Specific Non-Native Hydrophobic Interactions in a Hidden Folding Intermediate: Implications for Protein Folding

Hanqiao Feng,[‡] Jiro Takei,[‡,§] Rebecca Lipsitz,[||] Nico Tjandra,[||] and Yawen Bai*[,‡]

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, and Laboratory of Biophysical Chemistry, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892

Received September 1, 2003; Revised Manuscript Received September 24, 2003

ABSTRACT: Structures of intermediates and transition states in protein folding are usually characterized by amide hydrogen exchange and protein engineering methods and interpreted on the basis of the assumption that they have native-like conformations. We were able to stabilize and determine the high-resolution structure of a partially unfolded intermediate that exists after the rate-limiting step of a four-helix bundle protein, Rd-apocyt b_{562}, by multidimensional NMR methods. The intermediate has partial native-like secondary structure and backbone topology, consistent with our earlier native state hydrogen exchange results. However, non-native hydrophobic interactions exist throughout the structure. These and other results in the literature suggest that non-native hydrophobic interactions may occur generally in partially folded states. This can alter the interpretation of mutational protein engineering results in terms of native-like side chain interactions. In addition, since the intermediate exists after the rate-limiting step and Rd-apocyt b_{562} folds very rapidly ($k_f \sim 10^4$ s^{-1}), these results suggest that non-native hydrophobic interactions, in the absence of topological misfolding, are repaired too rapidly to slow folding and cause the accumulation of folding intermediates. More generally, these results illustrate an approach for determining the high-resolution structure of folding intermediates.

To understand the mechanism of protein folding, a great deal of effort has been devoted to characterizing the structures of folding intermediates (*1*−*3*). However, a high-resolution structure of a partially unfolded intermediate is still not available. The major difficulty is that intermediates only exist transiently during kinetic folding. To date, the structures of the intermediates mainly have been characterized by using amide hydrogen exchange (*4*−*10*) and mutation studies (*11*−*14*). Backbone structure has also been studied using multidimensional NMR (*15*−*19*). Although atomic detail on backbone structures has been obtained in some cases, information about main chain topology and especially about side chain tertiary interactions is lacking.

Since recent studies have indicated that the topology of a protein structure plays an important role in determining the folding rates of small apparent two-state proteins (*20*), it is important to obtain such structural information for intermediates so that their kinetic role can be addressed. So far, tertiary interactions in partially unfolded intermediates have been studied in a few cases, using the protein engineering procedure (ϕ analysis) (*11*−*14*, *21*). In this procedure, a parameter, ϕ, is measured as the change in free energy of intermediate or transition states divided by the change in the free energy of the native state for a given mutation. The interpretation of ϕ assumes that the important side chain interactions in intermediates and transition states are generally native-like.

We have used native state hydrogen exchange to identify two partially unfolded intermediates of a redesigned four-helix bundle protein, Rd-apocyt b_{562} (*22*). These intermediates were shown to form after the initial rate-limiting transition state. The second stable intermediate has the N-terminal helix unfolded. Given this information, a mutant protein was designed using four glycines and an aspartic acid residue as substitutes for hydrophobic core residues in the N-terminal helix (*18*). These changes selectively destabilize the native state and cause the second, more advanced intermediate to become the most stable form. In addition, these mutations help to make the intermediate more soluble. This intermediate mimic, termed 4GD7, is expressed well in *Escherichia coli*. It is highly soluble and has a well-resolved ^{1}H−^{15}N HSQC spectrum. ^{15}N dynamics studies have shown that 4GD7 has nearly the same T_1, T_2, and NOEs as the fully folded state except in the mutated region, indicating that it is monomeric. We determined a high-resolution structure of the intermediate. The N-terminal helix segment is highly mobile. The other three helices are fully formed (*22*) and native-like, but to our surprise, significant non-native hydrophobic interactions were found.

MATERIALS AND METHODS

NMR Sample Preparation. Expression and purification of the intermediate 4GD7 were carried out as described previ-

* To whom correspondence should be addressed. E-mail: yawen@helix.nih.gov.

‡ National Cancer Institute.

§ Present address: Graduate School of Integrated Science, Yokohama City University, Tsurumi, Yokohama 230-0045, Japan.

|| National Heart Lung and Blood Institute.

10.1021/bi035561s
Published on Web 10/09/2003

ously (*18*, *22*). Isotopically enriched proteins for NMR structure determination were grown on M9 minimal medium containing 1 g/L [U-^{15}N]$^{15}NH_4Cl$ and/or 4 g/L [U-^{13}C]glucose (Isotec) for doubly (^{13}C and ^{15}N) and/or singly (^{15}N or ^{13}C) labeled proteins as the sole sources of nitrogen and carbon. NMR samples included ^{13}C/^{15}N, ^{15}N, ^{13}C, and nonlabeled protein at concentrations of ~2−3 mM (95% H_2O and 5% D_2O) at pH 5.0 with 20 mM NaAc-d_4 as a buffer. For residual dipolar coupling (RDC) measurements, Pf1 phage (16 mg/mL) was used to slightly align the protein sample (*23*, *24*).

NMR Spectroscopy. NMR spectra were collected at 25 °C on Bruker (Bellerica, MA) DRX 500 and 800 MHz spectrometers equipped with a 5 mm *x*,*y*,*z*-shielded pulse field gradient triple-resonance probe. A series of three-dimensional spectra [CBCA(CO)NH, HNCACB, HCCH-TOCSY, ^{15}N-edited TOCSY, ^{15}N- and ^{13}C-edited NOSEY, and two-dimensional (2D) ^{1}H NOESY] were collected for complete assignments and NOE measurement. An HNHA experiment was used to determine $^3J_{\alpha N}$ coupling constants (*25*). Hydrogen exchange rates were measured at 25 °C and pD_{read} 5.0 in D_2O. Many amide protons in all three helices have a protection factor of $>10^4$, calculated in the usual way (*26*). Amide protons with protection factors of >100 were considered to have hydrogen bonds formed. Four sets of residual dipolar coupling data (N−H, C′−N, C′−C_α, and C_α−H) were collected in a dilute phage solution (see ref *23* and references therein). NMR data were processed using NMRPipe (*27*) and analyzed using NMRview (*28*).

Calculation of Structure. Structural calculations were carried out using NIH X-PLOR (*29*). An extended polypeptide chain of reasonable geometry was used as the initial template. Backbone dihedral angles were then randomized before each cycle of the simulated annealing (SA) protocol. Each SA structure was optimized by restrained refinement. Ten structures with low energy and few NOE and dihedral angle violations from 50 SA structures were selected for refinement with RDC constraints. For each such starting structure, 10 RDC-refined structures (i.e., 100 structures) were calculated. The ones with low energy and no NOE and dihedral angle violations from each set were used for a second round of RDC refinement to calculate another 10 structures. This refinement procedure was repeated once more to obtain the final 10 structures with the lowest energy and no violation of restraints. These structures were further checked with PROCHECK-NMR (version 3.5.4) (*30*).

RESULTS

Structure of the Intermediate. ^{1}H−^{15}N HSQC spectra of the intermediate show very good chemical shift dispersion, which allowed full assignment of proton resonances. Altogether, 1686 NOEs were identified unambiguously using three-dimensional (3D) ^{15}N- and ^{13}C-edited NOESY and 2D ^{1}H NOESY for protons in aromatic side chains. The structures were determined on the basis of the constraints from NOEs, backbone dihedral angles, and hydrogen bonding. In addition, four sets of residual dipolar coupling values (N−H, C_α−H, C′−N, and C_α−C′) were measured using a dilute Pf1 phage solution to slightly align the intermediate. Figure 1 illustrates the RDC values of N−H dipoles. The near-zero values in the N-terminal helix segment indicate that the N-terminal helix is highly mobile, presumably

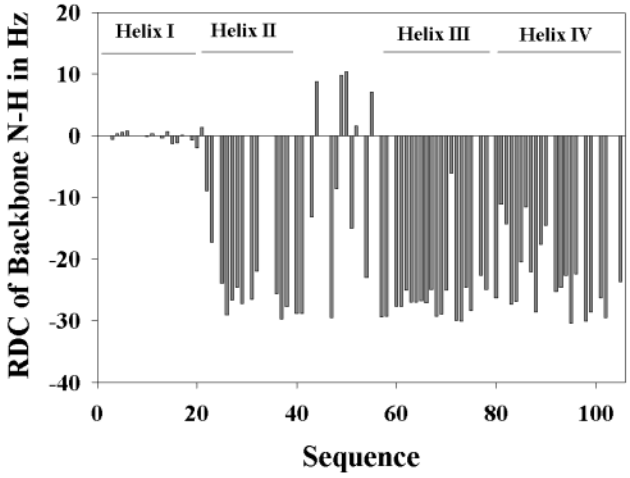

FIGURE 1: Residual dipolar coupling values of backbone N−H bonds measured at 25 °C and pH 5.0 in a dilute phage solution.

unfolded. The residual dipolar couplings for the other three helices are similar, indicating that they are close to being either parallel or antiparallel.

The structure of the folded region was well-defined with ~25 restraints per residue and 18 NOE constraints for each hydrophobic side chain. rmsds for the folded region in the intermediate are 0.35 and 1.11 Å for the backbone and all heavy atoms, respectively. Figure 2A shows the 10 structures superimposed on the backbone C_α atoms. Figure 2B illustrates the conformation of the hydrophobic residues in the same 10 structures but superimposed using all heavy atoms. The backbone and all hydrophobic side chains in the folded region except for Y105 are well-defined. The statistical parameters for the quality of the structure are listed in Table 1.

Non-Native Hydrophobic Interactions. Figure 3 superimposes the backbone structure of the intermediate on the native state. Some distortions exist, but the differences are minor. However, major differences in side chain tertiary packing are seen, as shown in Figure 4. In the native structure, the aromatic side chains of F61, F65, Y101, and Y105 face inside and form a hydrophobic cluster with W7 in helix I. In the intermediate, the side chains of F61 and F65 have rotated almost 180° away from the original hydrophobic core and are repacked in the region between helix II and helix III (Figure 4A,B). F65 forms a new hydrophobic contact with V69. Y101 and Y105 switch from the hydrophobic cluster to the interface between helix III and helix IV. In the native state, the hydrophobic core side chain of I98 interacts with W7 but not L68. In the intermediate, I98 and L68 move closer and form a new hydrophobic interaction. Repacking also occurs for hydrophobic residues at the other end of the four-helix bundle (Figure 4C,D). In the native state, L30 is in close contact with A79 and forms a hydrophobic bridge between residues V26 and L76. In the intermediate, L76 and V26 move closer and form a hydrophobic bridge between L30 and A79.

DISCUSSION

Non-Native Hydrophobic Interactions in Partially Folded States. Extensive previous studies on Rd-apocyt b_{562} folding intermediates used native state hydrogen exchange, NMR chemical shifts, dynamics analysis, and backbone NOEs.

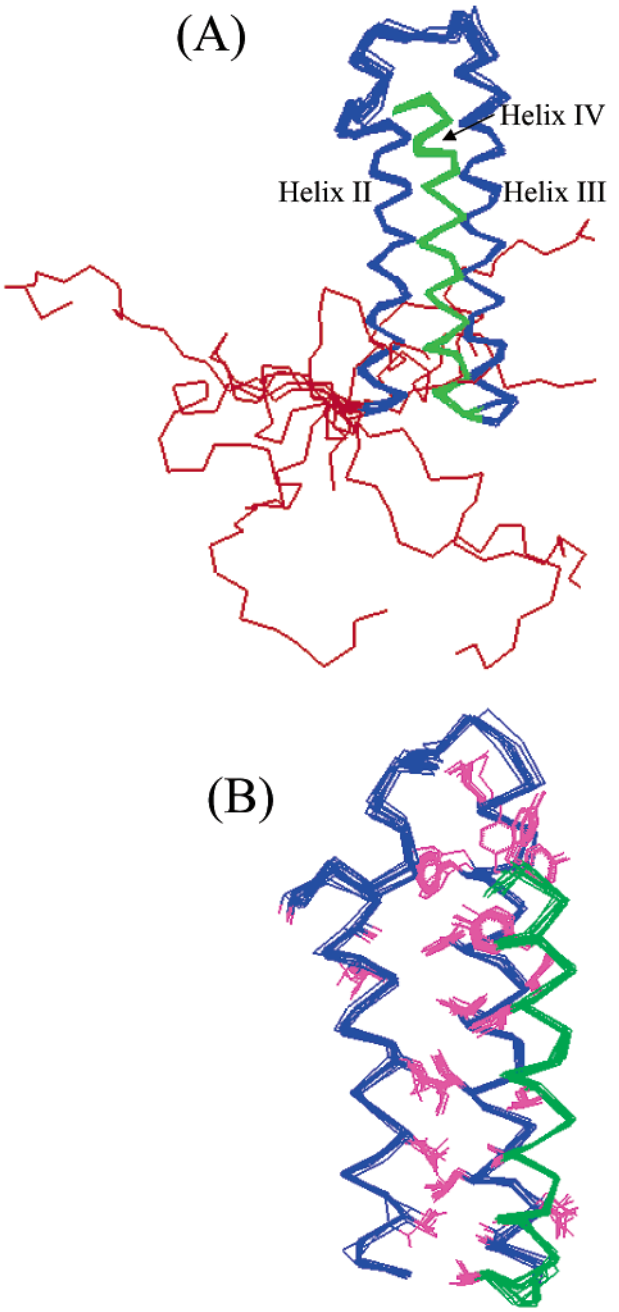

FIGURE 2: Structure of the partially unfolded folding intermediate of Rd-apocyt b_{562}. (A) C_α traces of 10 calculated structures superimposed on the folded region of the intermediate from residue 23 to residue 106. Helix I is in red. Helix II, helix III, and the loop between them are in blue. Helix IV is in green. (B) The same 10 structures were superimposed on all heavy atoms in the folded region of the intermediate. Hydrophobic side chains are shown in magenta.

Table 1: Parameters for the Structure of the Folded Region in the Partially Unfolded Intermediate

rmsd from ideal geometry	
bonds (Å)	0.00454 ± 0.00005
angles (deg)	0.623 ± 0.014
impropers	0.418 ± 0.013
rmsd from experimental constraints	
NOE (all) (Å)	0.0326 ± 0.005
dihedral (deg)	0.220 ± 0.06
rmsd from the mean structure	
(all residues/helical region)	
backbone atoms (Å)	0.35/0.30
all heavy atoms (deg)	1.11/1.11
experimental restraints	
NOEs	
intraresidue	502
sequential ($\|i - j\| = 1$)	1062
long-range ($\|i - j\| > 5$)	122
H-bonds	91
dihedral angles	67
RDCs	
$C'-C_\alpha$	80
$C_\alpha-H_\alpha$	55
$C_\alpha'-N$	68
N−H	65

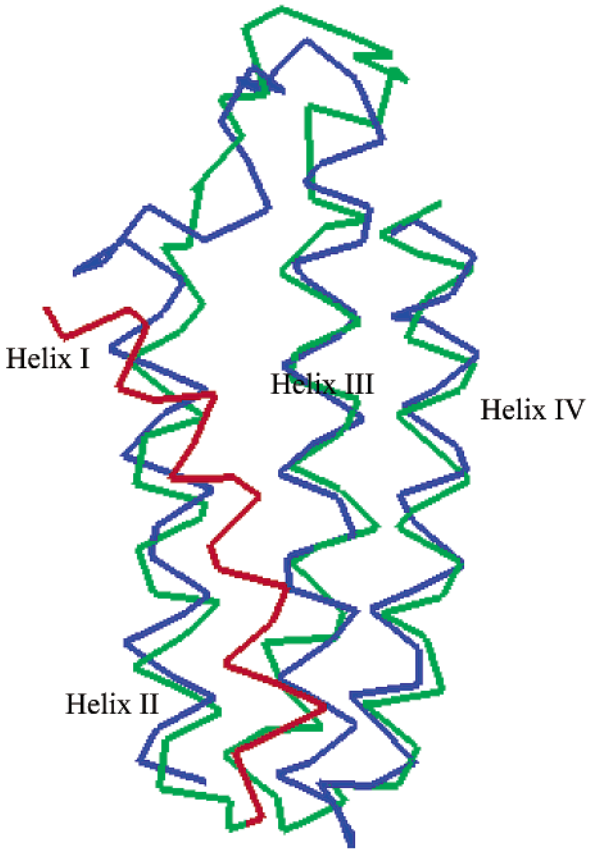

FIGURE 3: Comparison of the structures between the intermediate and fully folded Rd-apocyt b_{562}. The C_α trace corresponding to the folded region of the intermediate in a typical structure is superimposed on that of the native state (unpublished result). The intermediate is in blue, and the Rd-apocyt b_{562} is in green except helix I which is in red. The unfolded region (residues 1−22) of the intermediate is omitted.

These methods are insensitive to non-native side chain features (*18*, *22*). The high-resolution structure of the advanced intermediate studied here shows native-like secondary structure and backbone topology. However, non-native hydrophobic side chain interactions exist throughout the structure.

Many hydrophobic residues have solvent accessible surface areas similar to the native-like conformation but different interacting partners. L98 does not interact with L68 in the native state. However, the two residues are closely packed in the intermediate state (see Figure 4). F65 contacts Y101 in the native structure, but the two residues are completely separated and repacked in the folding intermediate. Similar changes of interacting partners occur for other residues such as L76 and V26, and L30 and A79. Overall, these changes

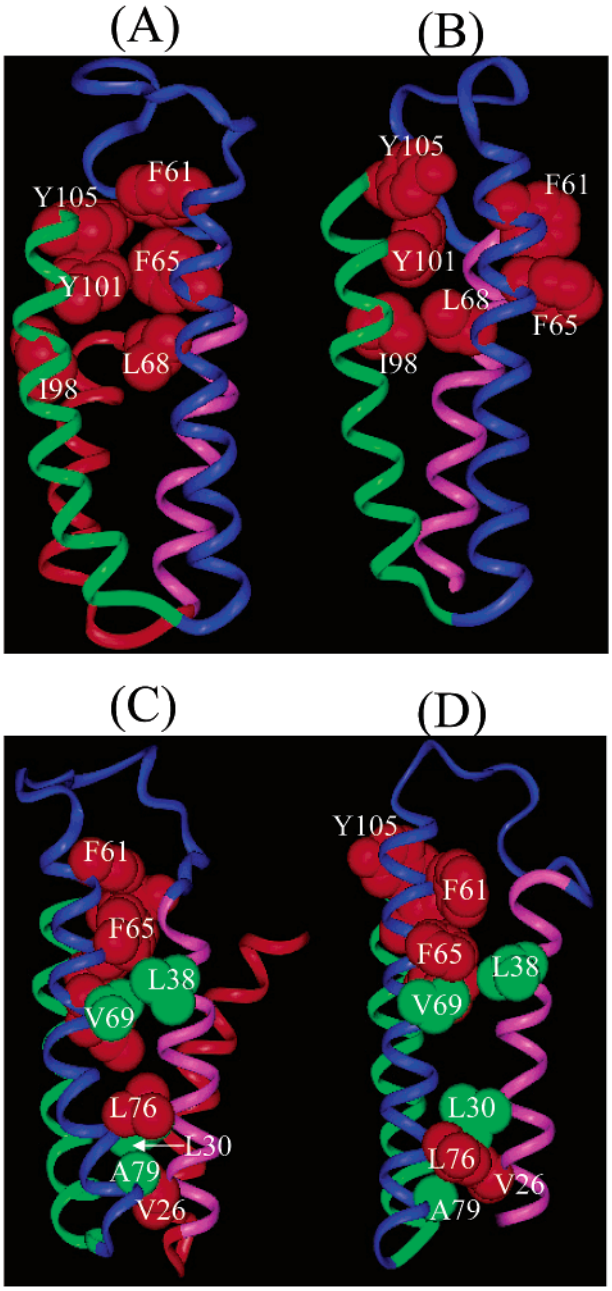

FIGURE 4: Illustration of the non-native hydrophobic interactions in the intermediate of Rd-apocyt b_{562}. (A) Structure of Rd-apocyt b_{562} with the side chains of residues F61, F65, V68, I98, Y101, and Y105 in CPK models. (B) Structure of the intermediate in the same orientation of the native state in panel A. (C) Structure of Rd-apocyt b_{562} with additional residues V26, L30, L79, and A79 from a different view. (D) Structure of the intermediate in the same orientation as the native state in panel C. The side chains in the native structure are better defined than the intermediate.

appear to represent a relaxation of the otherwise native-like structure which reorganizes the hydrophobic packing caused by the fact that helix I is not yet formed. These results suggest that non-native hydrophobic interactions may well occur in all partially unfolded transition and intermediate states. Because partial unfolding exposes hydrophobic core surfaces in the still folded region, a relaxation or reorganization of the partially folded structure to the most stable structure is likely to occur. Consistent with this possibility, non-native-like hydrophobic interactions have recently been suggested to exist in a folding intermediate of IM7 (*31*). Non-native hydrophobic interactions have also been detected in the unfolded states of lysozyme (*32*) and the initially collapsed state of Hpr (*33*).

Φ *Analysis*. ϕ analysis compares the mutational change in free energy for transition states or for intermediates with the change in free energy for the native state. The analysis assumes that the native interactions are conserved in the partially folded forms. The results described in detail here show that this assumption is questionable. The same question may arise for many theoretical studies that assume native-like contacts in partially unfolded transition states and use ϕ values to benchmark molecular dynamics simulation studies (*34*, *35*).

Non-Native Hydrophobic Interactions and the Population of Early Folding Intermediates. An important question in protein folding studies has been why folding intermediates are populated before rate-limiting steps (*7*, *9*, *36*). Two major hypotheses have been proposed. In the earlier studies, it was postulated that structure formation consists of at least two steps: sequential formation of secondary structures and tertiary side chain packing (*37*, *38*). It was suggested that formation of secondary structures is fast, whereas tertiary side chain packing is slow (*39*). Therefore, the population of molten globule-like folding intermediates before the rate-limiting step should occur generally (*40*). The generality of this kinetic framework or molten globule model, however, was challenged by later experimental results. It was found that some proteins could fold very rapidly in the absence of detectable early folding intermediates (*41*−*45*). These results have led to the new hypothesis that the population of early folding intermediates might not be an intrinsic folding event but rather is caused by some misfolding events in the intermediate (*41*, *45*−*50*). Misfolding involving non-native hydrophobic interactions in early folding intermediates has been suggested from both experiment and computer simulation studies (*43*, *48*). However, unlike the cases involving incorrect proline isomers (*51*), non-native disulfides (*52*), and misligation between the histidine side chain and Fe^{3+} in the heme (*41*, *53*), non-native hydrophobic interactions have not been convincingly shown to be the reason for population of early folding intermediate.

The structure of the late intermediate of Rd-apocyt b_{562} now shows that non-native hydrophobic interactions can also occur after the rate-limiting step, indicating that they are not unique to early folding intermediates. Therefore, non-native hydrophobic packing alone cannot be the reason for the population of the earlier folding intermediate. Moreover, Rd-apocyt b_{562} folds very rapidly with a rate constant of $\sim 10^4$ s^{-1} at 25 °C (*22*), showing that the repair of the hydrophobic mispacking is fast. It is therefore unlikely to become the rate-liming step needed for accumulation of early folding intermediates. This is consistent with the suggestion that topological misfolding is responsible for the barriers that cause intermediates to accumulate (*54*), and the recent finding that the folding rate of small proteins is dominantly controlled by the topological complexity of the native protein structure and only modulated by local energetic interactions (*20*).

Alternatively, population of early intermediates might arise from other causes rather than misfolding. For example, a protein with two independent domains will appear to fold through early folding intermediate if one domain folds faster than the other. Similarly, it is conceivable that proteins that

have two weakly associated domains or hydrophobic cores might also populate early folding intermediates. This is consistent with the observation that early folding intermediates are mostly observed in larger proteins with more than one hydrophobic core (*7*, *55*, *56*), suggesting that the size of a protein and the existence of multiple nuclei may play important roles in populating early folding intermediates. Structural determination of early folding intermediates will be needed to test these hypotheses.

CONCLUSIONS

The example presented here shows how intermediates previously characterized by hydrogen exchange can be populated under equilibrium conditions by protein engineering and studied at high resolution by NMR. Multiple non-native hydrophobic interactions were identified in the high-resolution structure of an otherwise hidden folding intermediate of Rd-apocyt b_{562}. The results have a number of implications. Non-native hydrophobic interactions may occur very generally in partially unfolded structures. The implication of native-like tertiary interactions from hydrogen exchange results and ϕ analysis of intermediates and transition states is questionable. High-resolution structures of folding intermediates are needed for benchmarking computer simulation studies on protein folding pathways. Non-native packing interactions alone, in the absence of topological misfolding, can be repaired too quickly to be responsible for the barriers that cause slow folding and intermediate accumulation.

REFERENCES

1. Kim, P. S., and Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459−489.
2. Matthews, C. R. (1993) *Annu. Rev. Biochem. 62*, 653−683.
3. Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 105−112.
4. Roder, H., Elove, G. A., and Englander, S. W. (1988) *Nature 335*, 700−704.
5. Udgaonkar, J. B., and Baldwin, R. L. (1988) *Nature 335*, 694−699.
6. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
7. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) *Nat. Struct. Biol. 3*, 782−787.
8. Englander, S. W. (2000) *Annu. Rev. Biophys. Biomol. Struct. 29*, 213−238.
9. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) *Science 249*, 1544−1548.
10. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
11. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
12. Barrick, D., and Baldwin, R. L. (1993) *Protein Sci. 2*, 869−876.
13. Bulaj, G., and Goldenberg, D. P. (2001) *Nat. Struct. Biol. 8*, 326−330.
14. Matouschek, A., Serrano, L., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 819−835.
15. Feng, Y., Sligar, S. G., and Wand, A. J. (1994) *Nat. Struct. Biol. 1*, 30−35.
16. Redfield, C., Smith, R. A., and Dobson, C. M. (1994) *Nat. Struct. Biol. 1*, 23−29.
17. Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry 39*, 2894−2901.
18. Takei, J., Pei, W., Vu, D., and Bai, Y. (2002) *Biochemistry 41*, 12308−12312.
19. Staley, J. P., and Kim, P. S. (1994) *Protein Sci. 3*, 1822−1832.
20. Makarov, D., and Plaxco, K. W. (2003) *Protein Sci. 12*, 17−26.
21. Goldenberg, D. P. (1999) *Nat. Struct. Biol. 6*, 987−990.
22. Chu, R. A., Pei, W. H., Takei, J., and Bai, Y. (2002) *Biochemistry 41*, 7998−8003.
23. Suzuki, M., Youle, R. J., and Tjandra, N. (2000) *Cell 103*, 645−654.
24. Hansen, M. R., Mueller, L., and Pardi, A. (1998) *Nat. Struct. Biol. 5*, 1065−1074.
25. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
26. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins 17*, 75−86.
27. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
28. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR 4*, 603−614.
29. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J. Magn. Reson. 160*, 65−73.
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
31. Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2002) *Nat. Struct. Biol. 9*, 209−216.
32. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Veda, T., Imoto, T., Dobson, C. M., and Schwalbe, H. (2002) *Science 295*, 1719−1722.
33. Canet, D., Lyon, C. E., Scheek, R. M., Robillard, G. T., Dobson, C. M., Hore, P. J., and van Nuland, N. A. J. (2003) *J. Mol. Biol. 330*, 397−407.
34. Daggett, V., and Fersht, A. R. (2003) *Trends Biochem. Sci. 28*, 18−25.
35. Mirny, L., and Shakhnovich, E. (2001) *Annu. Rev. Biophys. Biomol. Struct. 30*, 361−396.
36. Kim, P. S., and Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.
37. Shakhnovich, E. I., and Finkelstein, A. V. (1989) *Biopolymers 28*, 1667−1680.
38. Barrick, D., and Baldwin, R. L. (1993) *Protein Sci. 2*, 869−876.
39. Ptitsyn, O. B. (1994) *Protein Eng. 7*, 593−596.
40. Ptitsyn, O. B. (1995) *Curr. Opin. Struct. Biol. 5*, 74−78.
41. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
42. Jackson, S. E. (1998) *Folding Des. 3*, R81−R91.
43. Radford, S. E., Dobson, C. M., and Evans, P. A. (1992) *Nature 358*, 302−307.
44. Bai, Y., Karimi, A., Dyson, H. J., and Wright, P. E. (1997) *Protein Sci. 6*, 1449−1457.
45. Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9029−9033.
46. Bryngelson, J. D., and Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7524−7528.
47. Wolynes, P. G., Onuchic, J. N., and Thirumalai, O. D. (1995) *Science 267*, 1619−1620.
48. Guo, Z., and Thirumalai, D. (1995) *Biopolymers 36*, 83−102.
49. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 4*, 10−19.
50. Sali, A., Shakhnovich, E. I., and Karplus, M. (1994) *Nature 369*, 248−251.
51. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
52. Creighton, T. E. (1992) *BioEssays 14*, 195−199.
53. Elove, G. A., Bhuyan, A. K., and Roder, H. (1994) *Biochemistry 33*, 6925−6935.
54. Sosnick, T. R., Mayne, L., and Englander, S. W. (1996) *Proteins 24*, 413−426.
55. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
56. Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993) *Science 262*, 896−900.

BI035561S