
The dual role of a loop with low loop contact distance in folding and domain swapping

APICHART LINHANANTA,¹ HONGYI ZHOU,¹ AND YAOQI ZHOU

Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology and Biophysics, State University of New York at Buffalo, Buffalo, New York 14214, USA

(RECEIVED February 19, 2002; FINAL REVISION April 10, 2002; ACCEPTED April 10, 2002)

Abstract

α helices, β strands, and loops are the basic building blocks of protein structure. The folding kinetics of α helices and β strands have been investigated extensively. However, little is known about the formation of loops. Experimental studies show that for some proteins, the formation of a single loop is the rate-determining step for folding, whereas for others, a loop (or turn) can misfold to serve as the hinge loop region for domain-swapped species. Computer simulations of an all-atom model of fragment B of *Staphylococcal* protein A found that the formation of a single loop initiates the dominant folding pathway. On the other hand, the stability analysis of intermediates suggests that the same loop is a likely candidate to serve as a hinge loop for domain swapping. To interpret the simulation result, we developed a simple structural parameter: the loop contact distance (LCD), or the sequence distance of contacting residues between a loop and the rest of the protein. The parameter is applied to a number of other proteins, including SH3 domains and prion protein. The results suggest that a locally interacting loop (low LCD) can either promote folding or serve as the hinge region for domain swapping. Thus, there is an intimate connection between folding and domain swapping, a possible cause of misfolding and aggregation.

Keywords: Total contact distance; loop contact distance; protein folding; domain-swapping; loop formation; fragment B of protein A

This study focuses on the role of the loop regions in the folding kinetics of a model of fragment B of *Staphylococcal aureus* protein A (BpA), a three-helix bundle protein (H1, H2, and H3) connected by two loops (L1 and L2). The free-energy landscape of this protein has been obtained by an all-atom model that employs the CHARMM empirical force field in an explicit solvent (Boczko and Brooks 1995; Guo et al. 1997). BpA was also studied by high-temperature all-atom unfolding simulations in explicit solvent (Alonso and Daggett 2000), as well as by folding simulations using simple reduced models (Kolinski et al. 1998; Shea et al.

1999, 2000; Zhou and Karplus 1999; Berriz and Shakhnovich 2001; Favrin et al. 2002). Some studies indicated that loop formation is the rate-determining step (Shea et al. 1999; Berriz and Shakhnovich 2001), whereas others highlighted the importance of the stability of H3 (Guo et al. 1997; Alonso and Daggett 2000). Still others found on-pathway intermediates consisting of a two-helix microdomain (either H1-H2 [Boczko and Brooks 1995; Kolinski et al. 1998] or both H1-H2 and H2-H3 [Zhou and Karplus 1999]). Experimental studies have revealed some important information on the folding of BpA (Bottomley et al. 1994; Bai et al. 1997; Myers and Oas 2001) but were not detailed enough to provide the specifics of the folding mechanism.

All-atom folding simulations using empirical force fields are not yet possible using available computing power. We overcome this limitation by using a structure-based, all-atom (except nonpolar hydrogen atoms) model in which the atoms interact by discontinuous G \ddot{o} potentials (Zhou and Linhananta 2002a,b). The goal is to determine the folding

Reprint requests to: Dr. Yaoqi Zhou, Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology and Biophysics, State University of New York at Buffalo, 124 Sherman Hall, Buffalo, NY 14214, USA; e-mail: yqzhou@buffalo.edu; fax: (716) 829-2344.

¹These authors contributed equally to this work.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0205002>.

kinetics from the native structure. Thermodynamic analysis of the model (Zhou and Linhananta 2002a) found that the inclusion of side chains eliminates the molten-globule-like state often encountered in the C_α -based model (Zhou and Karplus 1997; Pande and Rokhsar 1998). The same all-atom model has also yielded a collapse-initiated folding mechanism for the second β -hairpin fragment of the Ig-binding domain B of *Streptococcal* protein G and revealed the essential role of both hydrophobic and hydrophilic residues. The results are highly consistent with available experimental data, as well as with other all-atom unfolding and equilibrium simulation studies of the same β -hairpin (Zhou and Linhananta 2002b).

The new model allows us to fold the 46-residue, 459-atom BpA. At a reduced temperature (T^*) of 2.5, 80 out of 197 (41%) independent trajectories folded to the native state in ~ 16 μ sec or ~ 50 h each on a 1-GHz pentium PC. Folding kinetic results suggest two folding pathways mediated by loop formations. In the dominant, fast-folding pathway, L2 forms first, resulting in an H2-H3 intermediate that rapidly folds to the native state. In the slower pathway, L1 initiates the formation of an H1-H2 intermediate, leaving L2 unformed. The long lifetime of H1-H2 suggests that it is a possible template for a domain-swapped dimer in which L2 is the hinge loop. These remarkable properties are attributed to the fact that L2 interacts weakly with the rest of the protein. To generalize, we quantify the topological connectivity between a loop and the rest of protein by a parameter called loop contact distance (LCD). Applications of LCD reveal that the different roles of a weakly interacting loop (low LCD) in the folding of the model BpA are also observed in many other proteins.

Results and Discussion

The probability distribution of the fractions of H1-H2 and H2-H3, nonlocal ($i - j > 3$), native atomic contacts generated from 197 folding simulations is shown in Figure 1a (see Materials and Methods). In addition to the native and coil-like states, there are two peaks that correspond to intermediates with well-formed H1-H2 and H2-H3 microdomains, respectively. The greater height of the peak corresponding to the H2-H3 microdomain to the peak of the H1-H2 microdomain indicates the dominance of this intermediate.

The folding mechanism of BpA is summarized in Figure 1b. BpA folds by parallel pathways via a diffusion-collision mechanism (Karplus and Weaver 1976). The folding is initiated by the partial, and strongly fluctuating, formations of H1, H2, and H3, with H3 as the most stable and first-formed helix (Linhananta and Zhou, in prep.). This is followed by the formation of an on-pathway intermediate. The dominant pathway (126 out of 197, or 64%) folds via an intermediate with a well-established H2-H3 microdomain (I23). In con-

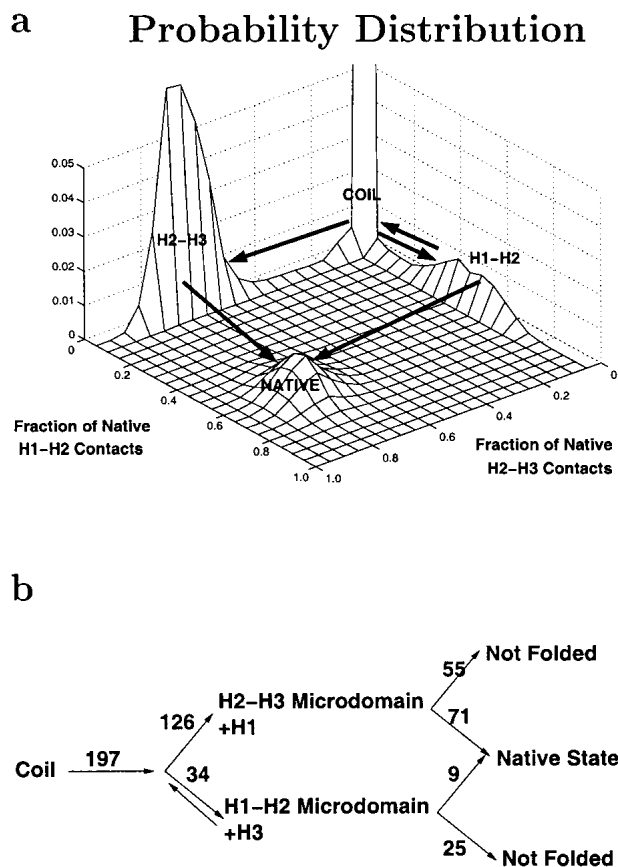


Fig. 1. (a) Probability distribution of the fractions of native H2-H3 and H1-H2 contacts obtained from 197 simulation trajectories. (b) A schematic plot for the folding pathway of BpA. The pathway via H2-H3 is dominant (126 of 197, 64%) over the one via H1-H2 (34 of 197, 17%). The free energy barrier for folding to the native state is greater from H1-H2 (9 of 34, 26%) than from H2-H3 (71 of 126, 56%).

trast, only 34 out of 197 trajectories (17%) involve an intermediate with a H1-H2 microdomain (I12). The rest (37 out of 197, 19%) remain in a coil-like structure for the entire simulation. The isolated H2-H3 microdomain has 31 more nonlocal native atomic contacts than the H1-H2 microdomain and hence I23 has a lower energy than I12. In addition, the structure of the H1-H2 microdomain fluctuates strongly, and the microdomain often dissociates back to the coil-like state. Once formed, the rate of folding to the native state from I23 (71 out of 126, 56%; Fig. 1b) is higher than from I12 (9 out of 34, 26%; Fig. 1b). This indicates that the free-energy barrier between the native and I12 states is higher than between the native and I23 states. This has been shown to be associated with the low stability of I12 (Linhananta and Zhou, in prep.). A typical folding trajectory in the dominant folding pathway is shown in Figure 2.

The dominant folding pathway found in this study is highly consistent with available experimental data. Hydrogen exchange data indicates that H3 is the most stable isolated helix (30% helical content; Bai et al. 1997). The domi-

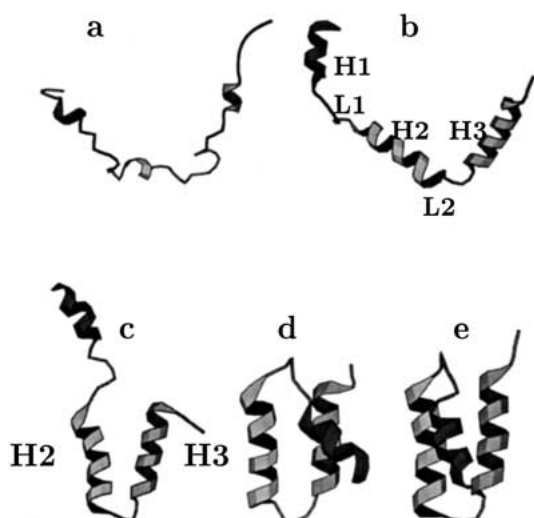


Fig. 2. The folding events of a typical trajectory that folds by the dominant pathway: the coil state ($t^* = 0$); the formation of helices ($t^* = 1000$); the formation of the H2-H3 microdomain ($t^* = 36,170$); the contact between loop 1 (L1) and helices H2 and H3 ($t^* = 47,990$); the final docking of H1 onto the H2-H3 domain ($t^* = 48,870$). Drawn using MOLSCRIPT (Kraulis 1991). L1, L2, H1, H2, and H3 are as labeled.

nant role of I23 is also consistent with the fact that a H2-H3 microdomain fragment contains 50% helix, compared to 12% for the H1-H2 fragment (Bai et al. 1997). Moreover, in an experimental denaturation study of BpA, Bottomley et al. (1994) stated that “H1 unfolded first, followed by H2 and H3 together”. This is exactly the reverse of the average folding pathway observed here. Furthermore, the diffusion-collision-like folding mechanism found here agrees with the fact that experimental kinetic data of BpA can be interpreted quantitatively by a diffusion-collision model (Karplus and Weaver 1976; Myers and Oas 2001) with experimentally measured fractional helicity (Bai et al. 1997) as input. The folding time (16 μ sec at 40% folded) is also to the same order of magnitude of the measured value (6 μ sec; Myers and Oas 2001).

Although this study suggests a specific pathway that differs from other theoretical works, some important aspects have been previously observed. For example, I12 was also found in the equilibrium free-energy surface analysis based on an all-atom CHARMM model in explicit solvent (Boczko and Brooks 1995; Guo et al. 1997). The faster formation of H3 in our model is consistent with the slower unfolding of H3 in an all-atom unfolding simulation study of BpA in explicit solvent (Alonso and Daggett 2000). In addition, loop formation as a rate-determining step was found in other theoretical studies of BpA as well (Shea et al. 1999; Berriz and Shakhnovich 2001). However, the focus of this paper is not on the accuracy of our results, but rather on the understanding of the high-resolution details obtained from the folding simulations of the model BpA. We will show that

our interpretation of the results has important general implication to the folding and misfolding of proteins.

The all-atom folding simulations reveal that L2 forms first most of the time. To interpret this, we define loop contact distance as the total contact distance (TCD) between the loop and the rest of the protein (see Materials and Methods). Similar to the observed correlation between high folding rates and low TCD values, loops with lower LCD values are expected to form earlier. Indeed, we found that the LCD value for L2 of BpA (0.22) is 4.3 times less than the value for L1 (0.94). This explains the dominance of the I23 pathway by a factor of 3.7 over the I12 pathway. L1 has many nonlocal long-distance contacts ($i - j > 3$) with H2 and H3, whereas L2 has only two short-distance, nonlocal contacts (Pro 39 with Ser 34 and Leu 35). A loop with a large number of nonlocal contacts (high LCD) must overcome a large entropic barrier, associated with the conformational search to make the contacts, and hence will take longer to form. The early formation of L2 was also observed in a C_α -based model of BpA that has an orientationally dependent native potential (Berriz and Shakhnovich 2001).

To further prove the utility of LCD, one needs to know the rates of the formation of different loops in a protein. Such information is not yet directly available. The role of loop formation in protein folding has been mostly derived from protein engineering experiments (Fersht 1995). In protein engineering experiments, the transition-state ensemble of a protein-folding reaction is characterized by ϕ values. For any residue, $\phi \sim 1$ means that native contacts involving the residue are mostly formed at the transition state, whereas the opposite is true for $\phi \sim 0$.

To address the question of whether or not a loop with the lowest LCD value is most likely to have the highest ϕ value, we compiled a list of proteins with high $\phi \sim 1$ values in the loop regions (Table 1). That is, we only surveyed proteins with a loop that is known to be formed at the transition state. We should emphasize that only a qualitative comparison can be made because a ϕ value is not a direct indicator for the rate of folding. The results are mixed.

For all- β proteins, the loop with $\phi \sim 1$, indeed, has the lowest LCD value (WW domain, α spectrin SH3, src-SH3) or a value close to the lowest (Sso7d-SH3 domain). There is no experimental result for all- α proteins. For α/β proteins, however, LCD values are predictive for loops with $\phi \sim 1$ with the stipulation that the comparison of LCD values is between the loops connecting β strands only (EE loops of Table 1). For example, the LCD values of the loop regions of β -hairpins 1 and 2 of protein L are 0.89 and 1.63, respectively. In other words, the first β -hairpin folds first, as observed experimentally (Kim et al. 2000). On the other hand, the LCD values of the loop regions of β -hairpins 1 and 2 of protein G are 0.98 and 0.74, respectively. Thus, consistent with experiments (McCallister et al. 2000), it is the second β -hairpin that folds first in protein G. More

Table 1. The values of loop contact distance versus loops (in boldface) that have a ϕ value close to 1

Protein	PDB ID	n_r	Loop:LCD ^a			
All- β proteins:						
WW Domain	1PIN	34	EE 16–21^b: 0.32	EE 27–31: 0.72		
α -SH3	1SHG	57	EE 16–21: 0.52	EE 26–69: 0.79	EE 36–40: 0.45	EE 47–48^c: 0.21
src-SH3	1SRL	56	EE 13–30: 1.1	EE 35–41: 0.34	EE 48–51^d: 0.30	EH 58–60: 1.9
Sso7d-SH3	1SSO	62	EE 7–10: 0.37	EE 16–20: 0.92	EE 26–27^e: 0.30	EE 34–40: <u>0.27</u> EH 46–54: 1.1
α/β proteins:						
Protein G	1GB1	56	EE 9–12: 0.98	EH 20–22: 0.66	EH 36–41: 0.69	EE 47–50^f: 0.74
	NUG1	56	EE 9–12^g: 0.66	EH 20–22: 0.58	EH 36–41: 0.61	EE 47–50: 1.0
	NUG2	56	EE 9–12^g: 0.90	EH 20–22: 0.58	EH 36–41: 0.74	EE 47–50: 1.0
Protein L	2PTL	62	EE 26–31^h: 0.89	EH 39–40: 1.2	EH 54–60: 0.59	EE 67–70: 1.63

^a EE: loop between two β strands; EH: loop between a β strand and a helix. For src-SH3 and Sso7d-SH3, EH: is the loop between a β strand and a 3_{10} helix. An underlined number denotes the lowest LCD value among EE loops.

^b Jäger et al. 2001.

^c Martínez and Serrano 1999.

^d Riddle et al. 1999.

^e Guerois and Serrano 2000.

^f McCallister et al. 2000.

^g Nauli et al. 2001.

^h Kim et al. 2000.

significantly, the respective roles of the two β -hairpins of protein G, which were reversed by engineering the turn region (Nauli et al. 2001), are also reflected by the reversal of the LCD values (NuG1 and NuG2 of Table 1). The probability that the observed correlation between the loop with the highest ϕ value and the EE loop with the lowest LCD value occurred by chance is 0.003.

It is not clear why LCD values are predictive only for the loop between two β strands (and possibly between two α helices for which there is no data). It could be that the transition state is more structured than indicated by ϕ -value analysis (Bulaj and Goldenberg 2001; Ozkan et al. 2001). Or, more likely, it suggests the limitation of the LCD parameter. After all, the TCD parameter (see Materials and Methods), from which the LCD parameter is derived, is an empirical parameter that has an approximate correlation with folding rate (Zhou and Zhou 2002). It cannot accurately describe the mutation-induced change in folding rate. Nevertheless, the results verify the concept that for many proteins, β -sheet proteins in particular, the loop with the lowest LCD value plays a key role in folding.

The results presented above also highlight the important role of the detailed native structure in determining folding pathways (Zhou and Linhananta 2002b), even for proteins with identical topology. For example, both proteins G and L are made of a four-stranded β -sheet that packs with an α helix. The difference in their folding pathways was attributed (McCallister et al. 2000) to the difference between the internal stabilities of the two β -hairpin loops. This is consistent with our findings here. A loop with a low LCD value is somewhat isolated from the rest of the protein and thus is likely to be more stable in isolation.

Another interesting observation regarding the folding simulations of BpA is that the structures of I12 and I23 have

domain-swapped forms, similar to those found in simpler C_α -based model (Zhou and Karplus 1999). Domain-swapping (Schlunegger et al. 1997) refers to the exchange of a domain of a protein with the same domain of a second identical protein. An interesting question is, what would be the most probable dimeric domain-swapped structure for the model BpA if it exists? Clearly, a longer-lived intermediate would have a higher probability of colliding with another intermediate to form a domain-swapped dimer. Because I12 has a higher barrier for folding (see Fig. 1b), it is more likely to serve as a template for a domain-swapped dimer. Physically, it can be understood as the lack of strong driving force for the formation of the weakly interacting L2 in the late stage of folding. This picture is consistent with the view that in the early stage of protein folding, an unfolded protein must overcome entropic barriers, whereas in the later stage, the barriers are more energetic (Šali et al. 1994).

The above interpretation suggests that a loop with low LCD may serve as the hinge region for domain swapping. To verify this hypothesis, we performed a literature search for proteins that have both monomeric and dimeric domain-swapped structures. Proteins with dimeric domain-swapped structures but without corresponding structures for same-sequence monomers are not included in the database. For these proteins, we decided not to use monomer structures from homologous proteins or mutants because LCD values may be significantly different even for proteins within the same structural family, as shown in Table 1. For example, the domain-swapping human cystatin C (Janowski et al. 2001) is not included in the database because only the monomer structure of chicken cystatin C is available.

The LCD values of proteins with both monomeric and domain-swapped dimeric structures are shown in Table 2. The LCD values correctly predict the hinge regions of prion,

Table 2. The values of loop contact distance calculated from monomeric structures are compared with the loops (in boldface) that serve as the hinge loop for the corresponding domain-swapped dimers

Protein	Monomer PDB ID	n_r	Loop: LCD ^a			
Prion	1QLX	104	EH 131–144: 1.1	HH 154–172: 0.97	HH 195–199^b: 0.20	
D _{9k}	1CB1	78	EH 18–24: 0.68	HH 39–50^c: 0.38	HH 56–63: 0.94	
protein L (V49A) ^d	1K50	62	EE 12–16: 0.94	EH 40–46: <u>0.63</u>	EE 53–56³: 0.93	
Cyanovirin-N	2EZN	101	EE 24–28: 0.56	EH 50–53^f: 0.37	EE 65–67: 0.72	EE 75–79: <u>0.30</u>
EPS8-SH3	1I0C	59	EE 12–21: 0.84	EE 26–29: 0.75	EE 34–39^g: 0.47	EE 45–49: 0.50
Grb2-SH2	1QG1	104	EH 9–11: 0.68	EH 19–26: 1.4	EE 32–40: 0.46	EH 58–68^h: 0.33 EH 83–93: 0.94
RNase A	2AAS	124	HH 14–24ⁱ: 0.77	EH 33–42: 0.83	EH 48–50: 1.1	EE 56–60: 0.85 EE 64–71: 0.53
			EE 75–78: 0.60	EE 87–96: 0.71	EE 112–115^j: 0.38	

^a The lowest LCD value is underlined. HH: loop between two helices.

^b Knaus et al. 2001.

^c Hakansson et al. 2001.

^d The residue index in 1K50 differs from that in 2PTL (Table 1) by 14.

^e O'Neill et al. 2001.

^f Yang et al. 1999.

^g Kishan et al. 2001.

^h Schiering et al. 2000.

ⁱ Liu et al. 1998.

^j Liu et al. 2001.

D_{9k}, EPS8-SH2 domain, and Grb2-SH2 domain. For RNase A, the loop with the lowest LCD value corresponds to the hinge region of the major domain-swapped dimer (Liu et al. 2001). For protein Cyanovirin-N, the hinge region has the second lowest LCD with a value close to the lowest value. One exception is the hinge region of the minor domain-swapped dimer of RNase A (Liu et al. 1998) that has an intermediate LCD value. This may be due to the fact that one of the hinges becomes helical in the dimeric form. A more serious exception is protein L V49A mutant, for which the EH loop (the loop between the helix and strand 3) has the lowest LCD value but the second EE loop (the second β -hairpin) is the hinge region. Nevertheless, the analysis shows that in most cases (subject to the limitation of the availability of experimental data), a loop with the lowest LCD value is prone to domain swapping. The probability that the observed correlation between the hinge loop and the loop with the lowest LCD value occurred by chance is 0.0003 (excluding the minor domain-swapped dimer of RNase A).

This study is the first that predicts the hinge loop of domain-swapped dimer from the monomeric native structure of a protein. In some cases, however, other factors can result in the failure of LCD to predict the correct domain-swapped form. For example, nonnative interaction can give rise to conformational changes, such as observed in the minor domain-swapped form of RNase A (Liu et al. 1998). In other words, nonnative interprotein interaction involving hinge regions can be important. This was seen in a recent protein-design study in which the monomeric form of protein L is converted to an obligated dimer by mutations that favor interprotein interaction between hinge loops (Kuhlman et al. 2001).

Hence, a weakly interacting loop can play dual, and opposite, roles. It can promote folding by bringing connecting secondary structure units together to form a transition state. Alternatively, it can participate in misfolding by serving as the hinge region for a domain-swapped dimer if it did not fold in the initial stage. More remarkable is the fact that these specific behaviors are encoded in the native structure and can be predicted in most cases by a simple structural parameter even for proteins with identical native topology. Moreover, the evidence reported in this work suggests that the mechanism of misfolding, as manifested in the formation of domain-swapped dimer, is also largely determined by the native structure.

Materials and methods

Model

The detail of the model has been described elsewhere (Zhou and Linhananta 2002a). The model BpA is composed of 459 heavy atoms and polar hydrogens (residues 10–55). The initial coordinates were obtained from the PDB databank (ID No. 1BDD). Two bonded atoms, any 1,3 angle-constrained pair and 1,4 aromatic carbon pair, as well as improper dihedral angles were constrained by infinitely deep square-well potentials. Nonbonded atoms interacted by square-well potentials. A Gō model in which the square-well depth is $-\epsilon$ for atomic pairs with overlap in the native structure and 0, otherwise was used. The hard-core and square-well diameters were 0.8 and 1.2 times the Van der Waals diameters obtained from the CHARMM parameter set 19 (Neria et al. 1996).

Kinetics

Discontinuous molecular-dynamics techniques were used for the simulations. There were 197 folding simulations started from the

equilibrated coil-like state at $T^* = k_B T / \varepsilon = 4$ and quenched to $T^* = 2.5$. (The folding transition temperature T_f^* is 3.3, obtained by a weighted histogram analysis of equilibrium simulation data at different temperatures [Zhou and Linhananta 2002a]. No significant changes in folding kinetics were observed at $T^* = 3$.) The physical time unit was obtained by setting 100 reduced time units, the approximate contact time between any two residues that are two residues apart, to be equal to the experimentally measured time of ~20 nsec (Bieri et al. 1999). Structures were saved every 100 reduced time units, for a total of 800 recorded structures for each simulation. The fractions of native H1-H2 and H2-H3 atomic contacts (i.e., the ratio between the number of native contacts in the structure to the number of native contacts in the native structure) were calculated for each structure. The two-dimensional probability distribution in Figure 1a was calculated by binning the fraction values of the 197×800 configurations into bins with widths of 0.05.

Loop contact distance

Loop contact distance LCD is derived from total contact distance, (TCD; Zhou and Zhou 2002) which improves over contact order (Plaxco et al. 1998) in predicting folding rates. TCD is defined as the contribution to the average sequence separation by contacting residues within a cutoff distance R_{cut} .

$$TCD = \frac{1}{n_r^2} \sum_{\substack{k=1 \\ |i-j| > l_{\text{cut}}}}^{n_c} |i - j|, \quad (1)$$

where the summation is over all n_c contacts between residues j and i and n_r is the number of amino acid residues of a protein (excluding disordered regions). A contact between two residues is defined by the existence of any heavy atom pairs within a cutoff distance of R_{cut} and separated by a residue separation cutoff of l_{cut} . The term n_r^2 is related to the number of pairs of residues in a pair distribution function (Zhou and Zhou 2002). TCD was found to be the most accurate parameter in predicting folding rate for a wide range of l_{cut} and R_{cut} values (Zhou and Zhou 2002). It differs from contact order (Plaxco et al. 1998) by a factor of n_r/n_c . LCD is defined as the total contact distance (Zhou and Zhou 2002) between the residues of a loop and the rest of the protein.

$$LCD = \frac{1}{n_i n_l} \sum_{\substack{k=1 \\ |i-j| > l_{\text{cut}}}}^{n_c} |i - j|, \quad (2)$$

where the summation is over all contacts between residue j and loop residue i and n_l is the number of residues in the loop. Here, we use $l_{\text{cut}} = 0$ and a residue-residue cutoff distance of $R_{\text{cut}} = 6 \text{ \AA}$ because these values yielded the best result in studies of total contact distance (Zhou and Zhou 2002). Other values were used but did not affect the results presented here. A loop is identified with the Kabsch-Sander algorithm (Kabsch and Sander 1983). For domain-swapping, because hinge loop lengths are 4–22 residues (Schlunegger et al. 1997), only loops with more than three residues are considered. Two loops are merged if they are separated by only two residues. For predicting high- ϕ loops, no such restriction was applied.

Acknowledgments

We thank S. Nauli and Prof. D. Baker for providing the computer-designed structures of NuG1 and NuG2. This work was supported

by a grant from HHMI to State University of New York (SUNY) at Buffalo and by the Center for Computational Research and the Keck Center for Computational Biology at SUNY-Buffalo.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Alonso, D.O.V. and Daggett, V. 2000. Staphylococcal protein A: Unfolded pathways, unfolded states, and differences between the B and E domains. *Proc. Natl. Acad. Sci.* **97**: 133–138.
- Bai, Y.W., Karimi, A., Dyson, H.J., and Wright, P.E. 1997. Absence of a stable intermediate on the folding pathway of protein A. *Protein Sci.* **6**: 1449–1457.
- Berriz, G.F. and Shakhnovich, E.I. 2001. Characterization of the folding kinetics of a three-helix bundle protein via a minimalist Langevin model. *J. Mol. Biol.* **310**: 673–685.
- Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., and Drewello, M. 1999. The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci.* **96**: 9597–9601.
- Boczko, E.M. and Brooks III, C.L. 1995. First principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269**: 393–396.
- Bottomley, S.P., Popplewell, A.G., Scawen, M., Wan, T., Sutton, B.J., and Gore, M.G. 1994. The stability and unfolding of an IgG binding protein based upon the B domain of protein A from *Staphylococcus aureus* probed by tryptophan substitution and fluorescence spectroscopy. *Protein Eng.* **7**: 1463–1470.
- Bulaj, G. and Goldenberg, D.P. 2001. ϕ -Values for BPTI folding intermediates and implications for transition state analysis. *Nat. Struct. Biol.* **8**: 326–330.
- Favrin, G., Irbäck, A., and Wallin, S. 2002. Folding of a small helical protein using hydrogen bonds and hydrophobicity forces. *Proteins* **42**: 99–105.
- Fersht, A.R. 1995. Characterizing transition states in protein folding: An essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**: 79–84.
- Guerois, R. and Serrano, L. 2000. The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **304**: 967–982.
- Guo, Z.Y., Brooks III, C.L., and Boczko, E.M. 1997. Exploring the folding free energy surface of a three-helix bundle protein. *Proc. Natl. Acad. Sci.* **94**: 10161–10166.
- Hakansson, M., Svensson, A., Fast, J., and Linse, S. 2001. An extended hydrophobic core induces EF-hand swapping. *Protein Sci.* **10**: 927–933.
- Jager, M., Nguyen, H., Crane, J.C., Kelly, J.W., and Gruebele, M. 2001. The folding mechanism of a β -sheet: The WW domain. *J. Mol. Biol.* **311**: 373–393.
- Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. 2001. Human cystein C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat. Struct. Biol.* **8**: 316–320.
- Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577–2637.
- Karplus, M. and Weaver, D.L. 1976. Protein-folding dynamics. *Nature* **260**: 404–406.
- Kim, D.E., Fisher, C., and Baker, D. 2000. A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* **298**: 971–984.
- Kishan, K.R., Newcomer, M.E., Rhodes, T.H., and Guillot, S.D. 2001. Effect of pH and salt bridges on structural assembly: Molecular structures of the monomer and intertwined dimer of the Eps8 SH3 domain. *Protein Sci.* **10**: 1046–1055.
- Knaus, K.J., Morillas, M., Swietnick, W., Malone, M., Surewicz, W.K., and Yee, V.C. 2001. Crystal structure of the human prion protein reveal a mechanism for oligomerization. *Nat. Struct. Biol.* **8**: 770–774.
- Kolinski, A., Galazka, W., and Skolnick, J. 1998. Monte Carlo studies of the thermodynamics and kinetics of reduced protein models—application to small helical, β , and α/β proteins. *J. Chem. Phys.* **108**: 2608–2617.
- Kraulis, P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Applied Cryst.* **24**: 946–950.
- Kuhlman, B., O'Neill, J.W., Kim, D.E., Zhang, K.Y.J., and Baker, D. 2001. Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc. Natl. Acad. Sci.* **98**: 10687–10691.
- Liu, Y., Hart, P.J., Schlunegger, M.P., and Eisenberg, D. 1998. The crystal

- structure of a 3D domain-swapped dimer of RNase A at a 2.1 Å resolution. *Proc. Natl. Acad. Sci.* **95**: 3437–3442.
- 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**: 211–214.
- Martínez, J.C. and Serrano, L. 1999. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* **6**: 1010–1016.
- McCallister, E.L., Alm, E., and Baker, D. 2000. Critical role of β -hairpin formation in protein G folding. *Nat. Struct. Biol.* **7**: 669–673.
- Myers, J.K. and Oas, T.G. 2001. Preorganized secondary structure as an important determinant of fast protein folding. *Nat. Struct. Biol.* **8**: 552–558.
- Nauli, S., Kuhlman, B., and Baker, D. 2001. Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **8**: 602–605.
- Neria, E., Fischer, S., and Karplus, M. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**: 1902–1921.
- O'Neill, J.W., Kim, D.E., Johnsen, K., Baker, D., and Zhang, K.Y.J. 2001. Single-site mutations induce 3D domain swapping in the B1 domain of protein L from *Peptostreptococcus magnus*. *Structure* **9**: 1017–1027.
- Ozkan, S.B., Bahar, I., and Dill, K.A. 2001. Transition states and the meaning of π -values in protein folding kinetics. *Nat. Struct. Biol.* **8**: 765–769.
- Pande, V.S. and Rokhsar, D.S. 1998. Is the molten globule a third phase of proteins? *Proc. Natl. Acad. Sci.* **95**: 1490–1494.
- Plaxco, K.W., Simons, K.T., and Baker, D. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**: 985–994.
- Riddle, D.S., Grantcharova, V.P., Santiago, J.V., Alm, E., Ruczinski, I., and Baker, D. 1999. Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* **6**: 1016–1024.
- Šali, A., Shakhnovich, E.I., and Karplus, M. 1994. How does a protein fold? *Nature* **369**: 248–251.
- Schiering, N., Casale, E., Caccia, P., Giordano, P., and Battistini, C. 2000. Dimer formation through domain swapping in the crystal structure of the Grb2-SH2-Ac-pYVNV complex. *Biochemistry* **39**: 13376–13382.
- Schlunegger, M., Bennett, M., and Eisenberg, D. 1997. Oligomer formation by 3D domain swapping: A model for protein assembly and misassembly. *Adv. Protein Chem.* **50**: 61–122.
- Shea, J.E., Onuchic, J.N., and Brooks, C.L. 1999. Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci.* **96**: 12512–12517.
- Shea, J.-E., Onuchic, J.N., and Brooks III, C.L. 2000. Energetic frustration and the nature of the transition state in protein folding. *J. Chem. Phys.* **113**: 7663–7671.
- Yang, F., Bewley, C.A., Louis, J.M., Gustafson, K.R., Boyd, M.R., Gronenborn, A.M., Clore, G.M., and Wlodawer, A. 1999. Crystal structure of Cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J. Mol. Biol.* **288**: 403–412.
- Zhou, Y. and Karplus, M. 1997. Folding thermodynamics of a model three-helix-bundle protein. *Proc. Natl. Acad. Sci.* **94**: 14429–14432.
- . 1999. Interpreting the folding kinetics of helical proteins. *Nature* **401**: 400–403.
- Zhou, Y. and Linhananta, A. 2002a. Thermodynamics of an all-atom off-lattice model of the fragment B of *staphylococcal* protein A: Implication for the origin of the cooperativity of protein folding. *J. Phys. Chem. B* **106**: 1481–1485.
- . 2002b. Role of hydrophilic and hydrophobic contacts in folding of the second β -hairpin fragment of protein G: Molecular dynamics simulation studies of an all-atom model. *Proteins* **47**: 154–162.
- Zhou, H. and Zhou, Y. 2002. Folding rate prediction using total contact distance. *Biophys. J.* **82**: 458–463.