Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerization and aggregation

2005 Phys. Biol. 2 S44

(http://iopscience.iop.org/1478-3975/2/2/S05)

INSTITUTE OF PHYSICS PUBLISHING PHYSICAL BIOLOGY

Phys. Biol. **2** (2005) S44–S55 doi:10.1088/1478-3975/2/2/S05

Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerization and aggregation

Samuel S Cho[1,2], **Yaakov Levy**[1,3], **José N Onuchic**[1,3] **and Peter G Wolynes**[1,2,3]

[1] Center for Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[2] Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[3] Department of Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

E-mail: pwolynes@chem.ucsd.edu

Received 30 March 2005
Accepted for publication 24 May 2005
Published 17 June 2005
Online at stacks.iop.org/PhysBio/2/S44

Abstract

The prevalence of domain-swapping in nature is a manifestation of the principle of minimal frustration in that the interactions designed by evolution to stabilize the protein are also involved in this mode of binding. We previously demonstrated that the Symmetrized-Go potential accurately predicts the experimentally observed domain-swapped structure of Eps8 based solely on the structure of the monomer. There can be, however, multiple modes of domain-swapping, reflecting a higher level of frustration, which is a consequence of symmetry. The human prion and cyanovirin-N are too frustrated to form unique domain-swapped structures on the basis of the Symmetrized-Go potential. However, supplementing the completely symmetric model with intermolecular and intramolecular disulfide bonds in the prion and cyanovirin-N proteins, respectively, yielded unique domain-swapped structures with a remarkable similarity to the experimentally observed ones. These results suggest that the disulfide bonds may sometimes be critical in overcoming the intrinsic frustration of the symmetrized energy landscapes for domain-swapping. We also discuss the implications of intermolecular disulfide bonds in the formation of mammalian prion aggregates.

M This article features online multimedia enhancements

Abbreviations

CI2	chymotrypsin inhibitor-2
CSU	Contacts of Structural Units
CV-N	cyanovirin-N
Eps8	epidermal growth factor receptor pathway substrate 8 SH3 domain
PDB	Protein Data Bank
PrP	prion protein
PrP^{C}	normal cellular form of PrP
PrP^{Sc}	infectious aggregate 'scrapie' form of PrP
RMSD	root mean square deviation
SH3	src homology domain 3

1. Introduction

The energy landscape theory provides a fundamental basis for understanding protein folding. By having evolved minimally frustrated sequences, nature has solved the generally intractable problem of achieving robust and efficient folding and oligomerization. Optimizing native structure-seeking interactions while destabilizing non-native structure-seeking interactions results in a partially rugged funneled energy landscape [1]. With such an energy landscape, the dynamics of natural proteins is quite distinct from the kinetics of amino acid heteropolymers with random sequences.

The latter generally have frustrating conflicts of different interactions that lead to energetic traps. In the approximation that the energy landscape is perfectly funneled (i.e., no energetic frustration), protein topology becomes the remaining key determinant of the observed folding kinetic mechanism [2, 3]. Native topology-based (Go) models that correspond to perfectly funneled energy landscapes have been successful in predicting both the gross and, very often, the finer features of the folding mechanisms of numerous single-domain proteins [2, 4–6], as well as the binding mechanisms of many homooligomers [7, 8]. The structure of a domain-swapped dimer can often be predicted from the monomer topology alone by generalizing the use of the principle of minimal frustration to find the interactions in dimers from those seen in monomers [9].

Domain-swapping is an unconventional mechanism of oligomerization such that the structural element, or 'domain', of one chain is interchanged with a corresponding element of its partner, resulting in an intertwined homooligomer. The intramolecular interactions that would normally stabilize the monomer are thus 'recruited' in swapping processes. These now become intermolecular interactions and define the interface of the complex [10, 11]. Since the notion of domain-swapping was formally introduced by Eisenberg [10], about 70 proteins with domain-swapped oligomers have been characterized by x-ray crystallography and/or solution NMR. A domain-swapping event is characterized by slow kinetics with a high activation energy barrier arising from the many strong native interactions that must be rearranged. Early studies of domain-swapped proteins suggested that prolines in the hinge region connecting the swapped region with the main body of a domain-swapped oligomer could be important factors in determining whether proteins can domain-swap [12, 13]. Further analysis shows, however, that this hypothesis cannot be generalized to all proteins that bind via domain-swapping. Domain-swapped proteins are typically observed and isolated under high concentration and low pH conditions [14], but in a growing number of cases swapping has been observed under physiological conditions [15, 16]. This has fueled speculations about the role of domain-swapping *in vivo*. In particular, great interest has been generated by the proposal that domain-swapping is a crucial part of the mechanism for amyloid aggregation [17–19]. Two amyloidogenic proteins, the human prion [20] and the human cystatin C [21], have been observed as domain-swapped dimers, suggesting that these dimers may be the building blocks for fibers, at least in their nascent state [14, 21]. The underlying mechanism and the main determinant(s) of domain-swapping can be understood in the context of the energy landscape theory. These insights may yield valuable clues about the role of domain-swapping in oligomerization and aggregation *in vivo*.

At a first glance, the domain-swapping phenomenon presents a contradiction to the idea of a funneled energy landscape. Since domain-swapping involves homooligomers, a direct outcome of there being a funneled energy landscape for the monomer is that the very interactions that stabilize the monomeric conformation must compete with symmetrically similar ones that provide corresponding intermolecular interactions in the dimer. In other words, for any given residue i that is in native monomeric contact with residue j, there is, at first sight, nothing to prevent the same residue i from favorably interacting intermolecularly with residue j' in its partner, since the physico-chemical nature of the intramolecular interaction is the same as that of intermolecular interaction. In principle, there is, thus, no reason why one region of the protein should be favored more to swap than any other region. With many different and potentially conflicting possibilities for intermolecular interactions, a frustrated energy landscape generally results for the dimer. There would then be no preference for a single domain-swapped configuration. Even if a single domain-swapped configuration is somehow preferred at equilibrium, there would appear to be no guarantee that the most stable structure should correspond to that observed in nature, which might be the result of kinetic control. Therefore, a funneled energy landscape for monomeric folding generally would seem to preclude the possibility of a perfectly funneled energy landscape for oligomerization by domain-swapping. In this case, how do domain-swapping proteins, with the potential for a frustrated energy landscape for oligomerization, discriminate against alternatives to find their way to a unique swapped conformation?

The simplest model capturing the essential issues raised above is the Symmetrized-Go potential. This model symmetrizes the intrachain interactions to form the interchain interactions. Using this potential, we have recently shown, with only the monomer structure of Eps8 as input, that despite the possibilities of frustration, the monomer topology alone determines which mode of swapping dominates. The inherent energetic frustration present in the model is more than balanced by chain entropy effects to yield an inherent preference for the experimentally observed domain-swapped dimer [9]. In this present study, we explore this model further. We study what the model would predict for proteins for which there are no structural data indicating that they undergo specific domain-swapping. We also examine the model's predictions for known domain-swapping proteins that also possess intramolecular or intermolecular disulfide bonds. For the latter, we show that these strong interactions can play a special role. The Symmetrized-Go model when used to study two proteins with no evidence of domain-swapping, 434 repressor and chymotrypsin inhibitor-2 (CI2), yields a topologically frustrated landscape for domain-swapping. No unique swapped structure was observed in simulating these proteins. On the other hand, cyanovirin-N (CV-N) and the human prion (PrP) have been shown to be specific domain-swapped proteins. These proteins both have disulfide bonds, and we show that those disulfide interactions are critical in generating a unique domain-swapped configuration rather than mixtures of dimerized configurations from the uniformly symmetrized model. The necessity of forming specific intermolecular disulfide bonds to direct the domain-swapping of the human prion suggests that such interactions may be a key factor in their aggregation as well; a possibility that has been previously suggested by earlier investigators [22, 23].

2. Methods

2.1. Sequence analysis of the hinge regions of domain-swapping proteins

Do the sequences of hinge regions primarily determine domain-swapping? To analyze the amino acid residue prevalence in the hinge region of domain-swapping proteins, we constructed two libraries: one of domain-swapped protein structures and another of a non-redundant set of the PDB (i.e., no two proteins in the library have more than 25% sequence homology). The purpose of the latter library is to provide a baseline containing the amino acid residue prevalences found in nature. Using PDBSelect [24], a set of 1834 non-redundant proteins (188 388 total residues) each with less than 200 amino acid residues was chosen, and the amino acid frequencies were calculated. The same calculation was carried out for the hinge region of the library of domain-swapped proteins. We used the same definition of the hinge region of domain-swapping proteins as was introduced by Eisenberg [14]. In that definition, the hinge loop is defined to include those residues with a RMSD change in backbone ϕ and ψ dihedral angles of more than 20° in the domain-swapped oligomer when compared to the corresponding angles found in the monomeric configuration plus any residues in the same loop that connect secondary structure elements.

2.2. Symmetrized-Go potential

In a Go model, from which the Symmetrized-Go potential is derived, the native topology is used as input to study the mechanism of folding. A detailed definition of the kind of Go model we use [2, 3] is given in the supplementary material[4], but we briefly describe the main features here. Each residue is described as a single bead, centered on the Cα position. The beads in an intact protein chain are connected to adjacent beads by bond, angle and dihedral potentials. Simulation of the resulting simplified model of a protein allows the observation of slow conformational changes, and usually provides an accurate description of the intermediate and transition states of the folding mechanism observed experimentally. The network of favorable tertiary interactions is defined by the protein's native topology while all other non-bonded interactions are repulsive. In the Symmetrized-Go potential for a two-chain system, each individual protein chain is represented likewise using a series of single beads, each centered on the Cα position, and the native monomeric configuration is still used to define the intramolecular interactions. However, the Symmetrized-Go potential for the two-chain system also contains intermolecular interactions. In the symmetrized potential, the observed intramolecular interactions of the monomer also introduce the favorable possible intermolecular interactions. No other interactions are introduced. This model can be readily generalized to study aggregation. Indeed, this model had been previously used by Ding *et al* to study the aggregation of SH3 [25]. This protocol introduces intermolecular energetic frustration into the energy function, making it a predictor of not only the mechanism of domain-swapping but also allows it to make predictions of the domain-swapped structure (if unique) using only the monomer structure as input.

The energy function for the Symmetrized-Go potential for a two-chain system (designated chain A and chain B) with configuration Γ can be written explicitly:

$$H(\Gamma, \Gamma_0) = H_{\text{backbone}} + H_{\text{intrachain}} + H_{\text{interchain}}$$

$$H_{\text{backbone}} = \sum_{\text{bonds}} K_r (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi^{(n)} [1 - \cos(n(\phi - \phi_0))]$$

$$H_{\text{intrachain}} = \sum_{\text{chain A}}^{i<j-3} \left\{ \varepsilon_1(i, j) \left[5 \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - 6 \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{10} \right] + \varepsilon_1(i, j) \left(\frac{\sigma_0}{r_{i,j}} \right)^{12} \right\} + \sum_{\text{chain B}}^{i'<j'-3} \left\{ \varepsilon_1(i', j') \left[5 \left(\frac{\sigma_{i',j'}}{r_{i',j'}} \right)^{12} - 6 \left(\frac{\sigma_{i',j'}}{r_{i',j'}} \right)^{10} \right] + \varepsilon_1(i', j') \left(\frac{\sigma_0}{r_{i',j'}} \right)^{12} \right\}$$

$$H_{\text{interchain}} = \sum_{\text{A}\to\text{B}}^{i<j'-3} \left\{ \varepsilon_2(i, j') \left[5 \left(\frac{\sigma_{i,j'}}{r_{i,j'}} \right)^{12} - 6 \left(\frac{\sigma_{i,j'}}{r_{i,j'}} \right)^{10} \right] + \varepsilon_2(i, j') \left(\frac{\sigma_0}{r_{i,j'}} \right)^{12} \right\} + \sum_{\text{B}\to\text{A}}^{i'<j-3} \left\{ \varepsilon_2(i', j) \left[5 \left(\frac{\sigma_{i',j}}{r_{i',j}} \right)^{12} - 6 \left(\frac{\sigma_{i',j}}{r_{i',j}} \right)^{10} \right] + \varepsilon_2(i', j) \left(\frac{\sigma_0}{r_{i',j}} \right)^{12} \right\}.$$

The local backbone interactions are contained in H_{backbone}, which applies to both chains. K_r, K_θ and K_ϕ are the force constants of the bonds, angles and dihedral angles, respectively. The r, θ and ϕ variables are the bond lengths, the angles and the dihedral angles, respectively. The same quantities with a subscript zero represent the corresponding values taken only from the native monomer configuration, Γ_0. The non-bonded contact interactions, $H_{\text{intrachain}}$ and $H_{\text{interchain}}$, contain Lennard–Jones 10–12 terms for the non-local 'native' intrachain and interchain interactions and a short-range repulsive term for the 'non-native' pairs. Strictly speaking, the 'native' interchain interactions that result from symmetrizing the intrachain interactions include not only interactions that are present in the experimentally observed dimer (i.e., native), but also interchain interactions that are not present (i.e., non-native or frustrated), which is why this is

[4] See online at stacks.iop.org/PhysBio/2/S44.

a non-trivial model when it comes to predicting the domain-swapped structure.

We chose as parameters of the energy function $K_r = 100\varepsilon$, $K_\theta = 20\varepsilon$ and $\varepsilon_1 = \varepsilon_2 = \varepsilon$. Forming disulfide bonds is effectively irreversible. Such bonding interactions were incorporated into the energy function by setting $\varepsilon_1 = 10\varepsilon$ or $\varepsilon_2 = 10\varepsilon$ for intramolecular or intermolecular disulfide interactions, respectively. The secondary structure biases are set as $K_\phi{}^{(1)} = \varepsilon$ and $K_\phi{}^{(3)} = 0.5\varepsilon$ if the residue was either α-helical or β-sheet in character according to the DSSP definition [26] and $K_\phi{}^{(1)} = 0.25\varepsilon$ and $K_\phi{}^{(3)} = 0.12\varepsilon$ otherwise. Using higher flexibility for all of the turns in the proteins allows for changes in the dihedral angles of hinge regions without biasing any specific region of the protein to swap. $\sigma_{i,j}$ is the distance between the pair of residues (i, j) in the native monomeric configuration and $\sigma_0 = 4.0$ Å for all non-native residue pairs.

A total of N native contact pairs for the monomeric conformation was determined using the CSU (Contacts of Structural Units) software [27]. Only native contact pairs with sequence distance $|i - j| > 3$ were used because any three or four contiguous residues already interact through the angle and dihedral terms. The $2N$ intramolecular interactions (N native interactions for each monomer) also define the $2N$ intermolecular interactions as follows: for each i and j intramolecular interaction that is native in the monomeric conformation we also define equal intermolecular interactions between i and $j' = j$. In total, $4N$ interactions are thus represented in the model. Therefore, there exists an energetic competition as to whether the pair of molecules should make any given contact intra- or inter-molecularly. Interchain interactions between helical residues where the sequence distance $|i - j|$ equals 4 were ignored because helical contacts are not expected to be involved in swapping.

We performed constant temperature molecular dynamics simulations of the protein systems with the Symmetrized-Go potential. We imposed an interchain center of mass constraint $E_{\text{cons}} = K(R - R_0)^2$ that becomes effective only when $R > R_0$. The minimum of the constraint, R_0, was set to the radius of gyration of the monomer conformation.

2.3. Studied protein systems

In our applications of the Symmetrized-Go potential, we used the monomer structure to construct the model and the domain-swapped structure for comparing our results. Two proteins that have not been observed in domain-swapped form were studied. These were the 434 repressor (PDB code 1R69) and CI2 (PDB code 2CI2). The engineered domain-swapped structure of CI2 (PDB code 1CQ4) was used to compare to the structures in a highly populated domain-swapped basin. As for the domain-swapped proteins used in our study, the Symmetrized-Go potential constructed from monomer conformations of Eps8 (PDB code 1I0C), CV-N (PDB code 2EZM) and PrP (PDB code 1QLX), and the domain-swapped conformation (PDB codes 1I07, 3EZM and 1I4M, respectively) were used to compare to simulation results. Among the domain-swapped proteins that we studied with prolines in the hinge region is p13suc1 (PDB code 1SCE).

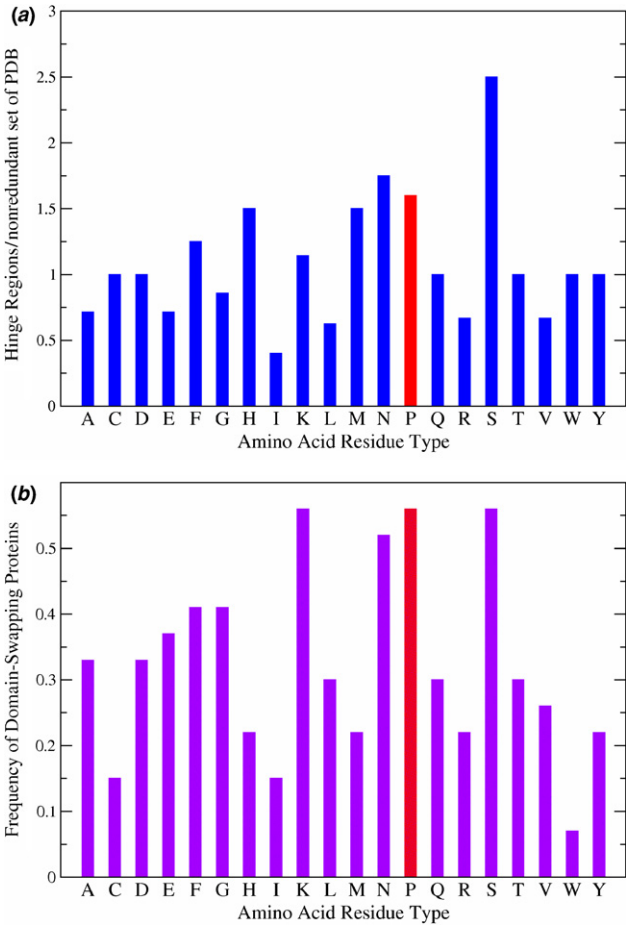

Figure 1. Evidence that prolines are not necessary as local signals to direct proteins to domain-swap. (*a*) A comparison between the distributions of amino acid residue frequency in the hinge region of domain-swapping proteins and a non-redundant set of the PDB. (*b*) The frequency of domain-swapping proteins with a certain residue in the hinge region.

3. Results and discussion

3.1. Bioinformatical survey of amino acid prevalence in the hinge region of domain-swapping proteins

It was observed, early on, from a survey of domain-swapping proteins conducted by Liu *et al* that there does not appear to be sequence homology between the swapping domains that is common to all domain-swapped proteins [14]. Further, the secondary structure cannot be a determining factor because the swapping region can range from a single α-helix or β-sheet to an entire tertiary domain [14]. One of the earliest and certainly most prominent hypotheses concerning the determinants of domain-swapping was that prolines play a pivotal role. This hypothesis was suggested largely because prolines seemed to be prevalent in the hinge regions of some of the first observed cases of domain-swapping proteins [12]. The apparent line of thought was the following: the cis–trans isomerization of prolines, which has a significantly lower energetic barrier than for other natural amino acids, is the rate-determining step in the folding rate of some proteins. Owing to this, prolines

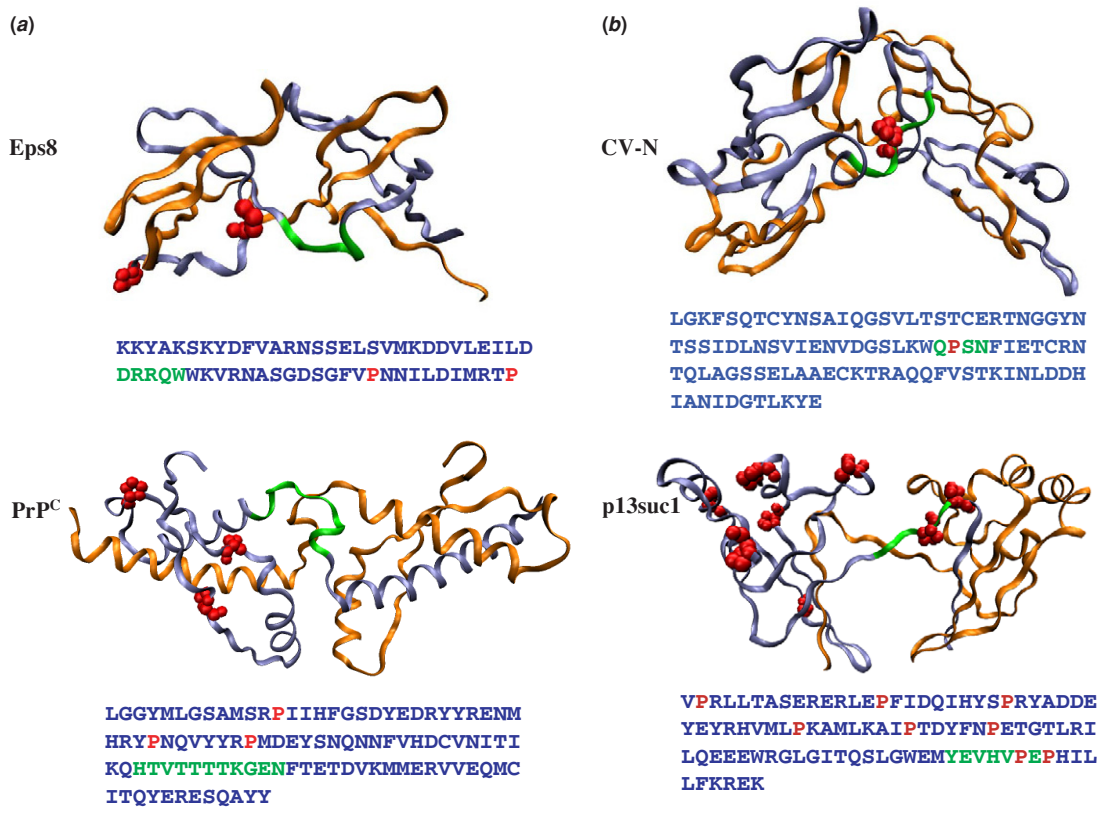

Figure 2. Examples of domain-swapping proteins and the proximity of the prolines to the hinge region. The structures of domain-swapping proteins without (*a*) Eps8 and PrP and with (*b*) CV-N and p13suc1 prolines in the hinge region are shown in a ribbon representation, with each monomer colored orange, or blue, and the hinge region colored green. The prolines found in the blue chain are shown in a red space-filled representation. The sequences of the proteins are shown below each structure, in which the prolines are colored red, the hinge region residues are colored green and the rest are colored blue.

often play a critical role in the observed folding kinetics, giving rise to many long-lived intermediates [28]. So, it was natural to suppose that prolines at or proximal to the hinge of domain-swapping proteins could act as local signals that would direct the global conformational change required to domain-swap with its identical partner. Experimental support for this hypothesis came in the form of a mutational study by Itzhaki *et al*, where it was found that two conserved prolines in the hinge region of p13suc1, a domain-swapping protein, controlled the monomer–dimer equilibrium in that system. Prolines made the hinge act like a 'loaded molecular spring' that shifts toward either the monomer or the domain-swapped conformation [13]. Itzhaki *et al* and others suggested that prolines more generally would be levers by which naturally monomeric proteins could be re-designed artificially to stabilize the domain-swapped state. This is no doubt true. One may go further, however, to posit that prolines in the hinge region are the main determinant of how proteins naturally oligomerize via domain-swapping.

To test this larger hypothesis on a broader basis in the naturally occurring proteins, we asked two questions: (1) is the prevalence of prolines in the hinge region of presently known domain-swapped proteins indeed significantly high? (2) Is the presence of prolines in or near the hinge region obligatory to oligomerize via domain-swapping? To answer these questions, we constructed appropriate libraries, as discussed earlier in the methods section. A comparison of the amino acid frequencies of the hinge regions of proteins with the frequencies in the library of the non-redundant proteins (figure 1(*a*)) shows that prolines are no longer prevalent in the hinges when compared to many other residues. We found that the frequency of prolines in domain-swapping proteins (figure 1(*b*)) is comparable to other kinds of residues. In fact, only about 50% of domain-swapping proteins have any prolines in their hinge region at all. In figure 2, we show two examples of domain-swapping proteins that do not contain prolines in their hinge regions (figure 2(*a*)) and two examples that have prolines in the hinge region (figure 2(*b*)). In both the examples in figure 2(*a*), the molecules do possess prolines that are absent from the hinge region, and indeed are distant from the hinge. For many domain-swapping proteins with prolines in the hinge region (figure 2(*b*)), as is the case of p13suc1, numerous prolines can also be found dispersed throughout the sequence, again even at positions very distant from the hinge region. There is, of course, no reason to challenge the contention that prolines significantly control the monomer–dimer equilibrium in p13suc1. It remains likely in our view that some proteins could be designed, by the addition of prolines, to favor domain-swapping. However, the examples explicitly show that the presence of prolines in the sequence

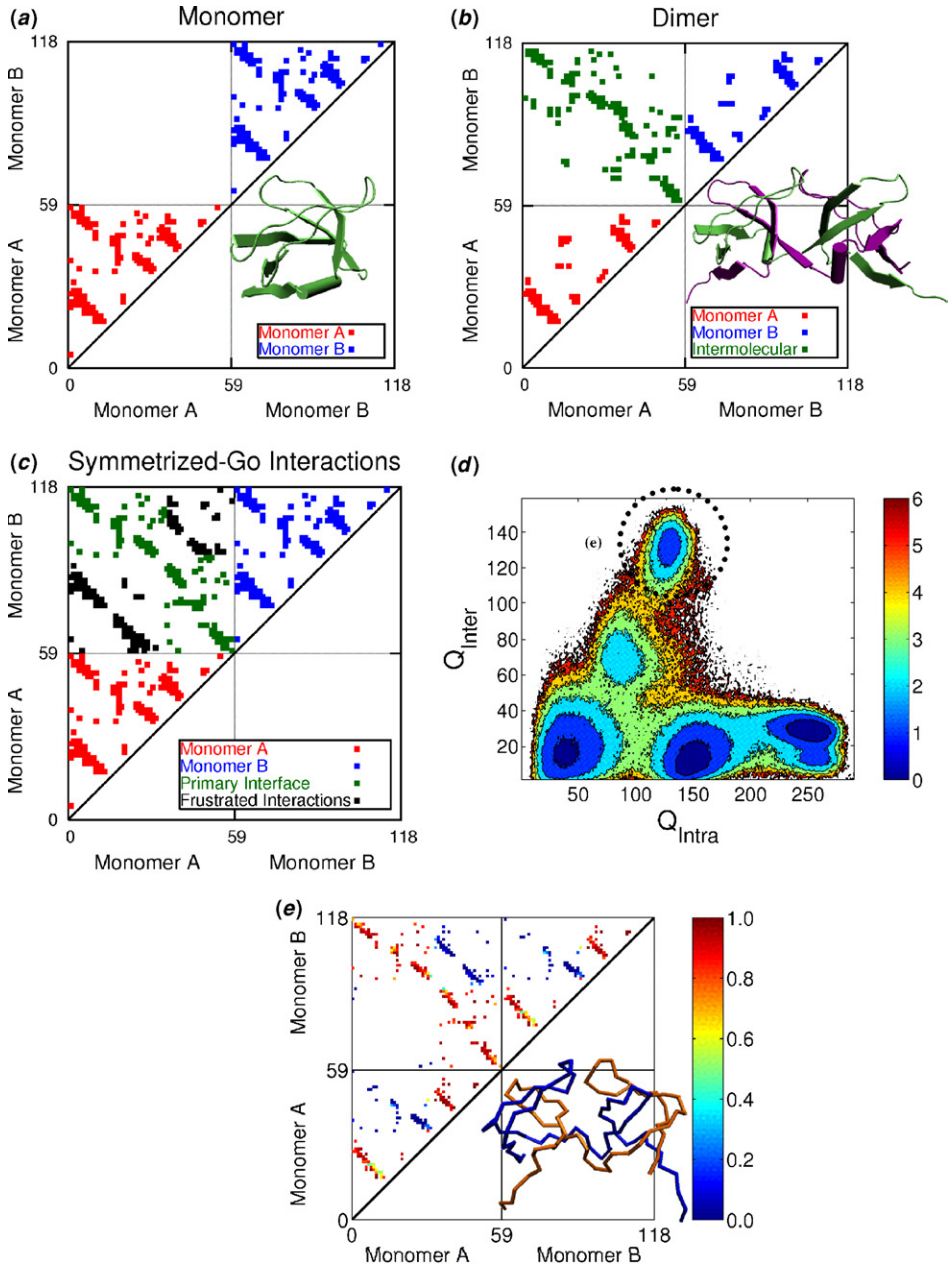

Figure 3. Application of the Symmetrized-Go potential to Eps8, a domain-swapping protein. The contact maps and the corresponding structures of the monomeric (*a*) and domain-swapped (*b*) Eps8 are shown. The represented favorable Symmetrized-Go interactions (*c*) include both the intramolecular and intermolecular interactions that have been derived from the monomeric conformation alone. The intermolecular interactions contained in the potential largely include the same interactions that are found in the experimentally observed dimer conformation (green), but there are also interactions that are not found in the experimentally observed dimer conformation (black). The free energy plot with respect to the number of intramolecular (Q_{Intra}) and intermolecular (Q_{Inter}) contacts (*d*) shows only a single stable domain-swapped conformation with an open-ended intermediate. The contact distribution plot of the minimum of the domain-swapped conformation (*e*) is shown as well as a representative structure from that minimum.

does not dictate whether a protein oligomerizes into a domain-swapped conformation.

To date, the primary strategy to engineer a protein to domain-swap has been to modify the hinge regions via mutations, additions or deletions. Two specific examples, however, also highlight the need to look outside of the hinge region. Mutagenic studies of BS-RNase demonstrated that Pro19, located in the hinge region, is not a significant factor in the domain-swapping mechanism. Instead, Leu28, which is located outside the hinge region, shifts the equilibrium toward the domain-swapped dimer by stabilizing the interface [29]. The sequences of two closely homologous proteins, the monomeric γB-Crystallin and the obligatory domain-swapped dimeric βB2-Crystallin, differ by the domain-swapped dimer having an acidic electrostatic repulsion between a residue in the hinge loop and a residue in the main body of the protein

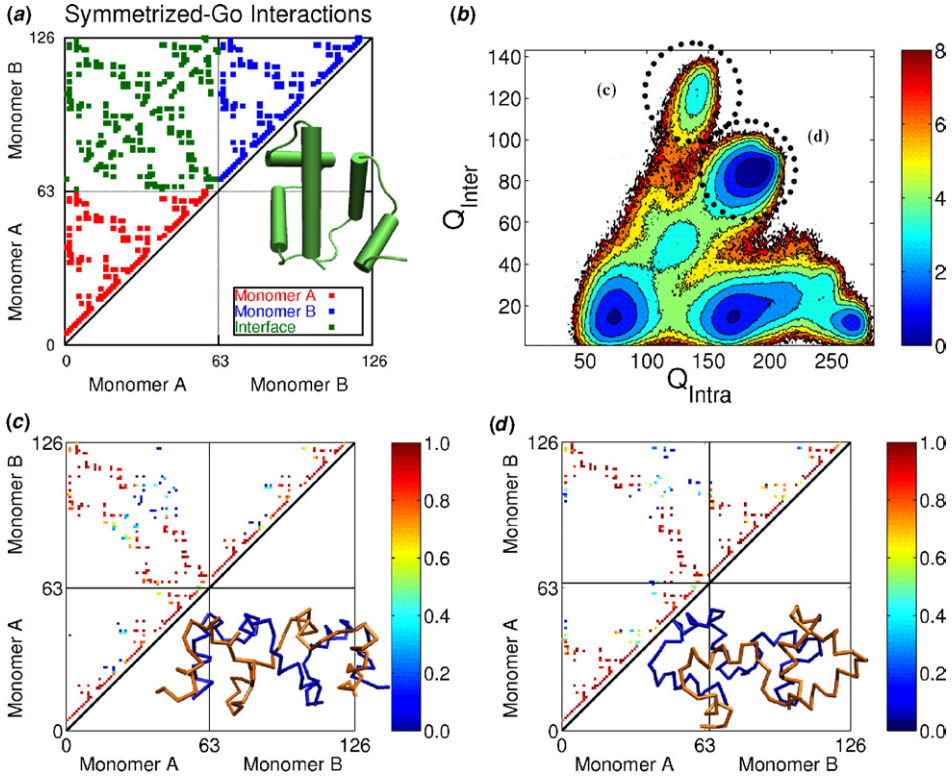

Figure 4. Application of the Symmetrized-Go potential to the 434 repressor, a dimeric protein showing no evidence of unique domain-swapping. The represented favorable Symmetrized-Go interactions (*a*) for the 434 repressor are shown with the corresponding structure of the monomer. The free-energy plot as a function of the number of intramolecular (Q_{Intra}) and intermolecular (Q_{Inter}) contacts (*b*) that was derived from our simulations shows two domain-swapped minima. The corresponding contact distribution plots of the two minima from (*b*) are shown in (*c*) and (*d*) as well as a representative structure from its respective minimum.

that prevents the formation of the monomer species [30]. Clearly, the network of interactions as a whole, not just those in the hinge region, must be considered in describing domain-swapping.

3.2. Unique and stable domain-swapped configuration

The first clue to direct the search for a unifying view of the domain-swapping mechanism is the somewhat tautological observation that the conformation of the swapped subunits in a domain-swapped oligomer bears a striking resemblance to the unswapped monomeric conformation (figures 3(*a*), (*b*)). Did evolution not only encode into the sequence information to fold a protein into its monomeric conformation but also instructions about whether it would oligomerize into a specific domain-swapped conformation? To ask whether the monomeric topology is sufficient for predicting how proteins oligomerize via the domain-swapping mechanism, we previously developed the Symmetrized-Go potential, as described in Yang *et al* [9] and the methods section. This model's formulation contains no information *a priori* that biases a specific swapping region. Also, the model contains no information concerning the secondary interface, i.e., there are no interactions corresponding to those new ones that would be formed upon domain-swapping that are not represented in the monomer conformation. The latter could potentially play a role in swapping. In principle, in the symmetrized model any region of the protein can exchange interactions with its partner and nothing would preclude even the possibility of there being multiple swapping regions. Does this energy function discriminate the experimentally observed domain-swapped structure from the energetic traps? If so, we can say that there already exists, encoded in the monomer topology, sufficient information to intrinsically choose the swapping region.

When we applied the Symmetrized-Go potential to Eps8 (epidermal growth factor receptor pathway substrate 8 SH3 domain), a domain-swapping protein, we found that despite the energetically frustrated intermolecular interactions, the model led to accurate prediction of the experimentally observed domain-swapped dimer as the most stable conformation. From our simulations, we can plot a free-energy surface as a function of the order parameters Q_{Intra} and Q_{Inter}, the number of native intramolecular and intermolecular contacts, respectively (figure 3(*d*)). Q_{Intra} indicates the degree of folding of the two monomers and Q_{Inter} indicates the degree of binding via swapping. At low Q_{Inter}, we found three basins, corresponding to two unfolded monomers, one unfolded monomer and one folded monomer, and two folded monomers. The basin with the highest Q_{Inter} corresponds to the fully swapped structure found via x-ray crystallography. At intermediate Q_{Inter}, there is a basin corresponding to one swapped and one unswapped conformation (i.e., partially domain-swapped intermediate). A contact probability plot of

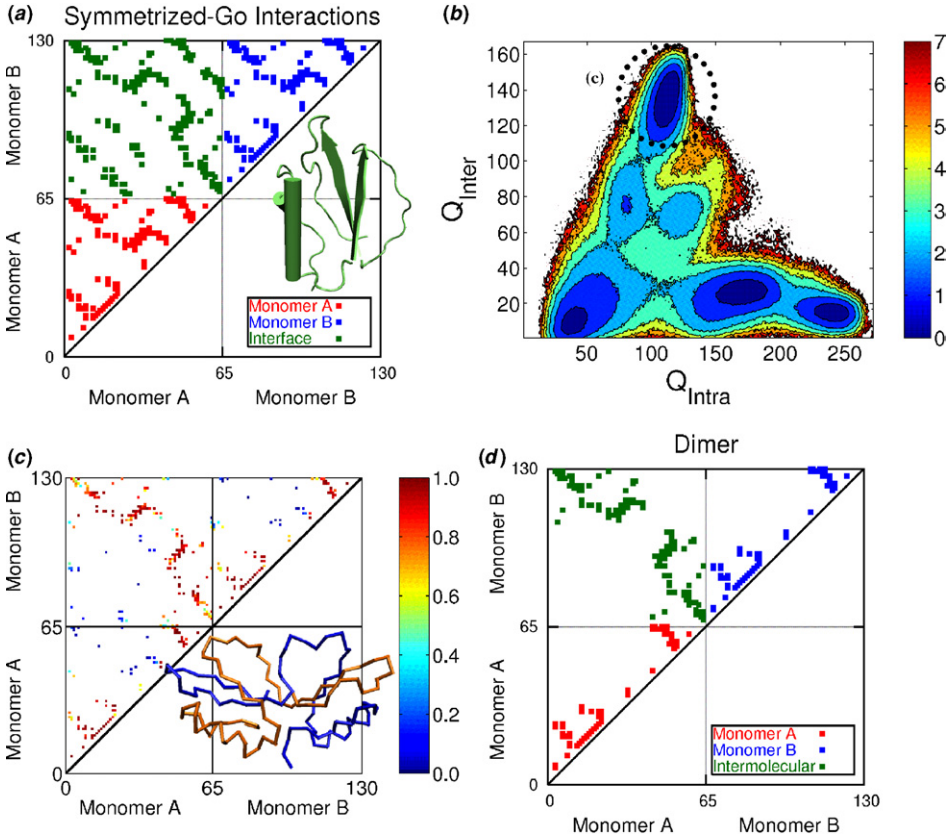

Figure 5. Application of the Symmetrized-Go potential to CI2, a naturally monomeric protein that has been artificially engineered to domain-swap via insertion of glutamine repeats. The represented favorable Symmetrized-Go interactions (*a*) for CI2 are shown with the corresponding structure of the monomer. The free-energy plot with respect to the number of intramolecular (Q_{Intra}) and intermolecular (Q_{Inter}) contacts (*b*) shows more than one domain-swapped minimum. The corresponding contact distribution plot of the deepest minimum from (*b*) is shown in (*c*) as well as a representative structure from its minimum. For comparison, the contact map depicting the swapping and main regions of the engineered domain-swapped of CI2 is shown in (*d*).

the basin of the domain-swapped conformation (figure 3(*e*)) shows that only the interactions found in the experimentally observed domain-swapped dimer are statistically favored. The other interactions, while favorable according to the symmetrized model, are either seldom represented or not found at all. Despite the energetic frustration that is present in the model, only the experimentally observed domain-swapped structure is found to be significantly populated.

3.3. Proteins that generally cannot stabilize into a unique domain-swapped configuration

We applied the Symmetrized-Go model to the 434 repressor, a well-studied dimeric protein for which no evidence of a unique domain-swapped form has been found to date. Just as with Eps8, we constructed a Symmetrized-Go potential from the conformation of a single monomer (figure 4(*a*)). The free-energy surface for the 434 repressor (figure 4(*b*)) shows two domain-swapped basins, reflecting a frustrated competition between the two states. This clearly contrasts with the free energy plot for Eps8, which has only one domain-swapped basin. A contact probability plot of the two basins yields two distinct domain-swapped structures (figures 4(*c*), (*d*)). One may note that these two swapped structures of the 434 repressor have a very similar number of contacts but differ in the degree of folding of the monomer and the interface size.

We further applied the Symmetrized-Go potential to CI2 (figure 5(*a*)), a protein that is found naturally as a monomer. While the wild-type protein is currently thought to be intrinsically monomeric, Perutz and colleagues have engineered a domain-swapped dimer by the insertion of glutamate repeats in a loop within the protein [31]. Similar to our study of the 434 repressor, we observed multiple minima of swapped structures when Q_{Inter} is high (figure 5(*b*)). Interestingly, the most stable of the minima had the highest number of intermolecular native contacts, and the ensemble of structures for this minimum (figure 5(*c*)) is similar to that structure found for the engineered domain-swapped protein (figure 5(*d*)). These observations indicate that further analysis of other naturally monomeric proteins using the Symmetrized-Go potential can predict which proteins might be most amenable to re-engineering into domain-swapping oligomers by appropriate hinge mutations. We note that the observation of non-specific domain-swapping of monomeric proteins is not simply an artifact of our model. Oliveberg observed 'transient aggregates' at high concentrations that cause deviations from two-state kinetics in protein folding [32], and we believe that they are the result of the unstable

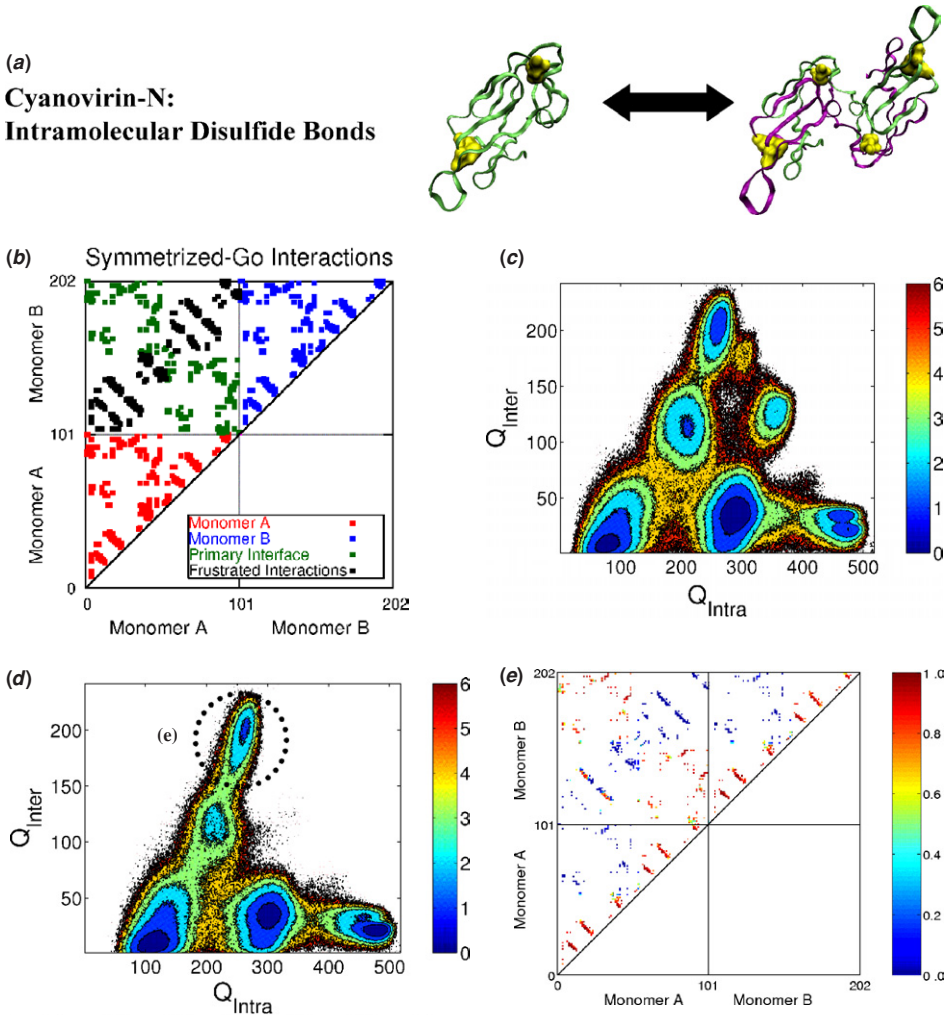

Figure 6. Application of the Symmetrized-Go potential to CV-N, a domain-swapping dimer with intramolecular disulfide bonds. The structures of the monomeric and domain-swapped conformations are shown (*a*) in a ribbon representation. The chains are colored green or purple, and the cysteine residues are shown colored yellow in a space-filled representation. The favorable Symmetrized-Go interactions of the domain-swapped dimers are shown (*b*). The free-energy plots as a function of the number of intramolecular (Q_{Intra}) and intermolecular (Q_{Inter}) contacts are shown, both without (*c*) and with (*d*) the explicit inclusion of disulfide bond interactions, along with a contact distribution plot of the domain-swapped basin (*e*).

domain-swapping we see in the Symmetrized-Go model. With multiple possibilities for domain-swapping, the protein is observed only in the monomeric conformation because of its higher specific concentration.

3.4. Disulfide bonds are essential for domain-swapping in CV-N and human prion

We now turn our attention to two other proteins with known domain-swapped structures: CV-N (figure 6(*a*)) and the human prion (PrP) (figure 7(*a*)). These have intramolecular and intermolecular disulfide bonds, respectively. CV-N has two intramolecular disulfide bonds: Cys 8–Cys 22 and Cys 58–73. The intramolecular disulfide bonds of CV-N are important for the stabilizing of the monomeric structure of CV-N. They are also critical to the anti-HIV activity of CV-N [33, 34]. The domain-swapped structure of CV-N has been resolved by both x-ray crystallography [33] and solution NMR [35]. The introduction of mutations to CV-N changed the energy landscape for folding to stabilize an intermediate [36]. Our Go-model simulations of wild-type CV-N as a monomer (see the supplementary material (footnote 4)) also revealed the existence of a high-energy intermediate. We had initially thought that this result indicated an actual intermediate that was, however, not able to be observed by current experimental techniques in the wild-type but was stabilized by incorporating mutations. However, when we introduced disulfide bonds into the topology of the Go-model, the high-energy intermediate was no longer in the free-energy profile. Retaining the disulfides changes the mechanism of folding.

How does the inclusion of disulfide bonds affect the energy landscape for domain-swapping? In the domain-swapped dimer conformation of CV-N, the disulfide bonds remain oxidized, so the conformational conversion does not require a reduction of the disulfide bonds. In Symmetrized-Go simulations of CV-N (figure 6(*b*)) without modifying

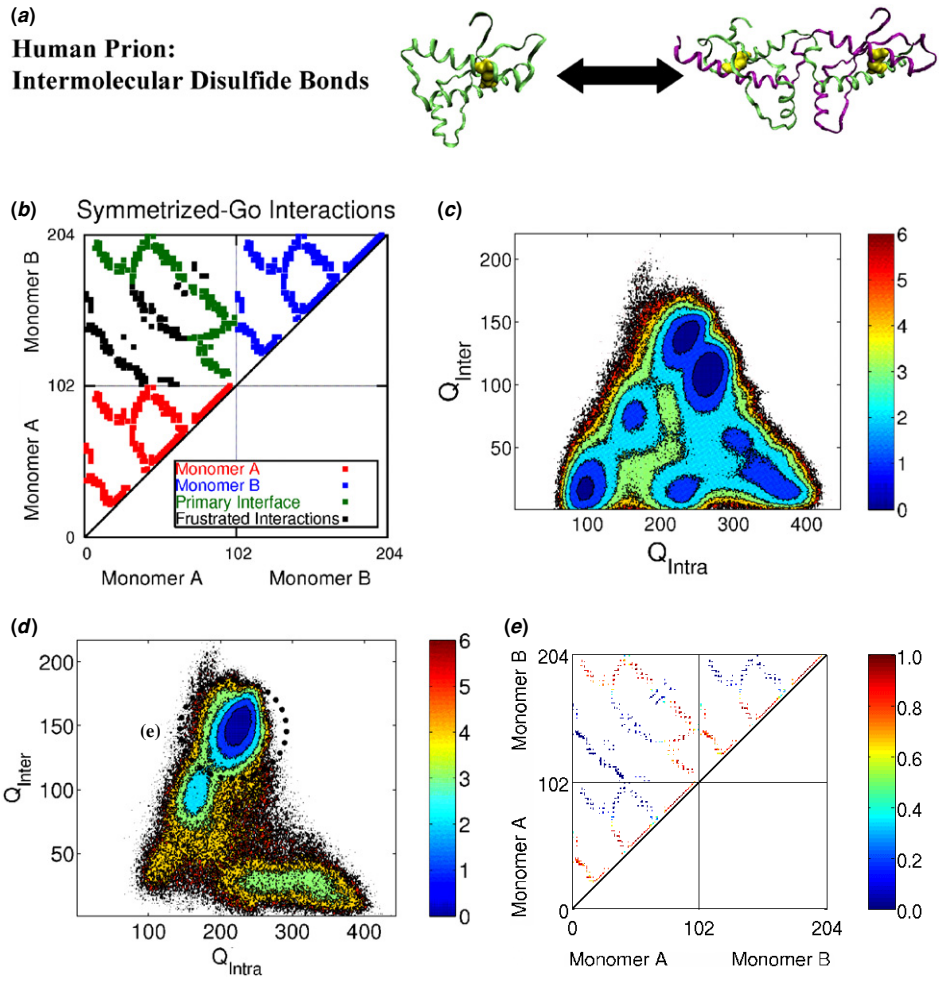

Figure 7. Application of the Symmetrized-Go potential to PrP, a domain-swapping dimer containing intermolecular disulfide bonds. The structures of the monomeric and domain-swapped conformations are shown (*a*) in a ribbon representation. The chains are colored green or purple, and the cysteine residues are shown colored yellow in a space-filled representation. The favorable Symmetrized-Go interactions of the domain-swapped dimers are shown (*b*). The free-energy plots as a function of the number of intramolecular (Q_{Intra}) and intermolecular (Q_{Inter}) contacts are shown, both without (*c*) and with (*d*) the explicit inclusion of disulfide bond interactions, along with a contact distribution plot of the domain-swapped basin (*e*).

the energetics of disulfide bonds to reflect their greater stability, the energy landscape for domain-swapping is clearly frustrated (figure 6(*c*)). However, once we included the stronger intramolecular disulfide bonds into the topology of CV-N, we found that the energy landscape for domain-swapping becomes effectively unfrustrated (figure 6(*d*)). A contact probability plot of the basin of the domain-swapped conformation (figure 6(*e*)) shows that only those interactions found in the experimentally observed domain-swapped dimer are now favored, just as we saw in the case of Eps8. The disulfide bonds not only act to stabilize the monomer conformation, but also they limit the possible states that are accessible for domain-swapping. With the permanent disulfide linkage, only one stable state becomes possible for the dimer.

Despite much progress and study, the detailed mechanism for the conversion of prions (PrP) from the normal cellular form (PrP^{C}) to the infectious aggregate form (PrP^{Sc}) remains elusive. The structures of PrP^{C} for several mammal proteins have been determined by solution NMR, and they all have the same basic monomeric structure, consisting of three long α-helices and two short β-sheet strands with a conserved disulfide bond between Cys 179 and Cys 214 that bridge helices 2 and 3. A domain-swapped dimer conformation of PrP was found experimentally in which there are intermolecular disulfide bonds between Cys 179 in one monomer and Cys 214 of its partner, bridging the helix 2 of one monomer with helix 3 in its partner [20]. In contrast to the case of CV-N, the conformational change of the PrP from the monomeric to the domain-swapped dimer forms must involve the reduction of the intramolecular disulfide bond and subsequent intermolecular reoxidation. The Symmetrized-Go simulations of the PrP (figure 7(*a*)) carried out without consideration of the disulfide bonds again revealed multiple possibilities for domain-swapping (figure 7(*b*)). It is only upon including effectively irreversible intermolecular disulfide bonds that the energy landscape for domain-swapping becomes topologically funneled (figure 7(*c*)) toward the experimentally observed domain-swapped state (figures 7(*d*), (*e*)). The pivotal role of intermolecular disulfide bonds in prion aggregation has been suggested both theoretically [22] and experimentally [23],

but there is some disagreement as to whether intermolecular disulfide bonds actually do occur in the large prion aggregate [37, 38]. While further study is clearly needed for a definitive answer, our present study would provide a structural basis for obligate intermolecular disulfide interactions in prion aggregation. If forming intermolecular disulfide bonds is critical for domain-swapping, these interactions may at least be transiently represented at the early stages of prion aggregation. The increase in local concentration of prion proteins caused first by domain-swapping may trigger the further conformational changes required to form PrP^{Sc}. We note that this hypothesis does not conflict with the current understanding of the structure of the PrP^{Sc} fiber [39] in which helices 2 and 3 of PrP^{Sc} and the disulfide bond between them remain intact. It has not escaped notice that transient disulfide oxidation isomerization and reduction, perhaps in different physiological compartments or conditions, would greatly modify the kinetics of aggregate formation and fragmentation from the predictions of simpler kinetic assembly models, which currently seem unable to account fully for the quantitative details of *in vivo* pathogenesis [40–43].

4. Conclusion and outlook

Owing to symmetry, the domain-swapping mechanism may lead to an energetically frustrated landscape for dimerization, potentially riddled with conflicting energetic contributions that could lead to traps. Those proteins that have multiple possibilities for domain-swapping only transiently aggregate but do not generally form a unique and stable domain-swapped structure which can be isolated. Domain-swapping proteins, on the other hand, appear to avoid the energetic frustration that is potentially present by having ways to narrow the selection of routes to unique stable domain-swapped conformations. In some cases, as illustrated by Eps8, special interactions are unnecessary; the topology of the fold already leads to the dominant unique and stable domain-swapped configuration. Unique domain-swapped structures can also form by making use of the strong, irreversibly formed disulfide bonds to overcome the energetically frustrated features of the dimer energy landscape. In the case of CV-N, we found that maintaining the intramolecular disulfide bonds restricts the conformational space accessible for domain-swapping such that only one state is possible. On the other hand, for the human prion, reducing intramolecular disulfide bonds and then re-oxidizing to form intermolecular disulfide bonds can lead to a unique dimeric state. The control of the oxidation and reduction of disulfide bonds can in turn control the balance between the monomeric and domain-swapped states. For mammalian prions, disulfide bonds may be a crucial kinetic factor in aggregation.

Acknowledgments

This work was funded by the National Science Foundation-sponsored Center for Theoretical Biological Physics (grants PHY-0216576 and 0225630) with additional support from NSF MCB-0084797. Computations were carried out at the University of California at San Diego Keck II computing facility (partially supported by the National Science Foundation, Division of Molecular and Cellular Biosciences). SSC is supported by a University of California at San Diego Molecular Biophysics Training Grant.

Glossary

Frustration. Conflicting interactions arising from competition between two or more states that minimize a local part of the free energy.

Principle of minimal frustration. Natural proteins are evolutionarily designed to have sequences that optimize native structure-seeking interactions and generally destabilize non-native structure-seeking interactions, resulting in a native state that is dominant and unique. These sequences are said to be minimally frustrated, resulting in a funneled energy landscape. In contrast, random sequences are typically highly frustrated, yielding a rough energy landscape with multiple trap states.

Native interactions. Interactions that are found in the folded state, typically defined from the x-ray crystallographic or NMR structure.

Non-native interactions. Possible interactions that are absent in the folded state.

Native topology-based model. A solvent-averaged energy potential that corresponds to a perfectly funneled energy landscape. It is mainly defined by attractive native state interactions and repulsive non-native interactions such that the lowest energy minimum is the native state. This type of model is often called a Go model.

Symmetrized-Go model. A solvent-averaged energy potential for describing the domain-swapping phenomenon. The intramolecular interactions are perfectly funneled to the monomeric state. Those interactions are symmetrized to become intermolecular interactions with its partner(s), resulting in a non-trivial predictive model of the domain-swapped structure, using only the monomeric state as input.

Domain-swapping. A mode of oligomerization in which a structural element, or a 'domain', of one chain is interchanged with a corresponding element of its partner, resulting in an intertwined homooligomer with at least one axis of symmetry.

Prion. Infectious apparently self-replicating protein particles thought to be the agent responsible for transmissible spongiform encephalopathies, Creutzfeldt–Jakob disease, and other neurological degenerative diseases.

References

[1] Bryngelson J D and Wolynes P G 1987 Spin glasses and the statistical mechanics of protein folding *Proc. Natl Acad. Sci. USA* **84** 7524–8

[2] Clementi C, Nymeyer H and Onuchic J N 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** 937–53

[3] Onuchic J N, Socci N D, Luthey-Schulten Z and Wolynes P G 1996 Protein folding funnels: the nature of the transition state ensemble *Fold Des.* **1** 441–50

[4] Koga N and Takada S 2001 Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model *J. Mol. Biol.* **313** 171–80

[5] Chavez L L, Onuchic J N and Clementi C 2004 Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates *J. Am. Chem. Soc.* **126** 8426–32

[6] Clementi C, Jennings P A and Onuchic J N 2000 How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta *Proc. Natl Acad. Sci. USA* **97** 5871–6

[7] Levy Y, Wolynes P G and Onuchic J N 2004 Protein topology determines binding mechanism *Proc. Natl Acad. Sci. USA* **101** 511–6

[8] Levy Y, Cho S S, Onuchic J N and Wolynes P G 2005 A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes *J. Mol. Biol.* **346** 1121–45

[9] Yang S *et al* 2004 Domain swapping is a consequence of minimal frustration *Proc. Natl Acad. Sci. USA* **101** 13786–91

[10] Bennett M J, Choe S and Eisenberg D 1994 Domain swapping: entangling alliances between proteins *Proc. Natl Acad. Sci. USA* **91** 3127–31

[11] Bennett M J, Schlunegger M P and Eisenberg D 1995 3D domain swapping: a mechanism for oligomer assembly *Protein Sci.* **4** 2455–68

[12] Bergdoll M, Eltis L D, Cameron A D, Dumas P and Bolin J T 1998 All in the family: structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly *Protein Sci.* **7** 1661–70

[13] Rousseau F, Schymkowitz J W, Wilkinson H R and Itzhaki L S 2001 Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues *Proc. Natl Acad. Sci. USA* **98** 5596–601

[14] Liu Y and Eisenberg D 2002 3D domain swapping: as domains continue to swap *Protein Sci.* **11** 1285–99

[15] Park C and Raines R T 2000 Dimer formation by a 'monomeric' protein *Protein Sci.* **9** 2026–33

[16] Botos I, Mori T, Cartner L K, Boyd M R and Wlodawer A 2002 Domain-swapped structure of a mutant of cyanovirin-N *Biochem. Biophys. Res. Commun.* **294** 184–90

[17] Liu Y, Gotte G, Libonati M and Eisenberg D 2001 A domain-swapped RNase A dimer with implications for amyloid formation *Nat Struct. Biol.* **8** 211–4

[18] Schlunegger M P, Bennett M J and Eisenberg D 1997 Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly *Adv. Protein Chem.* **50** 61–122

[19] Cohen F E and Prusiner S B 1998 Pathologic conformations of prion proteins *Annu. Rev. Biochem.* **67** 793–819

[20] Knaus K J, Morillas M, Swietnicki W, Malone M, Surewicz W K and Yee V C 2001 Crystal structure of the human prion protein reveals a mechanism for oligomerization *Nat. Struct. Biol.* **8** 770–4

[21] Janowski R *et al* 2001 Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping *Nat. Struct. Biol.* **8** 316–20

[22] Welker E, Wedemeyer W J and Scheraga H A 2001 A role for intermolecular disulfide bonds in prion diseases? *Proc. Natl Acad. Sci. USA* **98** 4334–6

[23] Lee S and Eisenberg D 2003 Seeded conversion of recombinant prion protein to a disulfide-bonded oligomer by a reduction-oxidation process *Nat. Struct. Biol.* **10** 725–30

[24] Hobohm U and Sander C 1994 Enlarged representative set of protein structures *Protein Sci.* **3** 522–4

[25] Ding F, Dokholyan N V, Buldyrev S V, Stanley H E and Shakhnovich E I 2002 Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism *J. Mol. Biol.* **324** 851–7

[26] Kabsch W and Sander C 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features *Biopolymers* **22** 2577–637

[27] Sobolev V, Sorokine A, Prilusky J, Abola E E and Edelman M 1999 Automated analysis of interatomic contacts in proteins *Bioinformatics* **15** 327–32

[28] Wedemeyer W J, Welker E and Scheraga H A 2002 Proline cis-trans isomerization and protein folding *Biochemistry* **41** 14637–44

[29] Picone D *et al* 2005 The role of the hinge loop in domain-swapping: the special case of Bovine seminal Ribonuclease *J. Biol. Chem.* **280** 13771–8

[30] Lapatto R *et al* 1991 High resolution structure of an oligomeric eye lens beta-crystallin. Loops, arches, linkers and interfaces in beta B2 dimer compared to a monomeric gamma-crystallin *J. Mol. Biol.* **222** 1067–83

[31] Chen Y W, Stott K and Perutz M F 1999 Crystal structure of a dimeric chymotrypsin inhibitor 2 mutant containing an inserted glutamine repeat *Proc. Natl Acad. Sci. USA* **96** 1257–61

[32] Oliveberg M 1998 Alternative explanations for 'multistate' kinetics in protein folding: transient aggregation and changing transition-state ensembles *Acc. Chem. Res.* **31** 765–72

[33] Yang F *et al* 1999 Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping *J. Mol. Biol.* **288** 403–12

[34] Mori T *et al* 1997 Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein *Biochem. Biophys. Res. Commun.* **238** 218–22

[35] Barrientos L G *et al* 2002 The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures *Structure (Camb.)* **10** 673–86

[36] Barrientos L G, Lasala F, Delgado R, Sanchez A and Gronenborn A M 2004 Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity *Structure (Camb.)* **12** 1799–807

[37] Welker E, Raymond L D, Scheraga H A and Caughey B 2002 Intramolecular versus intermolecular disulfide bonds in prion proteins *J. Biol. Chem.* **277** 33477–81

[38] May B C, Govaerts C, Prusiner S B and Cohen F E 2004 Prions: so many fibers, so little infectivity *Trends Biochem. Sci.* **29** 162–5

[39] Govaerts C, Wille H, Prusiner S B and Cohen F E 2004 Evidence for assembly of prions with left-handed beta-helices into trimers *Proc. Natl Acad. Sci. USA* **101** 8342–7

[40] Eigen M 1996 Prionics or the kinetic basis of prion diseases *Biophys. Chem.* **63** A1–18

[41] Masel J, Jansen V A and Nowak M A 1999 Quantifying the kinetic parameters of prion replication *Biophys. Chem.* **77** 139–52

[42] Feughelman M and Willis B K 2000 Thiol-disulfide interchange a potential key to conformational change associated with amyloid fibril formation *J. Theor. Biol.* **206** 313–5

[43] Tompa P, Tusnady G E, Friedrich P and Simon I 2002 The role of dimerization in prion replication *Biophys. J.* **82** 1711–8