we believe to be the first study of protein unfolding via progressive localized compression. Finally, we are able to explore the kinetics of mutually exclusive folding by determining the effect of linker length on rates of unfolding and refolding of each protein domain. Our results demonstrate that molecular simulations can provide seemingly novel biological insights on the behavior of individual protein domains, thereby aiding in the rational design of bifunctional switches.

INTRODUCTION

Protein conformational switches are one of the simplest molecular devices in nature, regulating a variety of biological processes. These switches are necessarily precise, adopting either “active” or “inactive” conformations in response to particular signals such as ligand binding (e.g., binding of the GTP ligand by GTPases) (1) and covalent modification (e.g., phosphorylation by kinases) (2). Understanding the mechanism of these switches is not only fundamental to biology, but could also be applied toward the design of artificial protein switches for a large number of applications, including biological imaging, biosensors, and therapeutic agents.

A number of strategies for designing protein conformational switches have yielded encouraging results (see (3–5) for reviews). A particularly elegant strategy is one pioneered by Radley et al. (6), Cutler and Loh (7), Cutler et al. (8), and Ha et al. (9) that involves the design of “mutually exclusive folding” via domain insertion, i.e., the insertion of a guest protein into the surface loop of a host protein. The sole requirement of this design is that the N- to C-terminal distance of the guest is much longer than the distance between the ends of the loop in the host (see Fig. 1 A). Fulfillment of this requirement leads to conformational strain in the

fusion protein such that the strain is expected to be only relieved through a thermodynamic “tug-of-war” between the proteins, where folding of one protein drives unfolding of the other; thus, the function of a protein is switched on or off by folding or unfolding, respectively. The state of the switch, as dictated by the tug-of-war, is controlled by factors that stabilize one protein over another (for example, mutations, ligand binding, or temperature). Although the mutually exclusive folding design and other related strategies (10–14) appear promising, the process of engineering an optimal two-protein fusion is still a major challenge; in particular, the choice of proteins, site of fusion, and addition of interdomain linker peptides are all critical for switch function. Moreover, it is not always possible to experimentally monitor the structural changes of each protein domain.

Computer simulations can be used to directly monitor the structural changes of each domain at the single-molecule level, potentially providing insights on switch optimization. Although all-atom simulations offer the most detailed view of protein dynamics, use of these simulations to fully explore the mechanism of mutually exclusive folding is currently computationally prohibitive (8). On the other hand, the use of residue-level simulations, along with a simple Go-type description of residue interactions (15,16) can generate a large ensemble of complete unfolding/refolding events within a week.

These types of simulations have been successfully used to model some key features of protein refolding events (see Clementi (17) for a review), providing a level of detail that may be sufficient for identifying the most promising fusion constructs. Here we explore the potential of these molecular simulations as virtual assays for switch activity (in this case, mutually exclusive folding) by focusing on

Submitted September 28, 2010, and accepted for publication December 17, 2010.

*Correspondence: ltchong@pitt.edu

This is an Open Access article distributed under the terms of the Creative Commons-Attribution Noncommercial License (<http://creativecommons.org/licenses/by-nc/2.0/>), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editor: Ruth Nussinov.

© 2011 by the Biophysical Society. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).
0006-3495/11/02/0756/9

doi: 10.1016/j.bpj.2010.12.3710

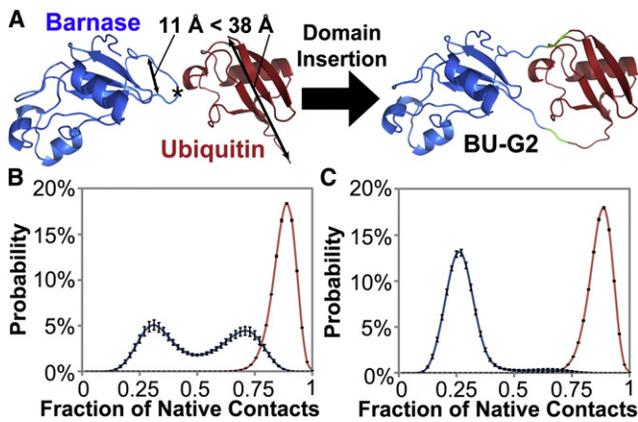


FIGURE 1 Effect of ubiquitin insertion on the stability of barnase in the BU-G2 fusion protein. (A) Design of mutually exclusive folding using barnase and ubiquitin as the host and guest proteins, respectively, for domain insertion. X-ray crystal structures of the host, barnase (blue) (18), and the guest, ubiquitin (red) (19), reveal that the C_{α} - C_{α} distance between the N- and C-termini of the guest is much longer than that between the ends of the insertion loop of the host (Pro⁶⁴ and Thr⁷⁰), thus fulfilling the requirement for mutually exclusive folding. The insertion site (between Lys⁶⁶ and Ser⁶⁷ of barnase) is indicated (asterisk). Interdomain glycine linkers in the BU-G2 fusion protein (shown in green). (B) Average distributions of the fraction of native contacts (Q) of free barnase and free ubiquitin (blue and red, respectively). (C) Average distributions of the fraction of native contacts of the barnase and ubiquitin domains of BU-G2 (blue and red, respectively). Distributions were determined from each of 10 independent 10- μ s simulations, then averaged, with error bars representing 1 SD ($N = 10$).

a set of barnase-ubiquitin (BU) fusion proteins where the toxic activity of the barnase domain is turned off by unfolding the domain. These simulations are the first, to our knowledge, to provide a complete molecular view of mutually exclusive folding.

METHODS

The protein model

All proteins were modeled at the residue level, with each residue represented by a pseudo-atom at the position of its C_{α} atom. C_{α} -models of both the folded and unfolded states of proteins were generated as starting conformations for simulation. Coordinates for the folded states were taken from x-ray crystal structures (PDB codes 1A2P (18), 1UBQ (19), and 1BRS (20), for barnase, ubiquitin, and barstar, respectively); in the case of the BU-G2/barstar complex, barstar was docked into the binding site of the barnase domain according to the crystal structure of the barnase-barstar complex (20). Coordinates for the unfolded states were taken from statistical coil conformations that were generated by the Unfolded State Server (<http://godzilla.uchicago.edu/cgi-bin/unfolded.cgi>) (21).

The conformational dynamics of the protein models are governed by a Gō-type potential energy function (15,16), in which bonded interactions between residues are modeled by standard molecular mechanics terms:

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

where r , θ , and φ are pseudo-bond lengths, pseudo-angles, and pseudo-dihedrals, respectively, and V_1 and V_3 are potential barriers for the dihedral terms. Equilibrium bond lengths r_{eq} , angles θ_{eq} , and dihedral phase angles, φ_1 and φ_3 , were taken from the crystal structures mentioned above. The force constants k_{bond} and k_{angle} were set to 20 kcal/mol/Å and 10 kcal/mol/rad, respectively.

Nonbonded interactions between residues (separated by four or more pseudo-bonds) were modeled using one of two different interaction potentials: a Lennard-Jones-like potential for residue-residue contacts that are present in the native, folded state and a purely repulsive potential for nonnative contacts. Native contacts were modeled using the potential

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

where ϵ^{native} is the energy well depth for the native interaction, r_{ij} is the distance between residues i and j in the simulation, and σ_{ij}^{native} is the distance between the C_{α} -atoms of residues i and j in the corresponding crystal structure of the native state. Two residues were considered to form a native contact if any of their heavy atoms are within 5.5 Å of each other in the crystal structure of the folded protein. The number of native contacts for barnase, ubiquitin, and barstar, were 335, 226, and 272, respectively; 98 native contacts between barnase and barstar were included for the barnase-barstar complex. Nonnative contacts were modeled using the potential

$$E_{ij}^{nonnative} = \epsilon^{nonnative} \left[\frac{\sigma_{ij}^{nonnative}}{r_{ij}} \right]^{12},$$

where $\epsilon^{nonnative}$ is set to 0.60 kcal/mol and $\sigma_{ij}^{nonnative}$ is set to 4.0 Å. These values were chosen so that the repulsive potential of a nonnative contact is of a similar magnitude as the attractive potential of a native contact. All contacts between the barnase and ubiquitin domains as well as those involving linker peptides were considered nonnative. For simplicity, we refer to ϵ^{native} as simply ϵ for the remainder of this article.

Parameterization of the model

We reproduced the experimental T_m values of each protein by optimizing the primary adjustable parameter in our Gō-type model: the well-depth ϵ for the potential of interaction between two residues that form a native contact. The energetic parameters ϵ , V_1 , and V_3 , were initialized to 0.57, 0.475, and 0.2375, respectively, because these values reproduce the experimental standard unfolding free energy of barnase (22). To determine the optimal parameters for each protein, we scaled these initial values until equal populations of unfolded and folded conformations were sampled in ten 10- μ s simulations. A single scaling factor was used because the energetic balance between nonlocal (controlled by ϵ) and local interactions (controlled by V_1 and V_3) can influence the cooperativity of folding equilibria simulated with Gō-type models (23). All dihedrals involving linker peptides were allowed to freely rotate by setting both V_1 and V_3 to zero. Optimal ϵ , V_1 , and V_3 values along with the corresponding free energy profiles for each protein are provided in the Supporting Material (Table S1 and Fig. S1). To model the effects of barstar binding to the barnase domain of BU-G2, we used a very large ϵ -value of 1.2 for the native contacts between barnase and barstar; the effects of other ϵ -values, i.e., 0.8 and 1.0 kcal/mol, are reported in Fig. S2.

Simulation details

All simulations were performed at the T_m of barnase (51.5°C) using a standard Brownian dynamics algorithm developed by Ermak and McCammon (24) with hydrodynamic interactions (25). To accelerate protein unfolding and refolding events, we used a small value of 1.5 Å for the hydrodynamic radii. A time step of 50 fs was used, constraining pseudo-bonds between

residues to their native bond lengths using the LINC algorithm (26). Nonbonded interactions were calculated only if r_{ij} was $< \sigma_{ij}^{\text{native}} + 6 \text{ \AA}$ for native contacts or 10 \AA for nonnative contacts; the list of pairwise interactions was updated every 20 time-steps.

Each simulation was carried out for $10 \mu\text{s}$, requiring ~ 4 days on a single core of a 2.66-GHz quad-core processor. To avoid bias toward the starting conformation, the first $0.25 \mu\text{s}$ of each simulation was omitted from analysis. All analysis was then performed on the remainder of the simulation, sampling conformations every 50 ps. The extent of folding of a protein at any point in the simulation was quantified using the fraction of native residue pairs (Q) that are in contact (i.e., within a distance of $1.2\sigma_{ij}^{\text{native}}$) (27,28). All simulations were converged, resulting in the same conformational distributions regardless of starting conformation (folded or unfolded).

As done by others (29–31), we performed potential domain-swapping simulations for each protein using the Gō-type potential energy function described above. The same simulation protocol as described above was used with the following modifications:

1. Among the two molecules of the protein, only the closer of the intra- and intermolecular versions of each native contact was treated as an attractive contact (with the other treated as repulsive).
2. A weak spherical confining potential with a harmonic spring constant of $1.0 \text{ kcal/mol/\AA}^2$ and 100 \AA radius was applied throughout to enable frequent collisions between the two molecules.

Fifty independent simulations were performed, starting from each of 50 randomly placed pairs of molecules, separated by at least 30 \AA . Each of these simulations was carried out for $2 \mu\text{s}$, requiring ~ 2 days on a single core of a 2.66-GHz quad-core processor. Analysis was performed on the latter $1.75 \mu\text{s}$ of each simulation, sampling conformations every 50 ps. Unfolded free energies were calculated from each of our simulations using $-RT \ln(N_{\text{unfold}}/N_{\text{fold}})$ where N_{unfold} and N_{fold} are the numbers of unfolded and folded conformations, respectively. Definitions of the unfolded and folded states ($Q_{\text{cross}} < 0.40$ and $Q_{\text{cross}} \geq 0.40$, respectively, where Q_{cross} indicates contacts between residues on opposite sides of the insertion site (see Fig. 1 A)) were taken from a free energy profile based on simulations of barnase (Fig. S1). The free energy of dimerization was estimated using

$$\Delta G_{\text{dimerize}} = -RT \ln \left[N_{\text{assoc}}^{\text{dimer}} / \left(N_{\text{assoc}}^{\text{mon}} \right)^2 \right],$$

where $N_{\text{assoc}}^{\text{dimer}}$ and $N_{\text{assoc}}^{\text{mon}}$ are the numbers of dimeric and monomeric conformations, respectively, with folded barnase domains. The barnase domain was considered folded if the fraction of native contacts between its segments before and after the point of ubiquitin insertion (residues 1–66 and 67–110, respectively) is > 0.40 , as determined from the free energy profile based on simulations of barnase (Fig. S1). The ensemble of 50 simulations for each protein provided converged free energies of dimerization, with standard deviations within 0.2 kcal/mol .

RESULTS AND DISCUSSION

Reproducing protein stabilities

An essential prerequisite to using molecular simulations to study conformational switching events is that the simulations reproduce known thermodynamic data on the switch's component parts. Thus, to simulate switching events of BU fusion proteins, we must first reproduce the stabilities of barnase and ubiquitin. In particular, we parameterized our molecular simulations to reproduce the experimental melting temperature (T_m) of each protein (8,32) (see Methods). Throughout this work, we assume that protein unfolding or refolding events can be adequately described

by the fraction of native contacts (Q), which are residue-residue contacts present in the folded conformation.

To sample both unfolding and refolding events of the barnase domain, all subsequent simulations were performed at the T_m value of barnase, 51.5°C ; however, full temperature scans (10 – 60°C) of all systems in this study are provided (see Fig. S3). We note that a major limitation of Gō-type energy functions is the neglect of stabilizing nonnative interactions, which is known to result in artificial folding mechanisms (27,28,33). Therefore, we focus on qualitative rather than quantitative comparisons with experiment. Although our simplified Gō-type description of residue interactions utilizes a number of approximations (see Approximations of the Simulation Model in the Supporting Material), our parameterized model may be sufficient for providing useful qualitative insights, e.g., ranking by potential switch activity.

Direct simulation of the tug-of-war

To determine the effect of domain insertion on the stability of barnase, we first performed 10 independent simulations each of barnase, ubiquitin, and BU-G2, which is the most conformationally strained, monomeric BU fusion protein (the term “G2” referring to interdomain linker peptides of two Gly residues each) (7). In their free states, barnase forms equal populations of unfolded and folded states whereas the much more stable ubiquitin is always folded (Fig. 1 B). Once the proteins are fused together, the ubiquitin domain remains folded during the simulation, whereas the barnase domain is almost always unfolded, with a small probability ($\sim 2\%$) of transiently refolding (Fig. 1 C). The insertion of ubiquitin into barnase therefore destabilizes barnase, which is consistent with thermodynamic parameters obtained from both GdnHCl and thermal denaturation experiments (8).

To directly test for mutually exclusive folding, we simulated the folding of the ubiquitin domain in the context of BU-G2 to see whether it drives unfolding of the barnase domain. Fifty such simulations were started from a conformation in which the ubiquitin domain is unfolded and the barnase domain is folded; as a control, an equal number of simulations were performed from the same conformation, but in the absence of conformational strain, with the ubiquitin domain artificially maintained in an unfolded conformation.

To obtain an ensemble-averaged view of the resulting dynamics, we monitored the average fraction of native contacts for each domain versus time. As shown in Fig. 2 A, the rapid refolding of the ubiquitin domain causes the barnase domain to be slightly more unfolded than it is in the control simulations (average Q -value of 0.28 ± 0.04 compared to $Q = 0.32 \pm 0.06$ for the control; error bars represent 1 SD), indicating a modest role of conformational strain in unfolding the barnase domain. This only modest role is likely due to the fact that the domains of BU-G2 are

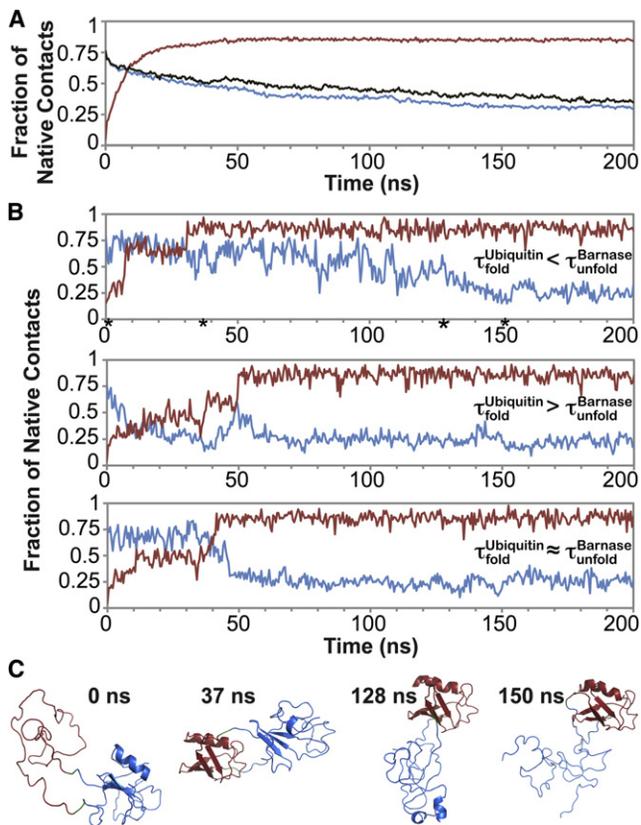


FIGURE 2 Effect of ubiquitin refolding on the stability of the barnase domain in the BU-G2 fusion protein. (A) Average fraction of native contacts as a function of time for the barnase and ubiquitin domains (blue and red, respectively) from 50 independent 1- μ s simulations involving the refolding of the ubiquitin domain starting from a conformation of the BU-G2 fusion protein in which the barnase domain is folded and the ubiquitin domain is unfolded. Additionally, the average fraction of native contacts as a function of time for the barnase domain from an equal number of control simulations in which the ubiquitin domain is artificially maintained in an unfolded conformation is provided (black). (B) Fraction of native contacts as a function of time for the barnase and ubiquitin domains (blue and red, respectively) for representative simulations of (top) ubiquitin refolding before barnase unfolds, (middle) ubiquitin refolding after barnase unfolds, and (bottom) simultaneous ubiquitin refolding and barnase unfolding. (C) Snapshots of conformations at the times highlighted with asterisks in panel B. Protein backbones were constructed from C_{α} coordinates using the BBQ program (45). Ribbon diagrams were created using PyMOL (46).

only coupled to an intermediate extent (8). Our simulations suggest that, with more closely coupled domains, the folding of the ubiquitin domain could drive the unfolding of the barnase domain.

In addition to providing an ensemble-averaged view of the protein dynamics, our simulations reveal the diversity of individual, single-molecule events. In 78% of the simulations, the refolding of the ubiquitin domain occurs before unfolding of the barnase domain, which is pulled apart as a result of the conformational strain (Fig. 2, B and C). However, ubiquitin refolds after barnase unfolds in 16% of the simulations and simultaneously (within 1 ns) with barnase unfolding in the remaining 6%. Thus, our simula-

tions predict a small probability of observing the two latter events in single-molecule experiments.

Effects of interdomain linker length

The extent of mutually exclusive folding in the BU fusion protein depends on how closely the folding of one domain is coupled to unfolding of the other. However, if the two domains are too closely coupled, the conformational strain is relieved through dimerization, potentially through domain swapping, rather than “switching-off” the barnase domain by unfolding it (Fig. 3 A) (8). One way to modulate the

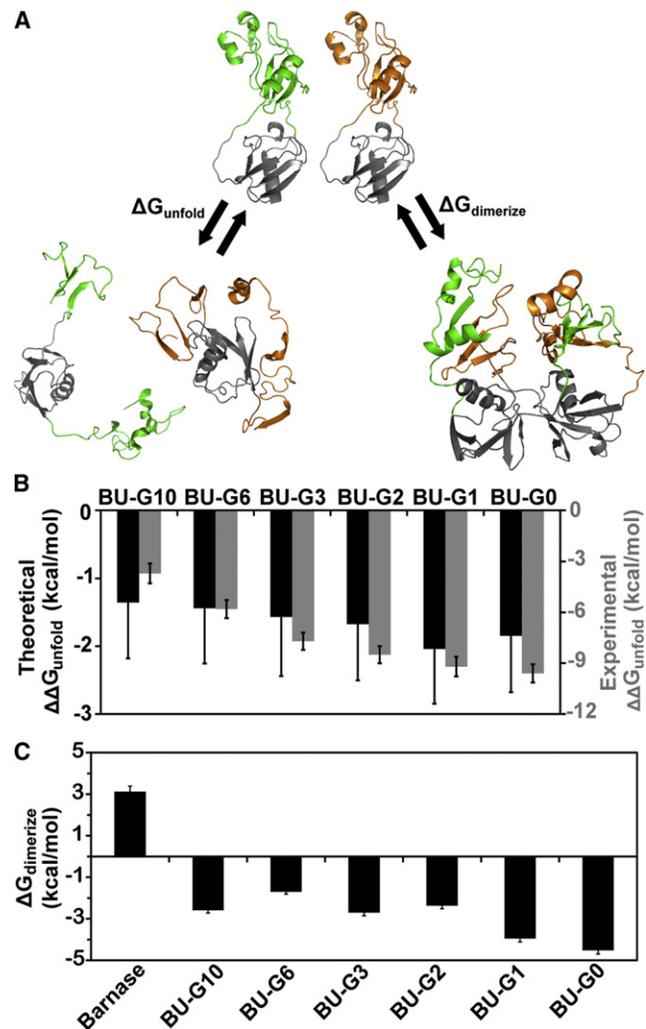


FIGURE 3 Effects of linker length on the stability and propensity of dimerization of the barnase domain in the BU fusion proteins. (A) Conformational strain can be relieved through either unfolding or dimerization, potentially through domain swapping, as illustrated with conformations from simulations of BU-G0 (see also Movie S1). (B) Average unfolding free energies of the barnase domain relative to that of free barnase ($\Delta\Delta G_{\text{unfold}}$) and (C) average dimerization free energy $\Delta G_{\text{dimerize}}$ of the barnase domain. For each protein, both $\Delta\Delta G_{\text{unfold}}$ and $\Delta G_{\text{dimerize}}$ values were computed using 50 independent 2- μ s simulations (see Methods). Experimental $\Delta\Delta G_{\text{unfold}}$ values due to GdnHCl denaturation (8) are provided in panel B as a qualitative comparison. Error bars represent 1 SD ($N \geq 3$ for the experimental $\Delta\Delta G_{\text{unfold}}$ values; and $N = 50$ for the theoretical $\Delta\Delta G_{\text{unfold}}$ and $\Delta G_{\text{dimerize}}$ values).

degree of coupling between domains is to introduce interdomain linker peptides. An important test of our simulations is to see, therefore, whether the effects of interdomain linker length on the switch activity of the BU fusion protein can be reliably reproduced. In particular, we performed simulations of the following BU fusions that contain interdomain linker peptides ranging from 0 to 10 Gly residues: BU-G0, BU-G1, BU-G2, BU-G3, BU-G6, and BU-G10. To enable dimerization events via domain swapping, simulations of each fusion protein involved two molecules of the protein at a high effective concentration (see [Methods](#)).

The ease of switching off the toxic activity of barnase in the BU fusion protein can be quantified as the unfolding free energy of the barnase domain relative to that of free barnase $\Delta\Delta G_{unfold}$ (see [Methods](#)). Interestingly, as shown in [Fig. 3 B](#), the insertion of the more stable ubiquitin domain drives the barnase domain toward the unfolded state ($\Delta\Delta G_{unfold} > 0$) even in the BU variant with the longest linker peptides, BU-G10. More importantly, $\Delta\Delta G_{unfold}$ increases as the linker length is shortened. This trend in $\Delta\Delta G_{unfold}$ is qualitatively consistent with that obtained from GdnHCl denaturation experiments involving Trp fluorescence (8), despite the much greater concentration of proteins in our simulations relative to the experiments (mM versus μ M) and the use of thermal instead of chemical denaturation ($\Delta\Delta G_{unfold}$ values were not obtainable from thermal denaturation experiments).

We also directly calculated the free energy of coupling between the domains, which represents an energetic penalty imposed on the folding of one domain by the native structure of the other ([Fig. S4](#)). These coupling free energies, which are reminiscent of those between pairs of residues (34), become increasingly unfavorable as the linkers are shortened. This result is qualitatively consistent with both GdnHCl and thermal denaturation experiments, where Co^{2+} binding to an engineered site in the ubiquitin domain progressively lowers the midpoint of transition for the barnase domain as the linker length is shortened (8). However, in the absence of denaturant, circular dichroism (CD) spectra show that both domains remain folded in all of the BU variants, even in the presence of Co^{2+} . Thus, despite the presence of conformational strain, none of the variants appears to behave as a perfect molecular switch. We note that the CD spectra were collected at 10°C to minimize aggregation. Although we primarily focus on our simulation results at the T_m value of barnase, we did perform simulations of all of the BU variants at 10°C; consistent with experiment, these simulations show that both domains are folded ([Fig. S3](#)).

Finally, we quantified the propensity for dimerization (via domain swapping) for each variant by computing free energies of dimerization ($\Delta G_{dimerize}$) from our simulations (see [Methods](#)). At the millimolar concentrations of proteins used in our simulations, there is a massive increase in dimerization for all BU variants, including BU-G10, relative to free barnase ([Fig. 3 C](#)). Although the inserted ubiquitin

domain can be viewed as a covalent linker between the two interrupted halves of the barnase chain, it is not merely a tether, which has been shown to stabilize the formation of dimeric proteins (35); instead, the ubiquitin domain adopts a folded structure that imposes a large distance between the two halves of barnase, decreasing their intramolecular concentration such that the intermolecular folding of barnase (dimerization) more effectively competes with its intramolecular folding.

Interestingly, the formation of fully domain-swapped dimers in our simulations is rare (<10%), with the majority of dimers resulting from partial domain swapping, meaning the formation of one folded barnase domain (as opposed to two); a representative simulation of BU-G0 which illustrates the formation of a fully domain swapped dimer is provided as [Movie S1](#) in the [Supporting Material](#). Another key result of our simulations is that BU-G0 and BU-G1 are significantly more likely to dimerize than the other BU variants, flagging these variants as having potential aggregation issues. Of the remaining variants, BU-G2 is the most conformationally strained ([Fig. 3 B](#)). This result is qualitatively consistent with the propensity revealed by analytical ultracentrifugation experiments in which BU-G0 and BU-G1 form dimers (or even higher oligomers), and BU-G2 and BU-G10 are monomers (8). Given the qualitative agreement with experiment, our simulations provide support for domain swapping as a potential mechanism for the formation of dimers/higher oligomers.

Kinetics of mutually exclusive folding

Although the thermodynamic aspects of mutually exclusive folding have been studied in detail (7,8), the kinetics of the process are relatively unexplored. The only kinetics study of mutually exclusive folding in artificial domain insertion proteins was conducted recently using stopped-flow techniques (36). In this study, the folding and unfolding kinetics of the host protein, GB1-L5, were monitored. Then, the kinetics of the guest protein, a mutant domain of titin, were inferred assuming correspondence of host folding and unfolding rates to guest unfolding and folding rates, respectively. This correspondence of unfolding and folding rates, however, is not required for mutually exclusive folding, which simply involves the folding of one protein triggering the complete unfolding of the other protein. Nonetheless, we tested for this correspondence in our simulations of BU-G2 involving refolding of the ubiquitin domain. The folding rate constant for the ubiquitin domain is $129 \pm 9 \mu\text{s}^{-1}$ while the unfolding rate constant for the barnase domain is $76 \pm 12 \mu\text{s}^{-1}$ (see [Fig. S5](#); *error bars* represent 1 mean \pm SE). Although the unfolding rate of the barnase domain does not correspond to the folding rate of the ubiquitin domain, the unfolding of the barnase domain is clearly triggered by the folding of the ubiquitin domain ([Fig. 2 A](#)).

We also investigated the effect that domain insertion has on the folding rates of the host and guest proteins. Domain insertion clearly destabilizes the host protein in our simulations, leading to a more favorable free energy of unfolding. To determine whether this destabilization is due to a faster rate of unfolding and/or slower rate of folding, we computed the rate constants for unfolding and folding of the barnase domain in each BU variant at the T_m value of free barnase (51.5°C), where we observe a large number of unfolding and folding events. Table 1 summarizes our computed rate constants. Relative to free barnase, the barnase domains of all the BU variants unfold more quickly (by at least 2×), increasing as the linker peptides are shortened; in contrast, their folding rates are dramatically slower (by at least 7×), decreasing as the linker peptides are shortened.

The fact that the increase in the barrier to folding is much more than the reduction in the barrier to unfolding suggests that unfolding of the barnase domain is driven more by stabilization of the unfolded state than destabilization of the folded state, although both effects are observed. The trend in folding rates with linker length is the opposite of that found in experiments involving loop insertions of linker peptides in various proteins (37–39); in these experiments, as the linker length is shortened, the folding rates increase while the unfolding rates decrease, reflecting a reduction in the entropic cost of loop closure. In the BU fusion proteins, which are designed to be conformationally strained, this entropic cost is apparently more than counterbalanced by the increasing degree of strain that results as the linker length is shortened.

Effects of barnase-barstar binding

One key characteristic of molecular switches is that they can be toggled by the binding of certain ligands. A critical test of our molecular simulations, therefore, is whether or not the simulations can reproduce the effects of ligand binding on the tug-of-war between the domains in a two-domain

TABLE 1 Effect of linker length on the unfolding and folding rate constants of the barnase domain

	k_{fold} (μs^{-1})	k_{unfold} (μs^{-1})	N
Free barnase	26.7 ± 1.7	24.9 ± 0.9	1270
BU-G10	3.8 ± 0.5	54 ± 4	319
BU-G6	3.8 ± 0.5	54 ± 3	323
BU-G3	2.8 ± 0.5	55 ± 5	217
BU-G2	2.7 ± 0.3	61 ± 4	358
BU-G1	1.6 ± 0.3	70 ± 8	126
BU-G0	1.4 ± 0.3	68 ± 8	109

Unfolding and folding rate constants were estimated by taking the inverse of the average unfolding and folding times, τ_{unfold} and τ_{fold} , respectively, from 10 independent 10- μs simulations. Definitions of the unfolded and folded states ($Q \leq 0.35$ and $Q \geq 0.61$, respectively) were taken from a free energy profile based on simulations of barnase (Fig. S1). Error bars represent mean \pm 1 SE calculated from the combined number of unfolding and folding events (N).

protein switch. We examined the effect of binding to the barnase domain of BU-G2 by its natural ligand, barstar (40,41), on the stability of the ubiquitin domain by performing 10 independent simulations starting from the fully folded conformation of BU-G2 with the barstar ligand bound to the barnase domain. In the absence of barstar, BU-G2 is primarily in the state with unfolded barnase and folded ubiquitin while all other possible states are populated to a minor extent.

Once the barstar ligand is introduced, the barnase domain is dramatically stabilized such that it is always folded. Although the state with both domains folded is the most populated, there is also a significant increase in the state with folded barnase and unfolded ubiquitin (Fig. 4 A). This result is consistent with CD spectroscopy data involving a similar BU variant (with Gly-Thr and Gly-Gly-Ser linker peptides added to the N- and C-termini of ubiquitin, respectively, instead of Gly-Gly peptides), which reveals unfolding of the ubiquitin domain upon barstar binding of the barnase domain (6). To provide a more detailed view of the resulting conformational changes in BU-G2, we determined the average fraction of native contacts formed by each residue (Fig. 4 B). Consistent with NMR spectroscopy and hydrogen-deuterium exchange data, the most dramatic conformational changes that occur in the barnase domain upon barstar binding occur in the binding site region (42); the unfolding of the ubiquitin domain is most dramatic in its β -sheet region.

In the context of the mutually exclusive folding design, one might expect that the host domain (barnase) unfolds the guest domain (ubiquitin) by compressing the termini of the guest domain; this mechanism is in contrast to that by which the guest domain might unfold the host domain: pulling apart the host domain. To further explore the effect of compressing the N- to C-terminal distance of ubiquitin on its stability, we gradually compressed the distance from 38 to 11 Å by performing simulations of free ubiquitin with distance restraints; we also performed simulations of free barnase with the distance between the ends of the insertion loop pulled apart from 11 to 38 Å (see Fig. 1 A).

Interestingly, ubiquitin is essentially unfazed by the compression until its N- to C-terminal distance is $< \sim 26$ Å, at which point the protein begins to unfold (Fig. 4 C). Expansion of the barnase protein necessarily allows for complete unfolding, as both fragments of the protein are eventually pulled apart beyond 30 Å (Fig. 4 D). These results suggest that distance-restrained simulations might be useful for identifying the distances required for compressing or pulling apart a protein, aiding the selection of optimal protein components for the design of mutually exclusive folding switches.

Distance-restrained simulations involving free ubiquitin suggest that the ubiquitin domain of BU-G2 may begin to unfold when the distance between its termini becomes < 26 Å due to compression by the folded barnase

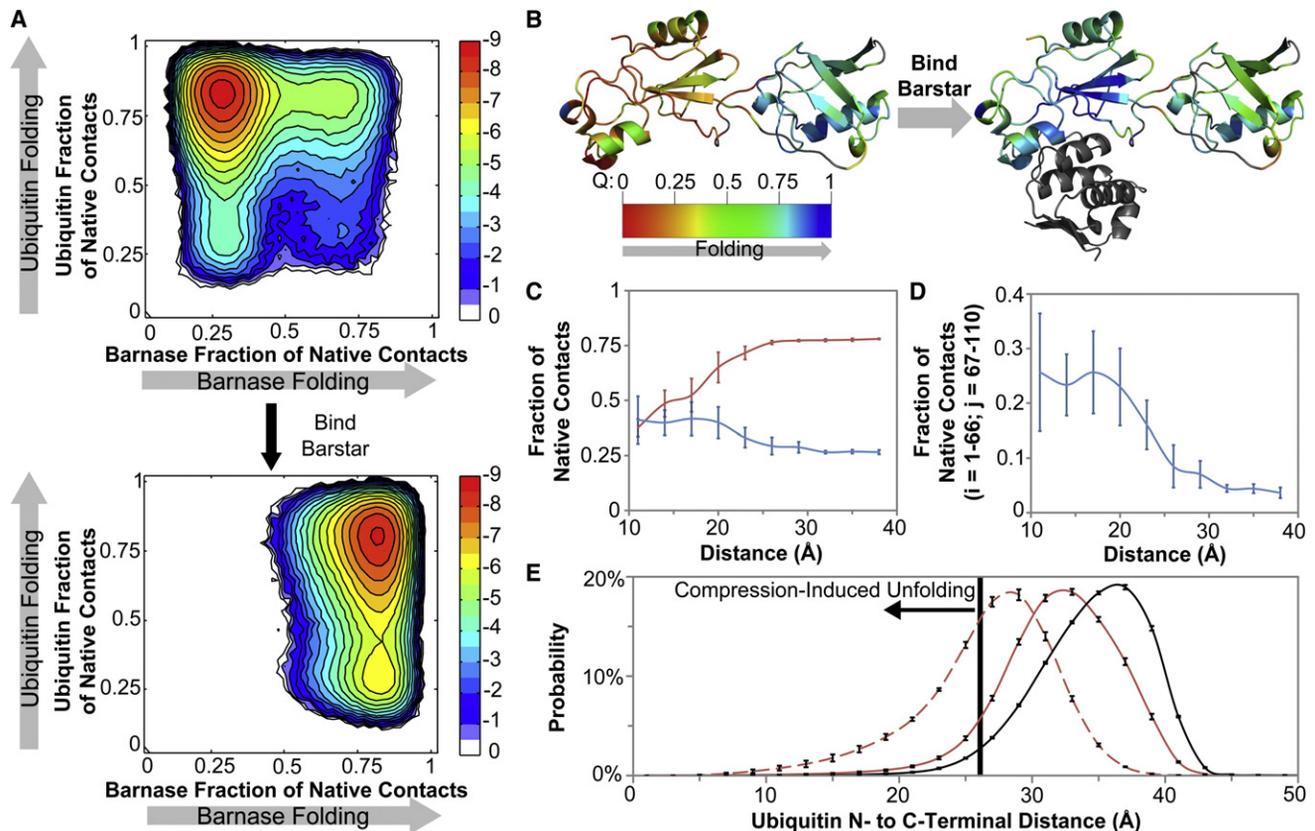


FIGURE 4 Effects of barstar binding on the stability of the barnase and ubiquitin domains in the BU-G2 fusion protein. (A) Potential of mean force surfaces for BU-G2 and the BU-G2/barstar complex as a function of the fraction of native contacts in the barnase and ubiquitin domains. Data shown for each system is based on 10 independent 10- μ s simulations. Contours are drawn at intervals of the available thermal energy, $0.5RT$. (B) Average fraction of native contacts formed by each residue in BU-G2 and the BU-G2/barstar complex. The barstar ligand (gray). (C) Average fraction of native contacts as a function of distance between 1), the termini of free ubiquitin and 2), the ends of the insertion loop of free barnase (see Fig. 1 A). Ten 1- μ s simulations were performed for each distance using restraints. (D) Average fraction of native barnase contacts between residues on opposite sides of the insertion site as a function of distance between the ends of the insertion loop in the simulations of free barnase described in panel C. (E) Average distribution of N- to C-terminal distances in free ubiquitin (black) and the ubiquitin domain of BU-G2 in the absence and presence of the barstar ligand (red, solid and red, dashed, respectively) from the 10 independent simulations described in panel A. The solid black line at ~ 26 Å indicates the N- to C-terminal distance at which ubiquitin begins to unfold due to compression as shown in panel C. All error bars represent 1 SD ($N = 10$).

domain. Indeed, the distribution of N- to C-terminal distances in the ubiquitin domain shifts toward shorter distances upon barstar binding to the barnase domain, resulting in a greater percentage of BU-G2 conformations with distances < 26 Å, increasing from 8 to 38% (Fig. 4 E). Thus, although the partial unfolding of the ubiquitin domain upon barstar binding of BU-G2 is apparently due to compression of its termini, the degree of compression is insufficient for complete unfolding. Localized compression of proteins, such as compression of its termini, has been an integral part of facilitating unfolding in engineered protein systems (43,44). We provide what we believe is the first molecular view of the extent of compression at the ends of a protein that is sufficient for unfolding.

CONCLUSIONS

We have performed what we believe to be the first direct simulations of mutually exclusive folding for a two-domain

protein switch, providing molecular views of each domain that are difficult to obtain using laboratory experiments in the context of the switch. In addition, we have demonstrated that our molecular simulation approach can reproduce the qualitative effects of linker length, including propensities for dimerization, and ligand binding as observed in experiments. Finally, our simulations provide what appear to be novel insights about the kinetics of mutually exclusive folding on the single-molecule level and the ease of unfolding a protein with localized compression. Although simulations and experiments both show that the BU variants are not perfect molecular switches, the fact that these variants can still be filtered by our simulations based on their extent of partial switch activity underscores their usefulness as sensitive virtual assays of switch activity.

It should be noted that BU-G2, despite being the best of the fusion constructs, lacks the degree of interdomain coupling that is required for the mutually exclusive folding design (8). For example, although barstar binding of its

barnase domain causes partial unfolding of the ubiquitin domain in our simulations, the most populated state consists of both domains in their folded states. For an optimal bifunctional switch, only one of the domains should be folded. Optimizing the degree of mutually exclusive folding is likely to involve more than just the degree of coupling between the domains; for example, the structural plasticity of the individual domains may also play a role.

Finally, given the likelihood of unfolding-induced oligomerization, the exploration of alternate design strategies that require less net unfolding may yield greater switch activity; one successful strategy fuses two proteins together end-to-end with an overlapping sequence (10–14).

SUPPORTING MATERIAL

Full details about the approximations made in our simulation model, along with Figs. S1–S5, Table S1, and Movie S1, are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(10\)05254-9](http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)05254-9).

We thank Stewart N. Loh (State University of New York Medical School) and Adrian H. Elcock (University of Iowa) for valuable discussions. We are also grateful to Adrian H. Elcock for making his Brownian dynamics simulation software available.

This work was supported in part by the National Science Foundation's CAREER award (No. MCB-0845216 to L.T.C.) and the University of Pittsburgh's Arts & Sciences Fellowship (to B.M.M.).

REFERENCES

- Bourne, H. R., D. A. Sanders, and F. McCormick. 1990. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature*. 348:125–132.
- Huse, M., and J. Kuriyan. 2002. The conformational plasticity of protein kinases. *Cell*. 109:275–282.
- Koide, S. 2009. Generation of new protein functions by nonhomologous combinations and rearrangements of domains and modules. *Curr. Opin. Biotechnol.* 20:398–404.
- Ostermeier, M. 2009. Designing switchable enzymes. *Curr. Opin. Struct. Biol.* 19:442–448.
- Ambroggio, X. I., and B. Kuhlman. 2006. Design of protein conformational switches. *Curr. Opin. Struct. Biol.* 16:525–530.
- Radley, T. L., A. I. Markowska, ..., S. N. Loh. 2003. Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol.* 332:529–536.
- Cutler, T. A., and S. N. Loh. 2007. Thermodynamic analysis of an antagonistic folding-unfolding equilibrium between two protein domains. *J. Mol. Biol.* 371:308–316.
- Cutler, T. A., B. M. Mills, ..., S. N. Loh. 2009. Effect of interdomain linker length on an antagonistic folding-unfolding equilibrium between two protein domains. *J. Mol. Biol.* 386:854–868.
- Ha, J. H., J. S. Butler, ..., S. N. Loh. 2006. Modular enzyme design: regulation by mutually exclusive protein folding. *J. Mol. Biol.* 357:1058–1062.
- Sallee, N. A., B. J. Yeh, and W. A. Lim. 2007. Engineering modular protein interaction switches by sequence overlap. *J. Am. Chem. Soc.* 129:4606–4611.
- Stratton, M. M., D. M. Mitrea, and S. N. Loh. 2008. A Ca²⁺-sensing molecular switch based on alternate frame protein folding. *ACS Chem. Biol.* 3:723–732.
- Strickland, D., K. Moffat, and T. R. Sosnick. 2008. Light-activated DNA binding in a designed allosteric protein. *Proc. Natl. Acad. Sci. USA*. 105:10709–10714.
- Mitrea, D. M., L. S. Parsons, and S. N. Loh. 2010. Engineering an artificial zymogen by alternate frame protein folding. *Proc. Natl. Acad. Sci. USA*. 107:2824–2829.
- Stratton, M. M., and S. N. Loh. 2010. On the mechanism of protein fold-switching by a molecular sensor. *Proteins*. 78:3260–3269.
- Gō, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12:183–210.
- Takada, S. 1999. Gō-ing for the prediction of protein folding mechanisms. *Proc. Natl. Acad. Sci. USA*. 96:11698–11700.
- Clementi, C. 2008. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* 18:10–15.
- Mauguen, Y., R. W. Hartley, ..., A. Jack. 1982. Molecular structure of a new family of ribonucleases. *Nature*. 297:162–164.
- Vijay-Kumar, S., C. E. Bugg, and W. J. Cook. 1987. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* 194:531–544.
- Buckle, A. M., G. Schreiber, and A. R. Fersht. 1994. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. *Biochemistry*. 33:8878–8889.
- Jha, A. K., A. Colubri, ..., T. R. Sosnick. 2005. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA*. 102:13099–13104.
- Elcock, A. H. 2006. Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLOS Comput. Biol.* 2:e98.
- Knott, M., H. Kaya, and H. S. Chan. 2004. Energetics of protein thermodynamic cooperativity: contributions of local and nonlocal interactions. *Polymer (Guildf.)*. 45:623–632.
- Ermak, D. L., and J. A. McCammon. 1978. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* 69:1352–1360.
- Frembgen-Kesner, T., and A. H. Elcock. 2009. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory Comput.* 5:242–256.
- Hess, B., H. Bekker, ..., J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
- 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.
- Koga, N., and S. Takada. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.* 313:171–180.
- Yang, S., S. S. Cho, ..., J. N. Onuchic. 2004. Domain swapping is a consequence of minimal frustration. *Proc. Natl. Acad. Sci. USA*. 101:13786–13791.
- Ding, F., K. C. Prutzman, ..., N. V. Dokholyan. 2006. Topological determinants of protein domain swapping. *Structure*. 14:5–14.
- Ding, F., N. V. Dokholyan, ..., E. I. Shakhnovich. 2002. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *J. Mol. Biol.* 324:851–857.
- Lazar, G. A., J. R. Desjarlais, and T. M. Handel. 1997. De novo design of the hydrophobic core of ubiquitin. *Protein Sci.* 6:1167–1178.
- Daggett, V., and A. R. Fersht. 2003. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* 4:497–502.
- Serrano, L., A. Horovitz, ..., A. R. Fersht. 1990. Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry*. 29:9343–9352.
- Zhou, H.-X. 2001. Single-chain versus dimeric protein folding: thermodynamic and kinetic consequences of covalent linkage. *J. Am. Chem. Soc.* 123:6730–6731.

36. Peng, Q., and H. Li. 2009. Direct observation of tug-of-war during the folding of a mutually exclusive protein. *J. Am. Chem. Soc.* 131:13347–13354.
37. Ladurner, A. G., and A. R. Fersht. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* 273:330–337.
38. Nagi, A. D., K. S. Anderson, and L. Regan. 1999. Using loop length variants to dissect the folding pathway of a four-helix-bundle protein. *J. Mol. Biol.* 286:257–265.
39. Viguera, A.-R., and L. Serrano. 1997. Loop length, intramolecular diffusion and protein folding. *Nat. Struct. Biol.* 4:939–946.
40. Hartley, R. W. 1993. Directed mutagenesis and barnase-barstar recognition. *Biochemistry.* 32:5978–5984.
41. Schreiber, G., and A. R. Fersht. 1993. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry.* 32:5145–5150.
42. Jones, D. N. M., M. Bycroft, ..., A. R. Fersht. 1993. Identification of the barstar binding site of barnase by NMR spectroscopy and hydrogen-deuterium exchange. *FEBS Lett.* 331:165–172.
43. Butler, J. S., D. M. Mitrea, ..., S. N. Loh. 2009. Structural and thermodynamic analysis of a conformationally strained circular permutant of barnase. *Biochemistry.* 48:3497–3507.
44. Blois, T. M., H. Hong, ..., J. U. Bowie. 2009. Protein unfolding with a steric trap. *J. Am. Chem. Soc.* 131:13914–13915.
45. Gront, D., S. Kmiecik, and A. Kolinski. 2007. Backbone building from quadrilaterals: a fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. *J. Comput. Chem.* 28:1593–1597.
46. DeLano, W. L. 2002. The PyMOL Molecular Graphics System, Ver. 1.2. DeLano Scientific, Palo Alto, CA.