PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Folding of multidomain proteins: Biophysical consequences of tethering even in apparently independent folding

Oshrit Arviv and Yaakov Levy*

Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

ABSTRACT

Most eukaryotic and a substantial fraction of prokaryotic proteins are composed of more than one domain. The tethering of these evolutionary, structural, and functional units raises, among others, questions regarding the folding process of conjugated domains. Studying the folding of multidomain proteins *in silico* enables one to identify and isolate the tethering-induced biophysical determinants that govern crosstalks generated between neighboring domains. For this purpose, we carried out coarse-grained and atomistic molecular dynamics simulations of two two-domain constructs from the immunoglobulin-like *β*-sandwich fold. Each of these was experimentally shown to behave as the "sum of its parts," that is, the thermodynamic and kinetic folding behavior of the constituent domains of these constructs seems to occur independently, with the folding of each domain uncoupled from the folding of its partner in the two-domain construct. We show that the properties of the individual domains can be significantly affected by conjugation to another domain. The tethering may be accompanied by stabilizing as well as destabilizing factors whose magnitude depends on the size of the interface, the length, and the flexibility of the linker, and the relative stability of the domains. Accordingly, the folding of a multidomain protein should not be viewed as the sum of the folding patterns of each of its parts, but rather, it involves abrogating several effects that lead to this outcome. An imbalance between these effects may result in either stabilization or destabilization owing to the tethering.

Proteins 2012; 80:2780–2798.

Key words: multidomain protein; conjugated protein; energy landscape; coarse-grained simulation; flexible linker.

INTRODUCTION

The modularity of protein structure plays a significant role in the divergence and adaptation of all life forms on earth.[1,2] In fact, analyses of sequenced genomes have shown that approximately 40–65% of prokaryotic proteins and more than 70% of eukaryotic proteins consist of more than one domain.[3–5] Around 95% of multidomain proteins contain two to five domains, and only a few comprise many more. Although a unique, consistent definition of a protein domain remains elusive,[6] it is frequently referred to as "a structural, functional, and evolutionary component of proteins, which can often be expressed as a single unit."[7,8] Therefore, domains sharing functional and structural features suggesting a common evolutionary origin can be grouped into superfamilies.[7] Protein domains of this limited repertoire are able to combine and therefore can facilitate the emergence of novel and complex functions.[1,9–11] The resulting domain combinations, which represent <0.5% of all possible superfamily combinations[1] (with most superfamilies having neighboring tethered domains from just one or two other superfamilies and almost always in the same N- to C-sequential order[12]), were clearly selected under strong evolutionary pressure. The existing domain combinations have evolved to sustain a variety of biological functions; yet, other types of selection constraints may have been involved as well. Indeed, the folding characteristics (e.g., structural stability and folding kinetics) of multidomain proteins stand out as an important constraint and suggest the existence of a funneled energy landscape in which certain outcomes are associated with reduced frustration.[13,14]

Nevertheless, the folding of tethered domains might be different from that of the isolated constituent

Grant sponsor: The Kimmelman Center for Macromolecular Assemblies; Grant sponsor: The Israeli Science Foundation
*Correspondence to: Yaakov Levy, Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: koby.levy@weizmann.ac.il
Received 8 April 2012; Revised 11 July 2012; Accepted 16 July 2012
Published online 14 August 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24161

domains.[15,16] The domains in a multidomain architecture may share significant interfaces and be attached to each other by either flexible or structured linkers of various lengths.[17] Therefore, the tethering of domains may affect the energy landscape of each of the component's domains (compared to its characteristics in isolation); hence, this may induce a variety of effects on the protein's thermodynamics and kinetics. The tethering effects may be minimized when the domains fold cotranslationally, one domain at a time. However, in the operative context, domains refold, after undergoing spontaneous (thermal) unfolding, several times during the lifetime of a multidomain protein. This suggests that energetically favorable folding pathways are also conserved from the perspective of these multidomain constructs. To investigate these effects, one has to study domains in isolation as well as in their multidomain architecture. However, in recent years most efforts in folding research have focused almost exclusively on studying single-domain proteins or small proteins in isolation, whereas the possible effects of interactions between domains on folding remain an understudied area of protein folding research.

Experimentally, the thermodynamics and kinetics of both isolated domains and conjugated constructs have been studied for only a few multidomain proteins (most of which were surveyed in comprehensive reviews[8,18]). The studied systems usually show that the folding characteristics of the individual domains depend on whether they fold as part of a multidomain complex or in isolation. These variations can potentially be traced back to differences in the crosstalks between tethered domains, which may also be reflected in their multidomain topologies. For example, it was suggested that an extensive helical linker domain correlates with measurable cooperative folding and large densely packed interfaces were shown to correlate with the stabilization of the tethered domains.[18] The observed enhanced thermostability, imposed by interfacial interactions that stabilize the native state, was also associated with slower unfolding rates. Interactions in the nonfully folded protein can be exploited to catalyze the folding of an adjacent domain, resulting in more efficient folding, or on the contrary, may significantly impede folding. A domain may also be affected by the unfolded neighboring domain (for more details, see Ref. [18]). However, obtaining a detailed account of possible different crosstalks is strongly limited by the scarcity of experimental data. Very few studies have been conducted with the aim of exploring differences in the behavior of tethered versus isolated domains; instead, they were motivated by a desire to study the features of specific proteins of interest. Reviews published in the field of multidomain proteins admirably aimed to generalize from the accumulated data, but they themselves emphasized the difficulty in deducing rules from the limited data set.[8,18]

The intuitive claim that tethering probably modifies the energy landscape of the protein domain is consistent with most cases studied,[8] which have shown folding behavior to be dependent on and/or affected by the neighboring domains. However, a surprising special scenario stands out, which will also be the focus of our investigation, in which a seemingly independent folding is observed. In this "sum of its parts" scenario,[19] the stability of each of the domains and the folding and unfolding rate constants were unaffected by the neighboring domain, and therefore were the same as those of each component of the multidomain protein in isolation. A multidomain protein that behaves as the "sum of its parts" lacks any crosstalk between constituent domains, that is, the interaction energy of a domain in a "sum of its parts" architecture with its native (folded) or unfolded neighbor was defined as zero.[18] As this is counterintuitive to the line of thought described earlier, it led us to wonder: Can tethering one domain to another produce a multidomain protein in which there is no communication between adjacent domains and no biophysical consequences? Moreover, from the perspective of seeking a more general understanding of the principles governing folding in a multidomain architecture, an intriguing question is: what differentiates the coupled domains folding scenario from this seemingly independent folding?

In pursuit of a better understanding, one can examine the nature of the interface and linker in these multidomain architectures, as there is ample evidence, suggesting that they are key features that may govern the degree of coupling in the folding of the tethered domains. The linkers, which interconnect the constituent domains, must be thought of as more than simple covalent connectors. The role of linkers in establishing the structural and functional[20,21] assembly of multidomain proteins was found to be of major importance.[22] Resolution of static 3D structures leaves unaccounted the dynamics of conformational changes, which are highly important if one wishes to better understand biological activity. The linkers perform the important task of establishing communication by directing the correlated movements of various domains. Indeed, a structured stiff linker can serve to orient domains in space and limit interactions to other domains that are not adjacent, whereas more flexible linkers can also enable different degrees of coupling between domains, as the linkers control the conformational changes required. Experimentally, extracting the ensemble of rapidly converting conformers is far from trivial, yet substantial efforts are being made.[23] Computational methods can aid in expanding the accessibility to linker dynamics.[24] Of course, the length and flexibility of linkers must also be correlated with the packing density across the interface and the extent of the sampling of interface topologies.[17] Intuitively, a large complementary interface will not only allow domains to adopt specific conformations relative to each other—it will also func-

tion to stabilize a protein. Specifically and in light of evolutionary pressure, tethered domains that have an extensive, densely packed interface will confer more stability than when the interface is small. However, the interactions' energy will depend on the interface topology and on the specificity of the interactions between the adjacent domains. Several multidomain systems surveyed demonstrated a tendency for the interface between domains to be largely hydrophobic,[8,25] which should influence the compatibility and strength of the interactions. Interestingly, in known cases where domains fold independently, the linkers between the domains are apparently short and flexible and the interfaces are small. Therefore, studying the "sum of its parts" scenario is consistent with reducing the complexity, which entails investigating the simplest multidomain architectures. In this article, we will concentrate on the "sum of its parts" scenario and attempt to provide a fully detailed microscopic explanation of its physical and chemical origin.

METHODS

Coarse-grained simulations

We applied a reduced model based on native topology to study the folding of multidomain proteins. We represented each amino acid (a.a.) by a single bead centered at the Cα. The molecular dynamic simulations were dictated by the set of forces exerted on the beads, based on the chosen force field—a native topology-based model (a Go-type potential energy function that portrays the perfect funneled energy landscape[26–30]). This native structure-based model provides a reduced description of the folding process of domains in isolation as well as in the presence of their tethered neighbor. We assumed that the energy landscapes of both single and multidomain proteins are funnel-like, with reduced frustration and that within the scope of our model, each of the domains maintains its overall topology, with its intracontacts still representing its structure in isolation. In the previous studies, it has been shown that the folding of multidomain or conjugated proteins is determined by topology, as a structure-based model was successful in capturing the essential features of their folding.[31–34]

The dynamics of the system was simulated by the Langevin equation, with a friction constant $\gamma = 0.01$. The Langevin thermostat for a system with fast and slow degrees of freedom (associated with the flexible linker and rigid structural domain, respectively) is important to avoid inhomogeneous distribution of the thermal energy.[35] For each system, multiple trajectories at various temperatures were simulated and numerous unfolding/folding events were obtained. The trajectories were analyzed by using the Weighted Histogram Analysis Method.[36] A consistency check was carried out by analyzing subgroups of trajectories and their differences are used to estimate the errors in the thermodynamic analyses. Free energy profiles as a function of the number of native contacts Q are generated.[37] A conformation was acknowledged as folded if its number of native contacts was higher than the number of contacts at the highest point of the transition barrier, and as unfolded if displayed otherwise. The same threshold was used to calculate free energy and enthalpy values for the folded or unfolded ensembles by summing all of the free energy values below or above the threshold separately. Entropic energy values were calculated by subtracting the enthalpy value from the free energy values. The values ΔH and $T\Delta S$ were calculated by subtracting the folded state enthalpy and entropy energy values from the unfolded state enthalpy and the entropy energy values from the corresponding values of the unfolded state. Additionally, as an indicator of the relative stability, the folding temperature, which is defined as the simulation temperature at which the ratio of folded to unfolded conformations is the closest to unity, was estimated. It is also correlated with the temperature at which the difference in free energy values between the folded and the unfolded states is zero. Rough estimations of relative folding and unfolding rates were obtained from the height of the respective transition barrier.

Attaining thermodynamic values for various systems (e.g., isolated vs. tethered variants) allows a detailed comparison of their folding characteristics. The validity of the comparison between simulations of modified systems is based on utilizing each domain's energy term but only for native contacts which are mutual to the compared systems, that is, one should take special care to ensure that identical domains in different constructs (e.g., as isolated vs. tethered or N- vs. C-domain in a homo multidomain protein) will comprise the same list of contacts. To dissect the contributions of various potential determinants of folding, in some of the modeled systems, we introduced fewer terms of the force field or simply altered the terms' coefficients. Modeling systems, which do not include interdomain interactions (namely, the interface), were obtained by removing the Lennard–Jones interactions between all i,j pairs of different domains. Studying the linker flexibility, in this context, was modeled by gradually decreasing the dihedral coefficient associated with the linker's beads. The crowding by adjacent domains was also studied by removing the repulsions between all i,j pairs of different domains. In this case, only the chain connectivity via bonds and angles represented the tethering of the domains, neglecting all other possible interactions between domains. Another change made was in the context of imitating the relatively high stability of one domain versus its tethered counterpart. As each of the tethered domains is expressed solely by its own topology (as shown by its coarse-grained model), the difference in T_f associated with each domain does not reflect their relative thermostability. However, the

high difference in thermostability in the case of fibronectin can be utilized to prevent the significantly more stable domain from undergoing unfolding. More specifically, it was modeled as a rigid body (petrified in the folded state) in which all its native contacts were assigned with a bond-like harmonic potential instead of Lennard–Jones interactions. Additionally, one should know that in the framework of this model, the native topology of the interface is translated to all attractive contributions as contacts (neglecting non-native interactions). Therefore, it is not surprising that an interface will influence the stabilization of the domains in their multidomain architecture. Moreover, the relative strength of the native attractive interaction is limited to one interaction (contact) per residue, ascribed to the coordinates of the Cα atom.

In this study, some of the simulations were supplemented by two other possible factors:

a. A virtual bond between each domain center of mass was added[27,38] to constrain two unlinked domains to a predetermined distance.
   $U_{\mathrm{CM\ constraint}} = K_{\mathrm{CM}}\,(R - R_0)^2$, where R is the center of mass distance of two domains; R_0 is the target distance; $K_{\mathrm{CM}} = 1$.
b. A chaperon-like cage[39,40] was included to impose geometrical constraints on the protein domains by confining their conformational space.

$$U_{\mathrm{cage}} = \sum_i K_{\mathrm{cage}} \left[\left(\frac{C}{2d_i}\right)^4 - 2\left(\frac{C}{2d_i}\right)^2 + 1 \right] H\left(\frac{C}{2} - d_i\right)$$

where d_i is the distance between the cage wall and the i bead; $H(x) = 1$ if $x > 0$ and $H(x) = 0$ if $x < 0$; $K_{\mathrm{cage}} = 100$. This is a purely repulsive interaction.

Atomistic simulations

Simulations at atomistic resolution have been carried out using the GROMACS package.[41] Briefly, the studied two-domain protein systems were solvated in a water box under periodic boundary conditions. Water layers between protein surfaces and walls prevented interactions between protein and its periodic image. Twenty-eight Na^+ counterions were added to make the systems electrically neutral. The solvated protein systems have up to 48,000 atoms. A standard equilibration procedure, which includes conjugate gradient minimization and a 200 ps simulation with position restraints, is followed to equilibrate each solvated system. The final configurations of equilibration are subsequently used as starting conformations for NPT simulations (1 bar and 300 K). For each studied system, we ran three trajectories, starting at different equilibrated configurations, and which were associated with different random seeds, of 500 nsec. Langevin dynamics ensured stochastic temperature coupling, and the pressure was coupled to a Berendsen barostat. The AMBER99 force field was used[42] with TIP3P water. A switching distance of 0.8–0.9 nm was set for the van der Waals interactions. The particle-mesh Ewald method was applied for electrostatic interactions. The time step was set to 2.0 fsec with bond lengths constrained using the linear constraint solver algorithm.

Studied systems

In this study, we focused on the following systems:

1. FNfn9–FNfn10—This is a natural two-fibronectin type III (*fnIII*) domain system from human fibronectin. The PDB 1fnf entry encompasses the 7th to 10th domains, from which we extract the FNfn9–FNfn10 structure. These domains are all-β protein and of the *immunoglobulin-like* β*-sandwich* fold. The FNfn9 domain includes 92 a.a. and the FNfn10 domain includes 94 a.a. The two a.a in the interconnecting region belong to both domains[43] and constitute the linker. In *in vitro* folding experiments, the FNfn9 domain was at first shown to be significantly less stable alone ($\Delta G_{\mathrm{D-N}} \sim 1.0$ kcal mol^{-1}, where D and N are the denatured and native states, respectively) than when tethered to FNfn10 ($\Delta G_{\mathrm{D-N}} \sim 3.5$ kcal mol^{-1}).[44] This added stabilization of the FNfn9 domain was, therefore, attributed to its conjugation to the highly stable FNfn10 domain.[45] However, when FNfn9 was lengthened by two residues, its stability was found to be independent of the presence of FNfn10. Therefore, it was concluded that the two residues at the C-terminus of FNfn9 and the N-terminus of FNfn10 structurally belong to both domains.[43] This illustrates the importance of correctly selecting boundaries. In the following report, we have accepted the above-clarified domain boundaries, and have referred to the additional two residues as part of the isolated FNfn9 definition. Our simulations, in agreement with the experimental data, showed that the FNfn10 domain has thermostability in isolation similar to when it is tethered to FNfn9.
2. I27–I27—This is an identically tethered two-immunoglobulin domain system. I27 is a domain from human titin, and its PDB entry *1tit* was used. I27 is an all-β protein of the *immunoglobulin-like beta-sandwich* fold; it includes 89 a.a. (Fig. 3). To *in silico* design, the homo conjugate construct, we used Pymol software. As in *in vitro* experiments, the linker between adjacent domains comprised two a.a.[43,46] In the coarse-grained model, we chose to compose the linker using two beads representing Ala. For studying linker flexibility, the structure of the linker was minimized in GROMACS by using the conjugate gradient algorithm. In all-atom simulations, we chose the most naive combination of Ala–Ser (among the less bulky and without the associated charge) of the actual combinations that were experimentally implemented.[43]

In the framework of the coarse-grained model, the tethered two-domain construct has been composed of two domains having the same native contact list as the isolated domain. No additional contacts were allowed between the linker and the domains, or for that matter, between the domains themselves, that is, all interfacial contacts were excluded. Therefore, this *homo* two-domain construct was conjugated solely by the linker's chain geometry. As the two tethered domains are identical, there is no need to worry about questions regarding the relative thermostability and any significant difference in the domains' calculated thermostability. With sufficient sampling, the break in the symmetry should be expressed in terms of links at different termini. In the absence of interfacial contacts, we controlled the linker's degree of flexibility by gradually changing the coefficient of the dihedral angle term of the force field ($\epsilon_{\text{dihedral}} = 0$, 0.25, 0.5, 0.75, and 1). In this case, the rigid variant corresponds to a fully extended linker (with tetrahedral angles and dihedral angles minimized to avoid spatial conflict). This described procedure enabled us to determine the extent of the influence of the linker's degree of flexibility.

3. TNfn3–TNfn3—This an identically tethered two-fibronectin type III (*fnIII*) domain system. TNfn3 is a domain from human tenascin. TNfn3 and FNfn10 have essentially the same structure (the backbone rmsd between pairs of structurally equivalent residues is 1.2 Å), but a low-sequence identity (24%).[47,48] We used the PDB entry *1ten,* which was extended by two residues at its C-terminus (Gly–Leu). Studies have shown that this boundary selection more accurately portrays the domain; as it includes two additional residues, it has a significant stabilizing effect.[49] A preceding minimization procedure was carried out first for this region. Overall, this domain comprises 92 residues. We have simulated this construct in an all-atom representation, to be used in comparison to the simulated I27–I27. The linker's composition was Ala–Ser, the same as described above.

RESULTS AND DISCUSSION

As the topologies of the various multidomain architectures confer strong constraints on potential crosstalks, they may play a determining role in the folding of the constituent domains. It was shown that the folding of multidomain proteins, similarly to single-domain proteins, is governed by their native topology and likewise follows the success of native topology-based models in capturing the essential features of their folding.[50–53] Additionally, it was shown that native topology-based models successfully revealed the effect of various conjugations on protein dynamics, resulting in branched proteins.[32] A similar model was also used to study the tethering of a protein at different locations and heights to a surface and its impact on folding.[54–56] Simulations of topology-based minimalist models not only enjoy the benefits of computational efficiency versus all-atom simulations—they can also be tailored to address specific questions because of their simplified representation. Indeed, in this study we implemented different minimalist representations tailored to the questions at hand to capture specific characteristics of multidomain folding, and we identified the biophysical determinants of the interplay between neighboring domains. Pinpointing the relevant factors and their degree of influence is an important initial step toward deciphering the manner in which the energy landscape of the domain in the modular architecture may be altered.

Here, we discuss the results obtained for two two-domain case study systems that experimentally have been shown to behave as the "sum of its parts": (1) FNfn9 and its *natural* neighbor FNfn10 (the 9th and 10th fnIII domains of fibronectin)[43]; (2) I27 (the titin 27th immunoglobulin domain) conjugated to another I27 domain.[46] Both these two-domain systems comprise domains of the same fold—the immunoglobulin-like β-sandwich. In attempting to determine the intermolecular forces involved in the interplay between the neighboring domains of these constructs, we examined the influence of several possible determinants. Each of the potential determinants was studied separately, which enabled us to quantify its effect in the specific architecture and also allowed us to infer its influence on governing the crosstalks between tethered domains.

We began our investigation with the FNfn9–FNfn10 system. Studying the folding dynamics of multidomain proteins using a native topology-based model requires a resolved 3D native structure. The structure of the FNfn9–FNfn10 construct is known (PDB entry: 1fnf—encompassing the 7th through 10th type III repeats), including the native details of the interdomain region, that is specific orientation and interactions (Fig. 1). The folding behavior of the FNfn9 domain in isolation and as part of the tethered two-domain construct was simulated and analyzed. Our model indicated that a tethered FNfn9 domain is thermodynamically destabilized when compared with its isolated variant (Fig. 2(A), blue line vs. gray circles). The lower stability of the tethered domain is manifested in its lower folding temperature compared with the isolated variant (the folding temperature, T_F, is the temperature at which $\Delta G_{U-F} \sim 0$, which is extracted from the peak of specific heat capacity vs. temperature). The statistically significant, yet quite surprising, destabilization occurred as a consequence of tethering FNfn9 to an FNfn10 domain, in contrast to the expected experimentally observed independence. This demonstrated that the straightforward representation of the simple, coarse-grained model incorrectly describes its folding behavior.

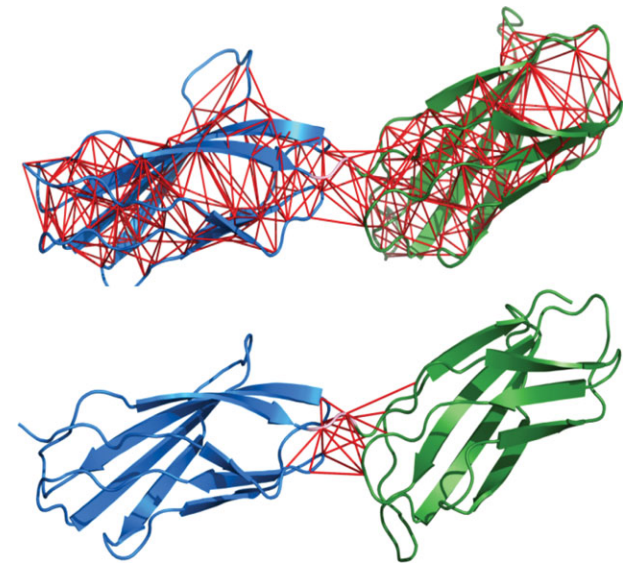

Figure 1

Structure of the ninth and tenth fibronectin type II domains in the FNfn9–FNfn10 two-domain protein (PDB ID code, 1fnf). The N- and C-domains are shown in blue and green, respectively, and the two-residue linker is in pink. (**A**) The pairwise native interactions are shown in red. (**B**) All interfacial pairwise interactions are shown in red.

We therefore concluded that there are specific factors influencing the folding of the two-domain construct, which were not taken into consideration. Thus, we sought to identify them, as those may aid us in deciphering the modifications needed in the native topology-based model to study multidomain proteins.

Deciphering folding determinants of a two-domain protein with "sum of its parts" thermodynamics

The effect of the interface between the constituent domains on the folding of multidomain proteins

If one wants to adopt a view of multidomain constructs as being simply large sole proteins as oppose to linked multiple protein units, then the relative density of interfacial contacts may serve to set the balance between these views, that is, the more extensive the interface will be, the more it will shift the dynamics and stability to be characteristic of the protein as a whole, and in divergence from a scenario of mild crosstalking protein domains.

In an attempt to illustrate the folding behavior of the simulated FNfn9–FNfn10 system, the dynamics of the interfacial contacts during folding and unfolding events was demonstrated in a 3D free energy surface projected on the degree of formation of interfacial contacts and intramolecular contacts of FNfn9 [Fig. 3(B)]. From this plot, it is clear that when the FNfn9 domain is folded, there is a substantial probability for forming interfacial contacts. In contrast, when the domain is unfolded, a substantial percentage of the interfacial contacts is not formed. Although folding of FNfn9 may involve the formation of interfacial contacts, it seems that its folding in this tethered construct is not coupled to the creation of interfacial contacts. Given that, in the overall structure of this two-domain protein, the domains are much more structured regions than the interface, apparently the constraints imposed by the intradomain contacts are stronger than those imposed by the interface.

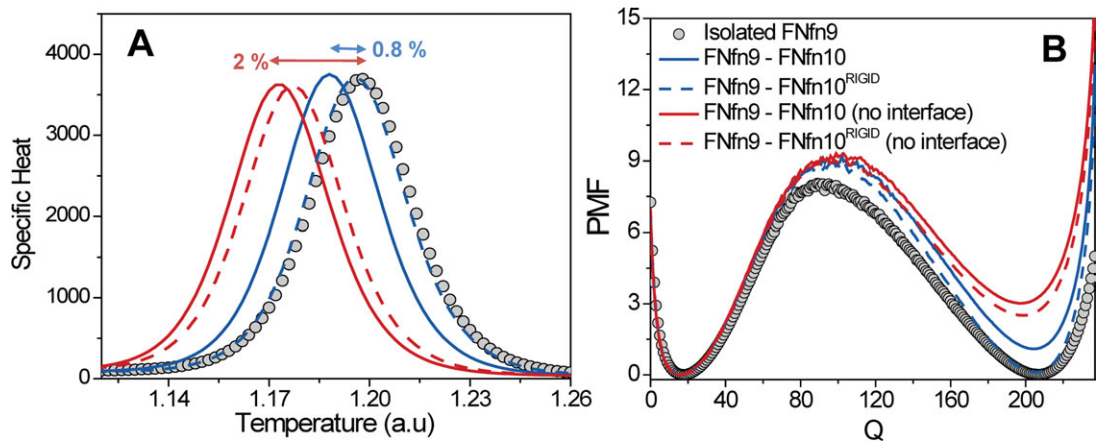

Figure 2

Thermostability of FNfn9 variants. (A) Specific heat versus temperature curves for several variants of FNfn9: isolated FNfn9 (gray circles), FNfn9 tethered to FNfn10 (full line), FNfn9 tethered to a permanently folded FNfn10 (dashed line); however, the interfacial interactions are included or excluded (blue and red lines, respectively). The peak of these curves corresponds to the transition folding temperature (T_F) at which proteins have zero stability (i.e., $\Delta G = 0$). The magnitudes of destabilization of the tethered variants with or without the interface are 0.8 and 2.0%, respectively. (B) Potential of mean force as a function of the number of native contacts, Q, of isolated or tethered FNfn9 variants.

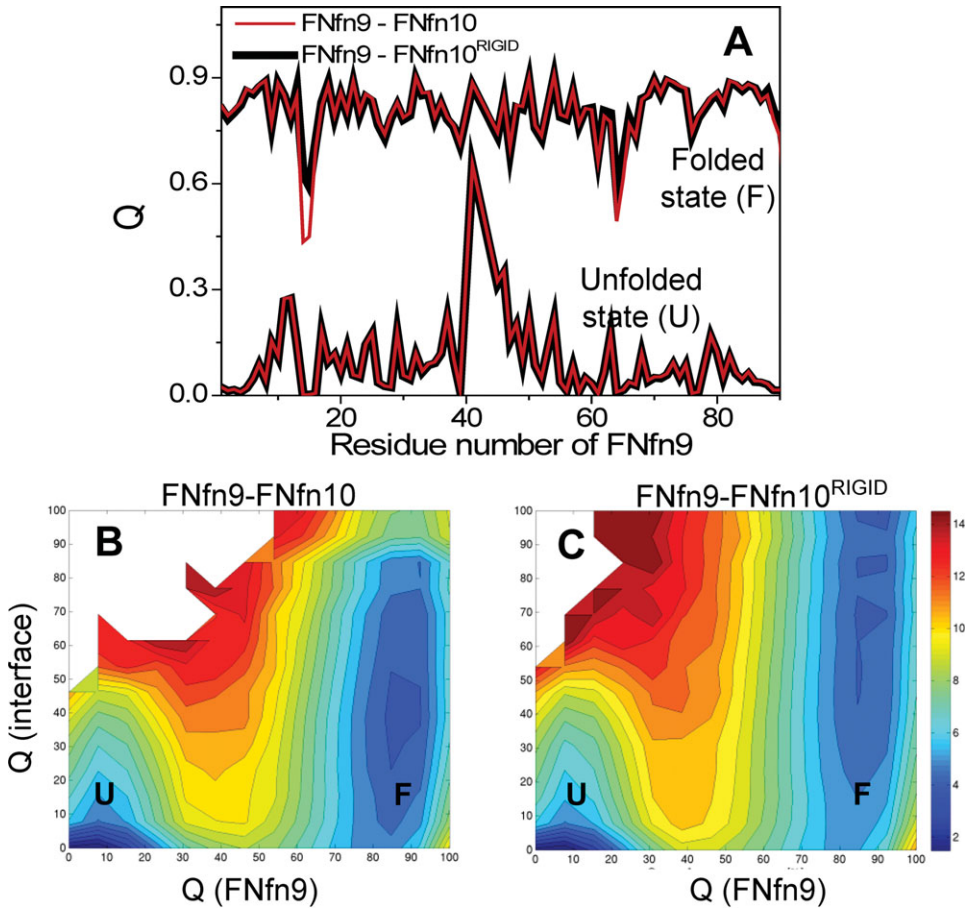

Figure 3

The effect of tethering on the structure of the folded and unfolded ensembles of FNfn9. (A) Structural variations of FNfn9 contacts in the folded and unfolded ensembles when they are tethered to a dynamic (red) or rigid (black) FNfn10. Indeed, interfacial contacts can be found surrounding residues 15, 65 and the tethered terminal of the FNfn9 domain. (B and C). Folding free energy landscapes of FNfn9 projected along the intramolecular contacts of FNfn9 [*Q*(FNfn9)] and the interfacial contacts it forms with FNfn10 [*Q*(interface)] when it is flexible (B) or rigid (C). The contours (black lines) represent isolines of (equal) probability.

Although the construction of the interface in FNfn9–FNfn10 does not substantially catalyze the folding of the constituent domains (Fig. 3(B) and Fig. 2(B), blue vs. red lines), as for instance, the C-domain catalyzes the folding of the N-domain in the lens protein γD-crystallin,[57] the interface of this construct is still expected to contribute to thermostabilization. As discussed earlier, the extent of interfaces formed between adjacent domains may be illustrated in their topology. The interface in the FNfn9–FNfn10 construct, as evident from its resolved crystal structure, is rather small. The interface in FNfn9–FNfn10 includes 19 contacts out of 495, using the Contacts of Structural Units analysis,[58] which constitute 3.8% of all native contacts (whereas large, tightly packed interfaces have contacts covering more than 10%). The relatively small interface in FNfn9–FNfn10 is also reflected by other measures such as packing density, as indicated by its local atomic density.[8,59]

In a desire to quantify the stabilizing effect of interfacial contacts in this construct, we simulated a variant of the FNfn9–FNfn10 construct with these contacts removed (and replaced them by repulsions between relevant beads). Figure 2(A) shows a substantial decrease in the stability of the FNfn9 domain when it is tethered to FNfn10 without the stabilizing contribution of the interfacial contacts. The relative decrease in stability between the tethered FNfn9 domain with the removal of interfacial contacts is significantly larger (T_F decreased by ~2.0% vs. the isolated variant) than the decrease observed following simply tethering the domains (in which the T_F decreased by ~0.8% vs. the isolated variant).[43] Therefore, a moderately sized interface (i.e., with more than the 19 contacts found in the X-ray structure of FNfn9–FNfn10) would have been able to compensate and mask out the above-demonstrated destabilization caused by tethering. Yet, contrary to the previously reported predictions concerning "sums of their

parts" constructs,[18] this small interface significantly contributes to the observed stability of the domains in this architecture, without which the calculated destabilization would have been substantially larger.

The effect of the relative thermostability on the constituent domains

The crosstalk between tethered domains can be greatly affected by differences in their relative thermostability. The tethered domains may interact differently when both are folded versus scenarios in which one or both of the domains are unfolded. In the FNfn9–FNfn10 construct, it has been experimentally shown that the FNfn10 domain is considerably more stable than the FNfn9 domain.[43,47] To account for the high thermal stability of FNfn10 compared with FNfn9,[32] we designed FNfn10 to be petrified in the folded state (for more details, see **METHODS** section), which means that any interplay between FNfn9 and its tethered neighbor will occur when the latter is permanently folded. This situation better distributes the thermal stabilities between the two domains of this construct. Figure 2 shows that when FNfn9 is tethered to a permanently folded FNfn10 (including interfacial contacts), the original decrease in stability seems to be compensated for. Moreover, the almost identical stability of this variant and that of an isolated FNfn9 is in agreement with the experimental results that call for "sum of its parts" thermodynamics for the FNfn9 domain.[19,43]

The probability of native contact formations throughout the simulations can aid in gaining a structural understanding of folding of tethered proteins. We carried out this analysis for the FNfn9 domain when it was tethered to the permanently folded FNfn10 and also when the FNfn10 was free to unfold. Figure 3(A) shows that the probabilities for contact formation are the same in both FNfn9 variants in the unfolded state and also in the folded state with the exception of those residues involved in interfacial contacts. Figure 3(A) shows that a folded FNfn10 assists in populating a higher percentage of interfacial contacts when the FNfn9 domain is folded and also, to some extent, when it is unfolded. Consequently, additional folding pathways that involve interfacial contacts emerge. Overall, it seems that being tethered to a folded domain (which is relatively more stable), as opposed to being tethered to a dynamically free domain, has a local effect on the FNfn9 domain.

The interface and relative thermostability: dependent effects, but not of identical origin

Both the involvement of interdomain interfacial contacts and being tethered to a permanently folded domain contribute to the stabilization of multidomain proteins. We have already shown that these effects are linked; however, do they share a common origin?

To address this question, we simulated another variant of the FNfn9–FNfn10 system in which the interfacial contacts are excluded, whereas FNfn10 is permanently folded (Fig. 2). A comparison of the specific heat curves of the different variants of FNfn9 [Fig. 2(A)] indicates that the difference in T_F between the tethered variants possessing or lacking interfacial contacts is not independent of the effect created by tethering to a folded domain, as was also demonstrated earlier. Although being tethered to permanently folded FNfn10 (i.e., rigid FNfn10—dashed blue line) significantly stabilized FNfn9 (reducing the destabilization caused by tethering from ~0.8% [blue line] to ~0.08% [dashed blue line], which is within our margin of error for the result obtained for the isolated variant [gray circles]), this effect is weaker when the interfacial contacts were excluded, amounting to about a third of the difference (reducing the destabilization of tethering without interfacial contacts from ~2.0% [red line] to ~1.7% [dashed red line]). Apparently, without the involvement of interfacial contacts, tethering to a permanently folded domain affects thermostability to a much more limited extent. This result is consistent with the locality we ascribed to the tethered to a permanently folded domain modification as being mostly limited to residues involved in interfacial contacts (Fig. 3). Yet there is an additional stabilizing factor that is contributed from being tethered to a permanently folded domain, which does not entail the involvement of interfacial contacts that we will discuss later in this article.

Folding of "sum of its parts" multidomain proteins: homo two-domain proteins

As we have shown, the stabilizing interfacial contacts and the relatively high thermostability of the neighboring domain can combine to compensate for a consistent, yet unaccounted for, destabilization caused by tethering. Continuing our investigation, we will attempt to further dissect the different factors involved in the biophysics underlying the folding of tethered two-domain systems. For this purpose, we decreased the complexity of the analysis by studying a *homo* two-domain system in which the component domains are covalently linked to each other without forming interfacial contacts. More specifically, we worked with two I27 domains interconnected by a two-residue linker. As the two tethered domains are identical, no relative thermostability issues arise. The two-domain model used was designed *in silico* by tethering two identical I27 domains that had the same native contact list as the isolated domain. No additional contacts were allowed between the linker and the domains, or for that matter, between the domains themselves, that is, all interfacial contacts were excluded. Therefore, this *homo* two-domain construct is conjugated solely by the linker's chain geometry.

The degree of flexibility of the linker

Linker dynamics may play a significant role in the biological function of multidomain proteins. Although the

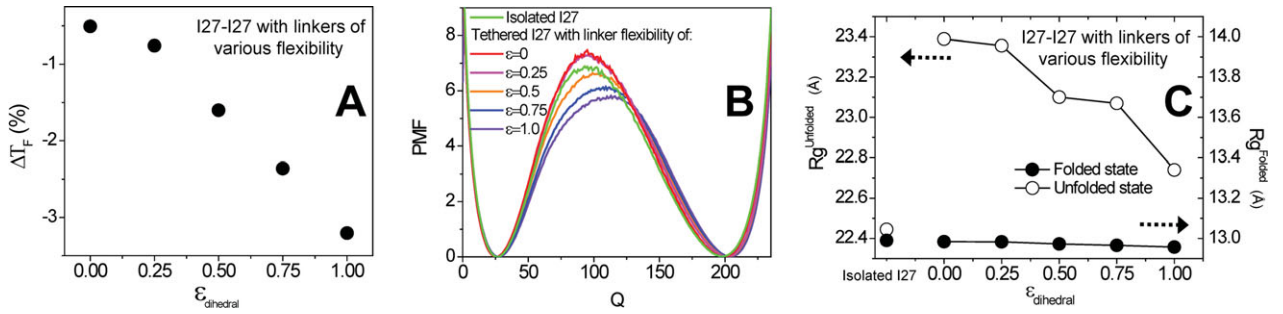

Figure 4

The effect of linker flexibility in I27–I27 on domain folding. (A) Correlation between the change in I27-tethered variants' thermostability and $\epsilon_{dihedral}$ of the dihedral angle of the linker (the lower the value of $\epsilon_{dihedral}$, the more flexible the linker is). The change in stability is calculated by $\Delta T_F = (T_F^{tethered} - T_F^{isolated})/T_F^{isolated}$. (B) Potential of the mean force profiles versus the number of native contacts at each system T_F. The position and height of the transition state seem to be mostly affected by the change in linker flexibility, corresponding to a higher barrier and less structure for more flexible linker. The isolated variant (marked in green) closely resembles a flexible linker. (C) The effect of the linker's flexibility on the compactness of the folded and unfolded ensembles of I27 when tethered to another I27 domain. A linear fit results in a negative slope for both the folded and unfolded state respectively (-0.03 $R^2 = 0.95$; -0.63 $R^2 = 0.89$).

structure of the I27–I27 two-domain system is unknown, homologous tethered domains in titin were shown to weakly interact with only a small interface. We attempted to quantify the effect of linker flexibility on the thermodynamics of tethered domains.

Figure 4 shows how changing the degree of flexibility affected the thermostability of each of the domains. There is a linear correlation ($R^2 = 0.97$) between increasing the rigidity of the linker and destabilizing the domain. The decrease in thermostability is quite substantial (>3% for $\epsilon_{dihedral} = 1$) [Fig. 4(A)]. Yet, even the most flexible variant ($\epsilon_{dihedral} = 0$) has a lower thermostability (~−0.5%) than the isolated I27 [Fig. 5(A)]. The free energy barriers are also affected by linker flexibility. Examining the PMF of all variants at their T_F suggests a change in the folding mechanism because the transition state has more free energy and becomes less structured with the more flexible linkers [Fig. 4(B)].

Plotting the R_g distribution of each individual domain shows a clear trend: the more rigid the linker is, the more compact both the unfolded and the folded ensembles of a single I27 domain are (more so if the neighboring domain is folded-data not shown) [Fig. 4(C)]. We suggest that a more rigid linker couples the folding behavior of the two tethered domains to a larger extent, resulting in a more structured folded and unfolded ensemble.

The effect of intrinsic self-crowding owing to tethering

The tethering of domains in a multidomain protein induces a high local concentration, because the neighboring domains are in immediate proximity of each other. A

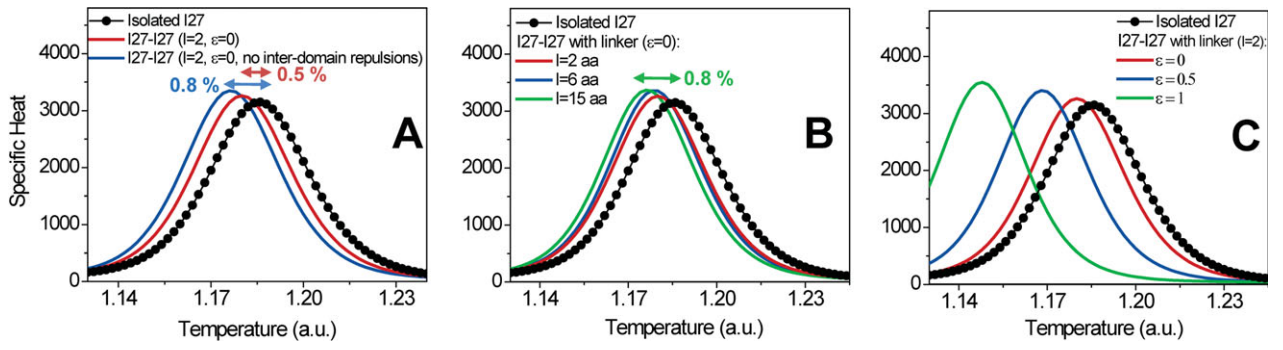

Figure 5

The effect of the linker's characteristics on the thermostability of I27 in I27–I27. (A) The effect of excluding the excluded-volume effects between the tethered I27 when the two-bead linker is flexible (no dihedral angle potential, $\epsilon = 0$). (B) The effect of the linker's length on thermostability. Linkers of 2, 6, and 15 residues were studied. (C) The effect of linker flexibility on thermostability (linker of two residues and an $\epsilon_{dihedral}$ of 0, 0.5, and 1).

possible outcome of this is that protein domains may have a higher tendency to aggregate while they are part of a multidomain architecture. Accordingly, it has been postulated that adjacent domains are under greater evolutionary pressure for their sequences to diverge than are nonadjacent domains. This claim was supported by experimental observations in which it was shown that tethered immunoglobulin domains with similar sequences (sequence identity, >40%) have a higher tendency to coaggregate[60] via swapping homologous structural elements.[61–64] Of course, the proximate presence of neighboring domains may have other more spatial consequences. Experimentally, covalently linked dimers via side chains of cysteine residues were shown to have lower stabilities relative to monomers.[65] This effect is sharply attenuated in dimers having longer linkers. The authors attributed the unexpected destabilization to a reduction in hydrophobic exposure (accompanying the reduction in the volume accessible to the unfolded state), which counteracts some of the expected stability gains from lost conformational entropy of the unfolded state. As our model does not account for non-native interactions, we can address the crowding feature separately. We examined to what extent the adjacent tethered domain acts as a crowding agent on its own neighbor.[66,67] Within this framework, we tried to account for the excluded volume attribute (the volume occupied by the domain tethered to the studied domain) and its influence on the folding behavior.

The potential volume exclusion by an adjacent domain can be studied in several ways; perhaps the most straightforward one is by removing the excluded volume interactions. Removal of such interdomain repulsions significantly alters the space occupied by the two-domain construct, because then basically each domain is transparent to its neighboring domain and the distance between their centers of mass can be quite short. However, as specific heat plots reveal, eliminating the volume exclusion results in only a small destabilization in comparison with the tethered variants that do exclude each other [Fig. 5(A)] (an additional destabilization, $\sim$ 0.3%). For the purpose of this comparison, we worked with a variant in which the linker, which was short and flexible, maximized the crowding effect (linker of two beads [l = 2] with entirely flexible dihedral angles [$\epsilon_{\text{dihedral}} = 0$]). Collisions between beads of adjacent domains are less likely within the variant of the fully extended rigid linker ($\epsilon_{\text{dihedral}} = 1$).

Extending the linker that bridges the two domains is another way to test the influence of crowding. Longer linkers, which are composed of a larger number of beads, increase the effective distance between the two domains and increase their similarity to isolated domains because they are then less crowded by the adjacent tethered domain. We tested the effect of using six-bead and 15-bead-long linkers, as opposed to the two-bead-long real-case linker [Fig. 5(B)]. Extending the linker consistently added only a small degree of destabilization. The variant with a six-bead linker is destabilized relative to the two-bead tethered domains by only an additional $\sim$0.2%. In the case of the 15-bead linker, we measured a slightly higher destabilization, which is the same (and also originates from the same enthalpic and entropic contributions) as the former case in which repulsions between domains were removed.

In further investigating the surprisingly limited effect of crowding by an adjacent domain, we restrained the distance between the centers of mass of two unlinked I27 domains by a harmonic constraint. In the previous simulations where a flexible two-bead linker was used, the average distances between the domains' centers of mass were 39 $\pm$ 7 and 46 $\pm$ 18 Å for the folded and unfolded ensembles, respectively. Accordingly, to examine crowding at shorter effective distances, we restrained the distance between the two unlinked domains to 40 Å (approximately the distance between the two folded linked domains of the previous simulations), 25, 20, and 15 Å and also to 350 Å (as a reference point for domains separated by a very large distance).

The stabilities of domains restrained to 40 Å are only slightly elevated in comparison with the two distant, greatly separated domains. Greater stabilities are found at a separation of 25 and 20 Å, but stability drops when a 15 Å separation distance is used [Fig. 6(A)]. However, by examining the enthalpic and entropic contributions [Fig. 6(B)], we can identify a trend of decreased enthalpy and entropy when the two domains are held closer together. Although the overall decrease in enthalpy governs stabilization at a separation of 25 and 20 Å, the decrease in entropy overcomes the enthalpy effect at 15 Å. This trend originates from a reduction in the residual structure of the unfolded state and, to a lesser extent, from an increase in the folded state (data not shown). The less structured unfolded state is associated with higher entropy, in contrast to the expected lower entropy of crowding conditions. This sort of architecture appears to interfere with the ability of the multidomain protein to maintain its residual structure as each domain unfolds. The system constrained to a distance of 40 Å (approximating the distance between two folded domains in our model system) does not follow this trend because it is situated at the opposite side of the reference at a distance of 350 Å (the origin 0,0) [Fig. 6(B)]. The small stabilization associated when the distance between the center of mass is constrained to a value of 40 Å, in comparison to a distance of 350 Å, originates primarily from a small reduction in the residual structure of the folded state accompanied by an increase that dominates its entropy (data not shown). We suggest that this effect is owing to the addition of collisions when the two domains are brought together, which slightly unwinds the folded state. Overall, the crowding effect, if any, is small, and the

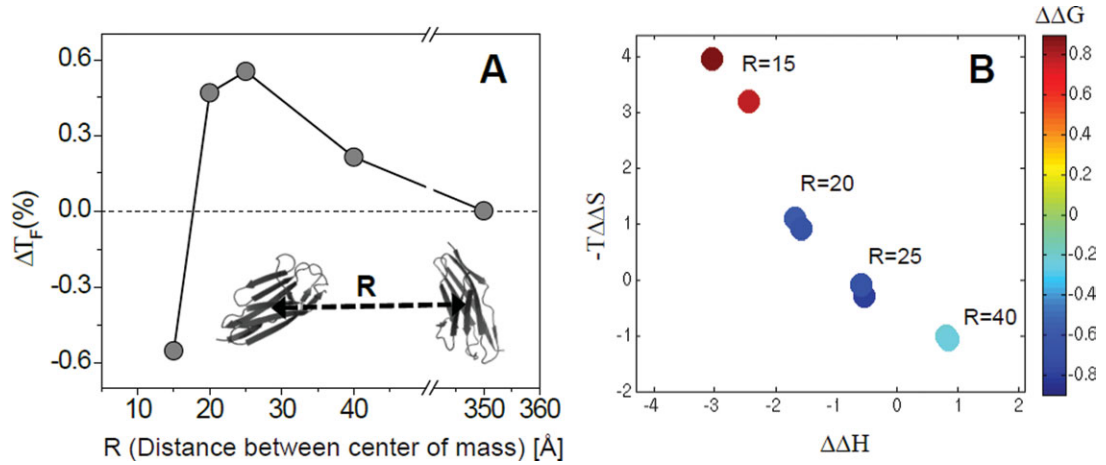

Figure 6

Crowding effect by increasing the local concentration of the subunits. Thermodynamic analysis as a function of center of mass distance between the two constituent domains of two-domain proteins. (A) The change in thermostability of I27 variants with different centers of mass constraint. The change in stability is shown by the percentages of $\Delta T_F = (T_F^{CM\ constrained} - T_F^{d\ =\ 350[Å]})/T_F^{d\ =\ 350[Å]}$. (**B**) Decomposition of free energy to enthalpic ($\Delta\Delta H$) and entropic ($-T\Delta\Delta S$) contributions, where $\Delta\Delta X = (\Delta X^{CM\ constrained} - \Delta X^{d\ =\ 350[Å]})$. The graphs are plotted for T_F at a distance 350 Å.

excluded volume by an adjacent domain corresponds to only a slight increase in stability. Why is that the case?

Crowding is expected to stabilize folded proteins by limiting the number of conformations the unfolded state ensemble can adopt, consequently, decreasing the entropy associated with the unfolded state and hence destabilizing it. As the unfolded state has many more conformational degrees of freedom than the folded state has, the effect of volume exclusion on the unfolded state is expected to govern the overall outcome and to result in protein stabilization. Therefore, one can reformulate the query regarding the small stabilization contributed by volume exclusion that our results raise by considering to what extent the unfolded state of a given domain is limited by its neighboring domain. If the space required for the unfolded state to react as if it was isolated is not significantly limited by the constraints of volume exclusion, then the expected stabilization will be small.

The domains are tethered at one of their termini, which do not impose crowding in all directions. As any stabilization arising solely from volume exclusion is relatively small, the conformational entropy of each domain by itself is only slightly altered, that is, the volume occupied by the neighboring domain only slightly affects the studied domain configurations. Therefore, the potential conflict with crowding theory can be simply resolved by differentiating between the two cases. In the first case, crowding is symmetrically applied from all directions. Here, the expected result will still be for the unfolded state ensemble to maintain a more residual structure. In the second case, directional crowding exists in which the asymmetric volume exclusion interferes with the structuredness of the unfolded state.

Linker-mediated dynamics coupled to the tethered domains

As we described in the **INTRODUCTION** section, the linker may also couple the dynamics of the domains. Movements of one domain or of the linker itself may be transferred to the other domain.[17] Therefore, examining how the average internal distances in a domain are affected solely by tethering (i.e., in comparison with isolation) can be used to probe the molecular communication between the two domains. This may facilitate investigating whether the disturbance caused by attaching another foldable domain accounts for the consistent destabilization, which was reported here.

The differences between the intramolecular distances of the tethered domain and the isolated I27 variant are manifested by the Δ distance matrices for the folded state ensemble and the unfolded state ensemble (Fig. 7). In the folded state, tethering results in local unwinding (extension relative to the isolated variant) at the termini of attachment. The region around the point of conjugation is locally affected by merely attaching a linker and a consecutive domain. In the unfolded state, a much larger distortion is observed, with more extensive and remote regions being drawn further away. We will refer to the observed expansion of the termini as a pulling effect, as the studied domain is being pulled by the adjacent domain at the terminus of attachment. We suggest that the high degree of locality of the pulling effect on the folded state, compared to the extensiveness of the effect on the unfolded state, is associated with the folded state structured nature, in which contacts, as scaffolds, prevent the ripple from progressing. This statement is also supported

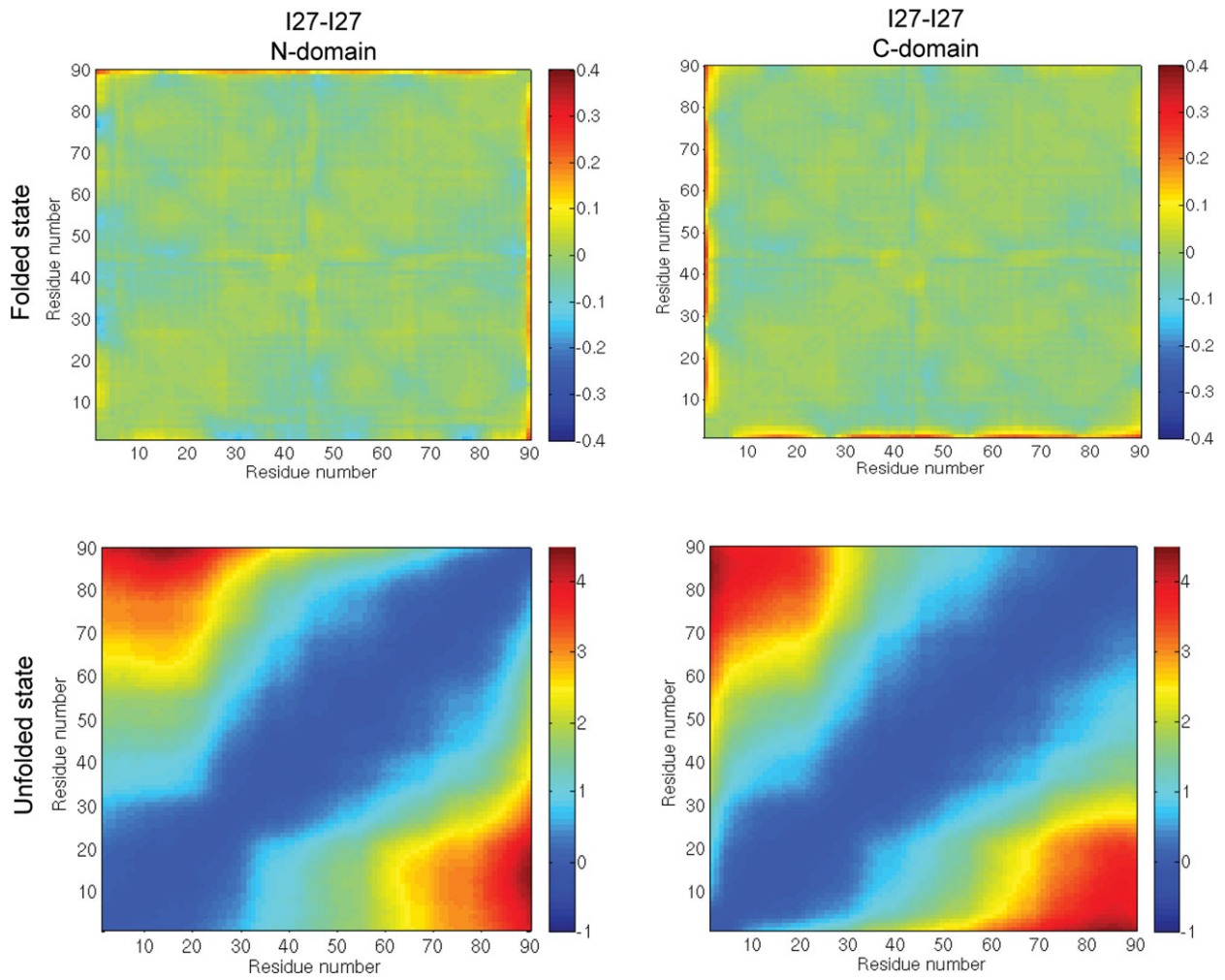

Figure 7

The effect of tethering on the dynamics of the folded and unfolded state of I27. Differences in dynamics between tethered and isolated I27 are analyzed separately for the folded and unfolded states (top and bottom panels, respectively) and for the N- and C-domains (left and right panels, respectively). The Δ distance matrix for state *X* (folded or unfolded state of a given domain) was calculated by $<\Delta_{ij}>=<R_{ij}^{X}> - <R_{ij}^{\text{Isolated I27}}>$; where $<R_{ij}>$ is the average distance between residues i and j. Matrices were calculated at the folding temperature, T_f, of the isolated I27. For an easier comparison, all differences in distances are shown in the range of −0.4 to +0.4 in the folded state, and −1 to +4.5 in the unfolded state (distances in Å are indicated by color).

by the findings of a recently published article[31] in which the stability of an ubiquitinated protein or of a cross-linked nuclease dimer[65] was strongly negatively correlated with their structural density at the tethering site, measured as the number of native contacts at these sites in the folded state. The reduced stability with high-structural density was attributed to a more easily distorted folded state and a more unwound and less-structured unfolded state, consequently, with a stronger pulling.

The pulling effect, as shown in Figure 7, also aids in clarifying our prior results. In the previous sections, we demonstrated consistent destabilization as a consequence of simply tethering two domains. This result is correlated with a dominant increase in unfolded state entropy (i.e., associated with a net decrease in entropy), as indicated by the Δ distance matrices (Fig. 8). In contrast, a reduction in the residual structure of the unfolded state and an increase in the residual structure of the folded state (note that some regions in the folded state are slightly closer together) are in agreement with a net decrease in enthalpy (Fig. 8). This behavior is also supported by the compactness of folded and unfolded states measured by the radius of gyration, whereas the unfolded state of the isolated variant seems to be the most compact, compared with the tethered variant, its folded state is the least compact (Fig. 4).

Indeed, the pulling effect is not limited to the flexible linker ($\epsilon_{\text{dihedral}} = 0$) variant, because for the variant with a more rigid linker ($\epsilon_{\text{dihedral}} = 1$) a similar effect can be seen (Fig. 9). However, consistent with the calculation of more compact unfolded and folded states versus the $\epsilon_{\text{dihedral}} = 0$ variant (Fig. 4), the pulling effect on the unfolded ensemble is smaller for the $\epsilon_{\text{dihedral}} = 1$ variant

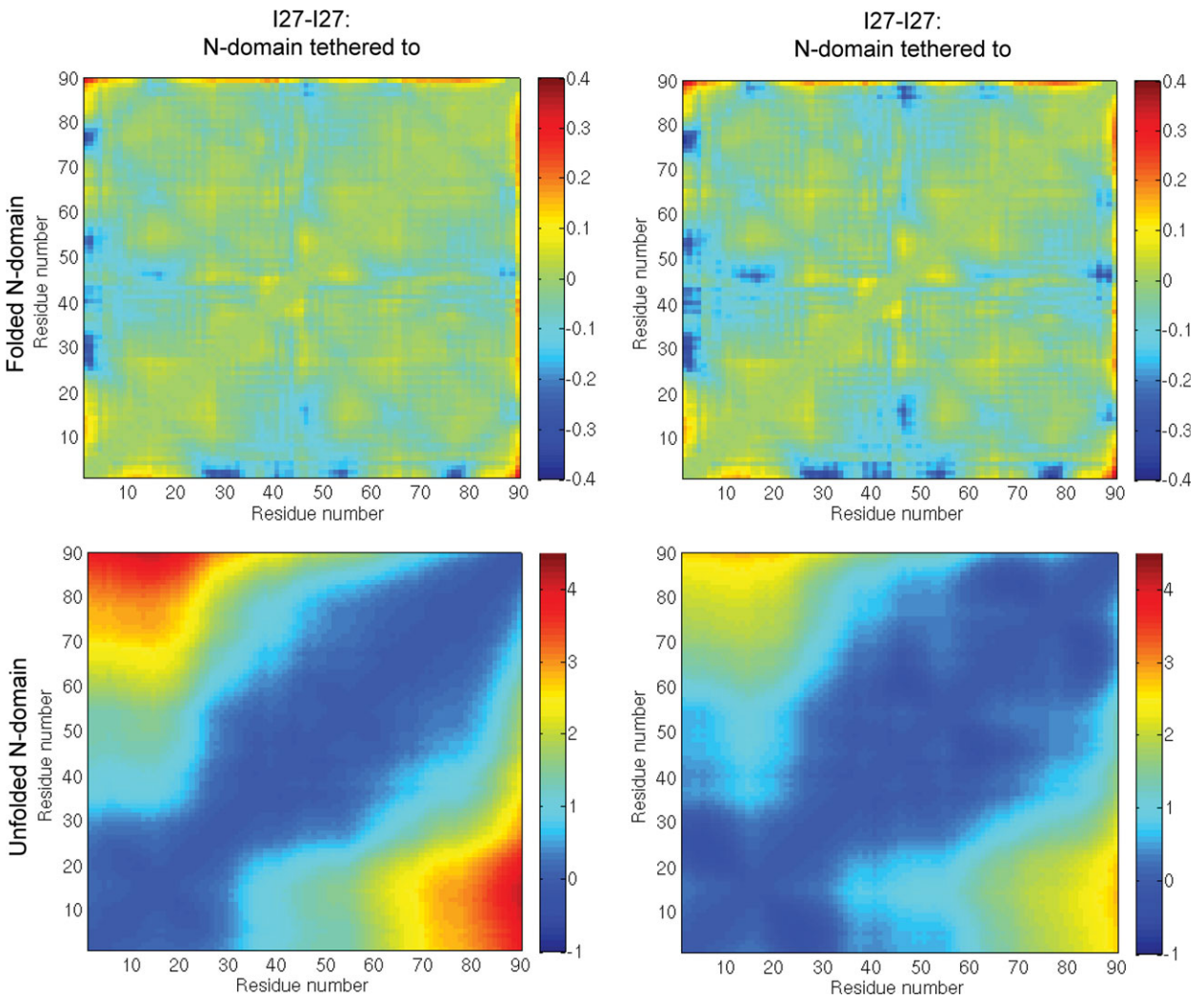

Figure 8

The effect of the state of the neighboring domain on the dynamics of an I27 domain. The Δ distance matrices of folded ensemble (top panels) and unfolded ensemble (bottom panels) of I27 (at the N-terminal) tethered to an unfolded (left column) or a folded (right column) I27 (located at the C-terminal). Matrices were calculated at the folding temperature T_F of the isolated I27. Differences in distances are shown from −0.4 to +0.4 in the folded state and −1 to +4.5 in the unfolded state, as shown in Figure 7.

(Fig. 9). This suggests that a rigid linker is less effective in mediating movements, thus it pulls less. Explicitly, a flexible linker can carry the motions of one domain and of the linker itself to the adjacent tethered domain more effectively than a stiffer linker.

The pulling effect is not, however, limited to the I27–I27 model; it can also be demonstrated in the FNfn9–FNfn10 construct. The Δ distance matrices of FNfn9 tethered to the permanently folded FNfn10 indicate that the pulling effect on the unfolded state results in less distortion (data not shown). This weaker pulling effect may account for the additional stabilizing effect of being tethered to a permanently folded adjacent domain (Fig. 2(A), red line vs. dashed red line, i.e. without the involvement of interfacial contacts). Accordingly, when examining the N-domain of the I27–I27 construct of Figure 7, further divided by the state of the adjacent domain, one can observe that a folded adjacent domain (Fig. 8, right panels) exerts less pulling than does an adjacent unfolded domain (Fig. 8, left panels). As a folded adjacent domain is more constrained in its motions than is its unfolded equivalent, it affects the studied domain less. Plotting the Δ distance matrices for a variant with the rigid linker, while implementing the same strategy of dividing by the state of the adjacent domain, also indicates that a larger pulling effect is associated with tethering to an unfolded domain (Fig. 9). However, an additional effect is revealed by these distance matrices regarding which residues come closer together to a larger extent while the adjacent domain is folded. We suggest that a rigid linker enables the adjacent

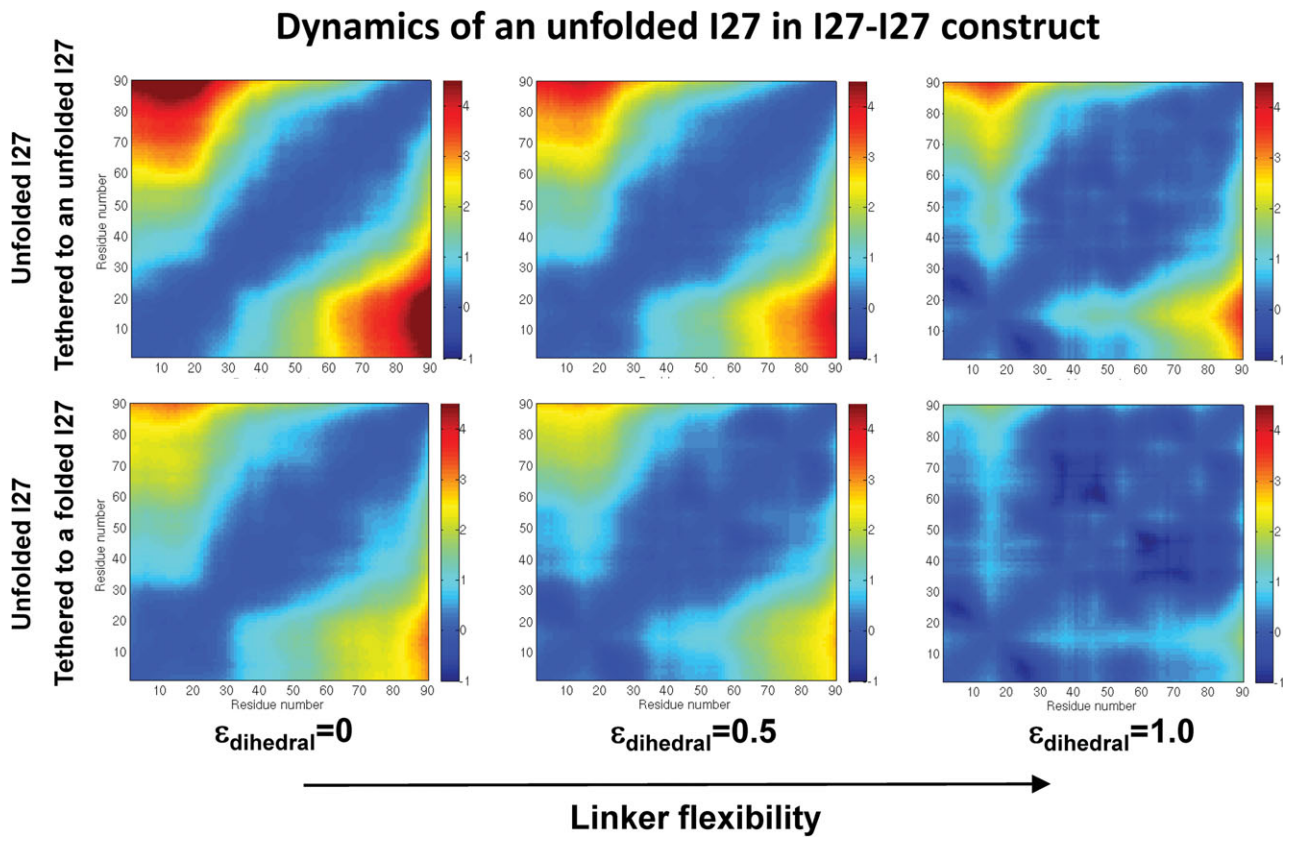

Figure 9

The effect of the linker's flexibility on the internal dynamics of an I27 domain in the I27–I27 construct. The Δ distance matrices of the unfolded ensemble of the N-domain of the I27–I27 construct tethered to the folded or unfolded C-domain I27. The Δ distance matrices are calculated relative to the pairwise distances in the unfolded state of isolated I27. Matrices were calculated at the folding temperature, T_F, of the isolated I27. All differences in distances are shown in ranges of −1 to +4.5 in the unfolded state. $\epsilon_{dihedral}$ is the coefficient of the dihedral angle term between residues of the linker, as they represent the linker's degree of flexibility; a higher $\epsilon_{dihedral}$ results in a more rigid linker.

domain to retain its structure and that the linker itself induces structure in the studied domain. Thus, although the pulling effect couples the movements of tethered domains and, as we previously stated, a rigid linker pulls less, the additional observed effect indicates an association between linker rigidity and the degree of structure retained by the tethered domains, such that the overall structure of the whole multidomain construct becomes more rigid itself.

Halting the pulling effect in tethered proteins

As we have shown that the pulling effect originates directly from the linker and its tethering points, we can attempt to halt its interference by artificially maintaining the contacts of the termini formed. We therefore carried out simulations in which we ascribed to the contacts of the two residues at each terminus a bond-like harmonic potential, ensuring that they remain constantly formed. These simulations produced the same stability as was seen for the isolated variant (namely, $\Delta T_F = 0\%$, Fig. 10). For controls, we ran simulations of permanently formed contacts in eight different regions of each domain other than the natural attachment termini of multidomain proteins. Figure 10 shows the uniqueness of the near linker termini. In fact, all other variants show destabilization resembling that of the unmodified tethered variant, regardless of the number of contacts associated with the modified residues. Apparently, preserving the contacts at the domains' attachment termini not only prevented pulling but also completely isolated the domains from sensing the adjacent domain. This result may suggest that the gatekeeper residues located at domain–domain boundaries play a different role. The unwinding, initialized at the termini by pulling, causes interference that progress throughout the protein chain. As this process is probably not correlated with the folding pathways, it may increase the propensity for misfolding and aggregation. For example, proline residues located at both fnIII domain–domain boundaries may set constraints on the backbone conformations, which presumably may isolate and prevent the linker from mediating movements. This scenario will also result in a diminished pulling effect.[43]

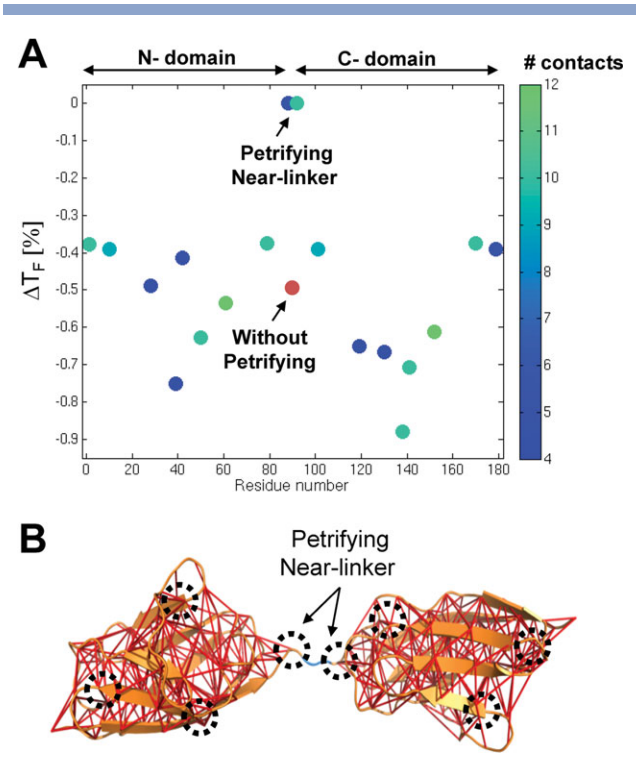

Figure 10

Thermostability of various tethered I27–I27 with constrained residues. (A) In each construct, the contacts associated with two beads (one from each domain) were petrified to examine their effect on the intrinsic destabilization induced by the tethering. The difference in T_f from a tethered system in which the contacts of two beads are permanently formed versus isolated I27 from the same contacts formed; $\Delta T_F = (T_F^{\text{tethered}''} - T_F^{\text{isolated}''})/T_F^{\text{isolated}''}$ as a function of the two-bead numbers. Color indicates the number of contacts associated with the constrained bead. The ΔT_F of the tethered I27 without petrifying (namely the unconstrained protein) is shown in red. The tethered protein with petrifying the residues at the C- and N-termini that directly involve in the tethering are indicated as petrifying near-linker. (B) A schematic illustration of some of the petrifying sites.

Another possible way to substantially balance the pulling effect can be achieved by artificially confining the unfolded state. By posing an external limit on the maximal size that the unfolded state may occupy, the pulling effect will be halted. To test the validity of this argument, we simulated two domains confined in a cage having a radius of 30 Å. This confined sphere radius corresponds to the tail of the distribution of two folded tethered domains, inflicting a real constraint on the unfolded ensemble. If our argument is valid, then two tethered domains or two unlinked domains confined to this cavity will exhibit similar stability, as the effect of tethering on the unfolded state will be negligible. This was indeed the result we obtained, with tethered domains in the cavity destabilized by only 0.2% compared with two unlinked domains when placed in the same confined cavity (data not shown). We wish to point out that tethering may, on the other hand, increase stability in cases where it may support the formation of an extensive interface between the two subunits,[68] as the tethering directs the preferable orientation between domains and therefore reduce the search time for the formation of interfacial contacts.

Estimating the size of the interface in multidomain proteins: atomistic simulations

Simulations at an all-atom resolution were carried out to determine the structure of the interdomain region of two conjugated I27 domains, that is, the I27–I27 model system, whose actual structure has not been resolved experimentally. The all-atom simulations revealed that the I27–I27 construct encompasses a relatively local crosstalk, which is limited to the interdomain connecting region and involves only residues near the linker (Fig. 11). In agreement with previously reported results that the I27–I27 domains weakly interact,[69,70] our results support the formation of an interface of limited size that strengthens the linker and related domain termini and therefore may minimize and compensate for a potential pulling effect.

Aiming to better understand the interdomain dynamics of two-domain protein systems, we used all-atom simulations to study two additional systems: FNfn9–1FNfn0 and TNfn3–TNfn3. Quantifying the size of the interface formed in these two systems is valuable because the former follows "sum of its parts" thermodynamics, whereas the latter is stabilized by tethering to another domain. For FNfn9–FNfn10, in particular, the 3D structure was solved. These all-atom simulations revealed complex behaviors across the interface. Following the number of interfacial contacts in these constructs, as a measure of the interface size, reveals that for substantial periods of the simulated trajectories the size of the interface of FNfn9–FNfn10 is considerably larger than that of the I27–I27 construct [Fig. 11(B)]. The size of the FNfn9FNfn10 interface is characterized by several modes, suggesting the existence of competing interactions. When the relative orientation of the domains in each of the constructs was compared, the interdomain region in these constructs was found to be rather flexible, in the spectrum of angles sampled (Fig. 11). However, unlike the I27–I27 construct, a preferred orientation for the FNfn9–FNfn10 domains exists. Overall, when this complex dynamics is considered in light of the similar "sum of its parts" end-result of both constructs, the considerably larger interface of FNfn9–FNfn10 versus the more flexible interdomain region of the I27–I27 construct dominates.

CONCLUSIONS

In this study, we have explored the underlying biophysical principles of the complex folding processes of multidomain proteins. Studying the folding of multido-

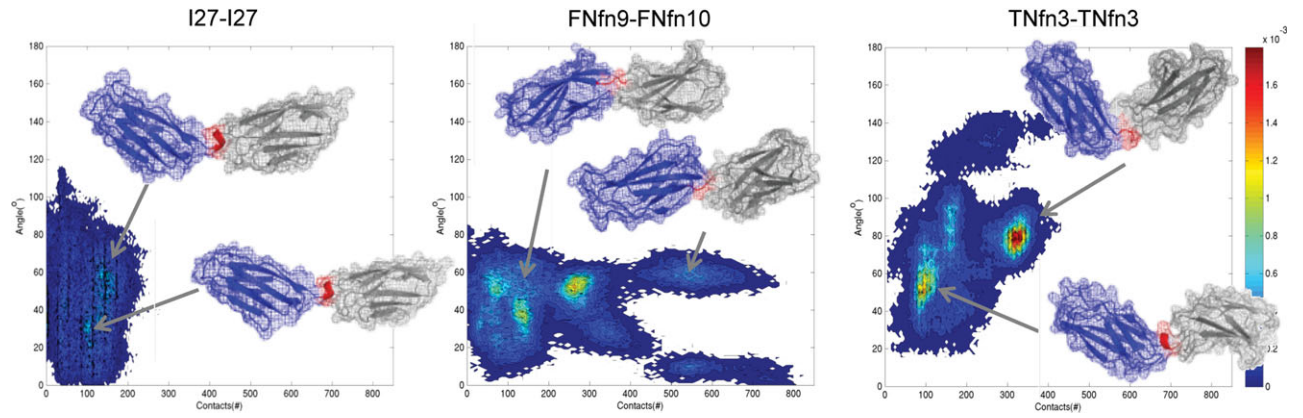

Figure 11

Atomistic modeling of the interface between three two-domain proteins. The interface in these two-domain proteins was analyzed by the number of interfacial contacts and the orientation angle between the two domains. A 3D probability histogram illustrates the angle formed between the vectors that connect the N- to C-terminal of each domain and the total number of interfacial contacts in a specific snapshot. Plots correspond to the last 80% (400 ns) of three trajectories obtained for each two-domain construct: I27–I27 (**A**) FNfn9–FNfn10 (**B**) and TNfn3–TNfn3 (**C**). The inset of each histogram shows representative conformations. Marked in red are the original two-residue linkers with an additional residue from each side. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

main protein systems using coarse-grained molecular dynamics allowed us to isolate the factors governing the crosstalk between tethered domains and hence learn the extent to which crosstalk influences the structure in these systems. Thus, we were able to show that tethering has various biophysical consequences that stem from microscopic origin.

Focusing on the "sum of its parts" two-domain proteins, we found that several competing effects come into play, which determine the thermostability of the constituent domains. This result is contrary to our current understanding, which assumes uncoupled folding behavior, with $\Delta G_{INT} = 0$.[18] Uncoupling between the thermodynamics of the neighboring domains can be considered as a reasonable mechanism, as in large multidomain proteins it might be advantageous for each domain to be individually stable, to avoid catastrophic domino effects upon mutation.[18] Moreover, and as is probably also important in an operative context, when domains are independent, the unfolding of one domain will not make the unfolding of its adjacently tethered neighbors more probable.[18] However, this argumentation is not supported by microscopic mechanistic evidence. The results reported here suggest the need to reconsider this view. In fact, we suggest that the experimental "tethered yet independent" perspective is the combined result of several folding determinants, which we have identified in this study, and which may mask out each other's effects. For example, in the "sum of its part" two-domain constructs, we studied an interface, which despite its small size, appears to be an essential contributor, as it can compensate for the disturbing effect of pulling. The all-atom simulations further indicate that, in the case of the I27–I27 construct, the linker region dynamics is limited, which can minimize the pulling effect in this architecture. Furthermore, with the FNfn9–FNfn10 construct, we observed an initial misfit to the experimental results, which was only overcome (and the experimental behavior reproduced) after we included the contributions of both the interface and the relative thermostability of the contributing domains.

The various determinants may combine to influence the crosstalks between domains in numerous ways and to various extents, depending on the specific characteristics of the inspected multidomain constructs. As a result, considerable diversity is expected. Therefore, identifying correlations between topology-driven crosstalks and measurable folding observables, such as thermostability and folding rates, is of major importance if one wishes to better understand the folding process in relation to multidomain proteins. In fact, the size and geometry of interfaces in a multidomain architecture may alter the folding pathways of tethered domains until they become very different from those of the isolated domains.[71] For example, in the yeast phosphoglycerate kinase (PGK), domain–domain interactions were shown to direct the folding of the N-domain along a completely different pathway from that of an isolated N-domain.[72] The closely interacting C-terminal helix and the N-domain are also responsible for placing the constituent domains in their preferable relative orientation. Consequently, this resulting orientation and topology set constrains on the dynamics, including the chosen pathways for folding. However, with the FNfn9–FNfn10 system, we showed that, although the interface contributes some stabilization, to a large extent interfacial contacts do not participate in the folding process, that is, the probable folding pathways do not involve these contacts. As a result, differences in the relative thermostabilities of the two tethered

domains can, by their nature, have only a local influence. However, as we demonstrated in this study, a relatively stable tethered neighbor increases the likelihood of altered folding pathways (Fig. 3). This effect will remain rather minimal unless the folding pathways of the protein domains have a tendency to be altered by the presence of their tethered neighbor. In such cases, the relative thermostability of the tethered domains can have a substantial influence. However, an open question is: why is the folding pathway of the N-domain of PGK altered, which does not occur in the FNfn9–FNfn10 construct. We suggest that the answer lies in the multidomain topology, specifically, in the interdomain interactions (not only regarding the extent but also regarding their distribution and geometry).

We also investigated the characteristics of the linker that interconnects the tethered domains and showed that a more flexible linker results in greater constituent domain stability, with decreasing flexibility linearly correlated with decreasing stability. It should be of interest to experimentally examine how the rigidity of the linker influences the biophysics of tethered domains by, for instance, incorporating double bonds or by relaxing interactions involving the linker. Such an approach may be of particular value from the perspective of protein design. Furthermore, we also discussed the role of the linker in mediating movement between the tethered domains, as manifested by a pulling effect. The structure of the linker also influences this role, as we demonstrated with our model system. For example, the rigid variant of the short linker is coupled to the movements of the domains to a lesser extent. However, an issue that remains to be investigated is will a structured linker be able to mediate movements or will it simply be able to dampen the carried movements. For instance, an intriguing question is whether a helical hinge, such as the linker in the PGK protein, mediates movements or whether it simply dampens the carried movements, as it is structured itself. We described other possibilities for the damping of the movements, which are carried out by the linker including the halting of pulling. From a protein design perspective, we suggest adding covalent bonds between the near-linker residues and their contacting residues (e.g., disulfide bonds), which will also remain in contact in the unfolded state ensemble and could resist the pulling effect.

Overall, it seems that tethering of domains may play a biophysical role in addition to enriching the functional diversity. Thus, the modified biophysics of domains, as part of multidomain architectures, should also be regarded as a result of selection under evolutional pressure. Attaining a better understanding of the complicated dynamics and folding behaviors of multidomain proteins should be considered as an important step toward approaching a more correct and realistic view of protein folding in the cell and the associated protein functions. The principles thereby revealed may also contribute to the design of new multifunctional proteins.

ACKNOWLEDGMENTS

The authors thank the members of our research group, in particular, Tzachi Hagai and Dalit Shental, for many insightful discussions. Y.L. is the incumbent of the Lillian and George Lyttle Career Development Chair.

REFERENCES

1. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 2004;14:208–216.
2. Wang M, Caetano-Anolles G. The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. Structure 2009;17:66–78.
3. Ekman D, Bjorklund AK, Frey-Skott J, Elofsson A. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. J Mol Biol 2005;348:231–243.
4. Apic G, Gough J, Teichmann SA. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. J Mol Biol 2001;310:311–325.
5. Levitt M. Nature of the protein universe. Proc Natl Acad Sci USA 2009;106:11079–11084.
6. Veretnik , Bourne PE, Alexandrov NN, Shindyalov IN. Toward consistent assignment of structural domains in proteins. J Mol Biol 2004;339:647–678.
7. Murzin A, Brenner S, Hubbard T, Chothia C. Scop—a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
8. Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J. The folding and evolution of multidomain proteins. Nat Rev Mol Cell Biol 2007;8:319–330.
9. Marsh JA, Teichmann SA. How do proteins gain new domains? Genome Biol 2010;11:126.
10. Vogel C, Berzuini C, Bashton M, Gough J, Teichmann SA. Supradomains: evolutionary units larger than single protein domains. J Mol Biol 2004;336:809–823.
11. Del Sol A, Arauzo-Bravo MJ, Amoros D, Nussinov R. Modular architecture of protein structures and allosteric communications: potential implications for signaling proteins and regulatory linkages. Genome Biol 2007;8:R92.
12. Bashton M, Chothia C. The geometry of domain combination in proteins. J Mol Biol 2002;315:927–939.
13. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:539–594.
14. Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol 2004;14:70–75.
15. Batey S, Scott KA, Clarke J. Complex folding kinetics of a multidomain protein. Biophys J 2006;90:2120–2130.
16. Steward A, Chen Q, Chapman RI, Borgia MB, Rogers JM, Wojtala A, Wilmanns M, Clarke J. Two immunoglobulin tandem proteins with a linking beta-strand reveal unexpected differences in cooperativity and folding pathways. J Mol Biol 2012;416:137–147.
17. Yuwen T, Post CB, Skrynnikov NR. Domain cooperativity in multidomain proteins: what can we learn from molecular alignment in anisotropic media? J Biomol NMR 2011;51:131–150.
18. Batey S, Nickson AA, Clarke J. Studying the folding of multidomain proteins. HSFP J 2008;2:365–377.
19. Scott KA, Steward A, Fowler SB, Clarke J. Titin; a multidomain protein that behaves as the sum of its parts. J Mol Biol 2002;315:819–829.
20. Gokhale RS, Khosla C. Role of linkers in communication between protein modules. Curr Opin Chem Biol 2000;4:22–27.
21. Wriggers W, Chakravarty S, Jennings PA. Control of protein functional dynamics by peptide linkers. Biopolymers 2005;80:736–746.

22. Ma B, Tsai CJ, Haliloglu T, Nussinov R. Dynamic allostery: linkers are not merely flexible. Structure 2011;19:907–917.
23. Blackledge M. Mapping the conformational mobility of multidomain proteins. Biophys J 2010;98:2043–2044.
24. Kim HJ, Choi MY, Kim HJ, Llinas M. Conformational dynamics and ligand binding in the multi-domain protein PDC109. PloS One 2010;5:e9180.
25. Hua L, Huang X, Liu P, Zhou R, Berne BJ. Nanoscale dewetting transition in protein complex folding. J Phys Chem 2007;111:9069–9077.
26. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. Proteins 2009;75:430–441.
27. Schug A, Whitford PC, Levy Y, Onuchic JN. Mutations as trapdoors to two competing native conformations of the Rop-dimer. Proc Natl Acad Sci USA 2007;104:17674–17679.
28. Schug A, Onuchic JN. From protein folding to protein function and biomolecular binding by energy landscape theory. Curr Opin Pharmacol 2010;10:709–714.
29. Gosavi S, Whitford PC, Jennings PA, Onuchic JN. Extracting function from a beta-trefoil folding motif. Proc Natl Acad Sci USA 2008;105:10384–10389.
30. Cho SS, Levy Y, Wolynes P. Quantitative criteria for native energetic heterogeneity influences in the prediction of protein folding kinetics. Proc Natl Acad Sci USA 2009;106:434–439.
31. Hagai T, Levy Y. Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. Proc Natl Acad Sci USA 2010;107:2001–2006.
32. Shental D, Arviv O, Hagai T, Levy Y. Folding of conjugated protein. Annu Rep Comput Chem 2010;6:263–277.
33. Shental-Bechor D, Levy Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. Proc Natl Acad Sci USA 2008;105:8256–8261.
34. Shental-Bechor D, Levy Y. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. Curr Opin Struct Biol 2009;19:524–533.
35. Mor A, Ziv G, Levy Y. Simulations of proteins with inhomogeneous degrees of freedom: the effect of thermostats. J Comput Chem 2008;29:1992–1998.
36. Kumar S, Rosenberg ,JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 1992;13:1011–1021.
37. Cho SS, Levy Y, Wolynes PG. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. Proc Natl Acad Sci USA 2006;103:586–591.
38. Yang SC, Cho SS, Levy Y, Cheung MS, Levine H, Wolynes PG, Onuchic JN. Domain swapping is a consequence of minimal frustration. Proc Natl Acad Sci USA 2004;101:13786–13791.
39. Takagi F, Koga N, Takada S. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. Proc Natl Acad Sci USA 2003;100:11367–11372.
40. Xu WX, Wang J, Wang W. Folding behavior of chaperonin-mediated substrate protein. Proteins 2005;61:777–794.
41. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–1718.
42. Sorin EJ, Pande VS. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophys J 2005;88:2472–2493.
43. Steward A, Adhya S, Clarke J. Sequence conservation in Ig-like domains: the role of highly conserved proline residues in the fibronectin type III superfamily. J Mol Biol 2002;318:935–940.
44. Spitzfaden C, Grant R, Mardon H, Cambell I. Module-module interactions in the cell binding region of fibronectin: stability, flexibility and specificity. J Mol Biol 1997;265:565–579.
45. Clarke J, Cota E, Fowler SB, Hamill SJ. Folding studies of immunoglobulin-like b-sandwich proteins suggest that they share a common folding pathway. Structure 1999;7:1145–1153.
46. Rounsevell RW, Steward A, Clarke J. Biophysical investigations of engineered polyproteins: implications for force data. Biophys J 2005;88:2022–2029.
47. Billings KS, Best RB, Rutherford TJ, Clarke J. Crosstalk between the protein surface and hydrophobic core in a core-swapped fibronectin type III domain. J Mol Biol 2008;375:560–571.
48. Ng SP, Rounsevell RW, Steward A, Geierhaas CD, Williams PM, Paci E, Clarke J. Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. J Mol Biol 2005;350:776–789.
49. Hamill SJ, Meekhof AE, Clarke J. The effect of boundary selection on the stability and folding of the third fibronectin type III domain from human tenascin. Biochemistry 1998;37:8071–8079.
50. Itoh K, Sasai M. Cooperativity, connectivity, and folding pathways of multidomain proteins. Proc Natl Acad Sci USA 2008;105:13865–13870.
51. Whitford P, Miyashita O, Levy Y, Onuchic JN. The conformational transition of adenylate kinase: switching by cracking. J Mol Biol 2006;366:1661–1671.
52. Hills RD, Jr, Brooks CL, 3rd. Subdomain competition, cooperativity, and topological frustration in the folding of CheY. J Mol Biol 2008;382:485–495.
53. Raman EP, Barsegov V, Klimov DK. Folding of tandem-linked domains. Proteins 2007;67:795–810.
54. Friedel M, Baumketner A, Shea JE. Effects of surface tethering on protein folding mechanisms. Proc Natl Acad Sci USA 2006;103:8396–8401.
55. Friedel M, Baumketner A, Shea JE. Stability of a protein tethered to a surface. J Chem Phys 2007;126:095101.
56. Zhuang Z, Jewett AI, Soto P, Shea JE. The effect of surface tethering on the folding of the src-SH3 protein domain. Phys Biol 2009;6:15004.
57. Mills IA, Flaugh SL, Kosinski-Collins MS, King JA. Folding and stability of the isolated Greek key domains of the long-lived human lens proteins gammaD-crystallin and gammaS-crystallin. Protein Sci 2007;16:2427–2444.
58. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327–332.
59. Bahadur RP, Chakrabarti P, Rodier F, Janin J. A dissection of specific and non-specific protein-protein interfaces. J Mol Biol 2004;336:943–955.
60. Wright CF, Teichmann SA, Clarke J, Dobson CM. The importance of sequence diversity in the aggregation and evolution of proteins. Nature 2005;438:878–881.
61. Borgia MB, Borgia A, Best RB, Steward A, Nettels D, Wunderlich B, Schuler B, Clarke J. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. Nature 2011;474:662–665.
62. Xia F, Thirumalai D, Grater F. Minimum energy compact structures in force-quench polyubiquitin folding are domain swapped. Proc Natl Acad Sci USA 2011;108:6963–6968.
63. Yang S, Cho SS, Levy Y, Cheung MS, Levine H, Wolynes PG, Onuchic JN. Domain swapping is a consequence of minimal frustration. Proc Natl Acad Sci USA 2004;101:13786–13791.
64. Cho SS, Levy Y, Onuchic JN, Wolynes PG. Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerization and aggregation. Phys Biol 2005;2:S44–S55.
65. Kim YH, Stites WE. Effects of excluded volume upon protein stability in covalently cross-linked proteins with variable linker lengths. Biochemistry 2008;47:8804–8814.
66. Tokuriki N, Kinjo M, Negi S, Hoshino M, Goto Y, Urabe I, Yomo T. Protein folding by the effects of macromolecular crowding. Protein Sci 2004;13:125–133.

67. van den Berg B, Ellis RJ, Dobson CM. Effects of macromolecular crowding on protein folding and aggregation. EMBO J 1999;18:6927–6933.
68. Wang W, Xu WX, Levy Y, Trizac E, Wolynes PG. Confinement effects on the kinetics and thermodynamics of protein dimerization. Proc Natl Acad Sci USA 2009;106:5517–5522.
69. Improta S, Krueger JK, Gautel M, Atkinson RA, Lefevre JF, Moulton S, Trewhella J, Pastore A. The assembly of immunoglobulin-like modules in titin: implications for muscle elasticity. J Mol Biol 1998;284:761–777.
70. Politou AS, Gautel M, Improta S, Vangelista L, Pastore A. The elastic I-band region of titin is assembled in a "modular" fashion by weakly interacting Ig-like domains. J Mol Biol 1996;255:604–616.
71. Sikora M, Cieplak M. Mechanical stability of multidomain proteins and novel mechanical clamps. Proteins 2011;79:1786–1799.
72. Osvath S, Kohler G, Zavodszky P, Fidy J. Asymmetric effect of domain interactions on the kinetics of folding in yeast phosphoglycerate kinase. Protein Sci 2005;14:1609–1616.