Hierarchical Folding of Intestinal Fatty Acid Binding Protein†

Syun-Ru Yeh,‡ Ira J. Ropson,§ and Denis L. Rousseau*,‡

Department of Physiology and Biophysics, Albert Einstein College of Medicine, Morris Park Avenue, Bronx, New York 10461, and Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033

Received January 16, 2001

ABSTRACT: Intestinal fatty acid binding protein (IFABP) is a member of the lipid binding protein family, members of which have a clam shell type of motif formed by two five-stranded β-sheets. Understanding the folding mechanism of these proteins has been hindered by the presence of an unresolved burst phase. By initiating the reaction with a sub-millisecond mixer and following its progression by Trp fluorescence, we discovered three distinct phases in the folding reaction of the W6Y mutant of IFABP from which we postulate the following sequence of events. The first phase ($k_1 > 10\,000\ s^{-1}$) involves collapse of the polypeptide chain around a hydrophobic core. During the second phase ($k_2 \sim 1500\ s^{-1}$), β-strands B−G, mostly located on the top half of the clam shell structure, propagate from this hydrophobic core. It is followed by the final phase ($k_3 \sim 5\ s^{-1}$) involving the formation of the last three β-strands on the bottom half of the clam shell and the establishment of the native hydrogen bonding network throughout the protein molecule.

The protein folding problem is one of the most important questions in biology. Most folding studies have focused on proteins that have predominantly α-helical structures. The folding of β-sheet proteins is relatively unexplored, and could display quite different mechanisms (*1*, *2*). In α-helical proteins, the structures are stabilized by local hydrogen bonds between the CO group of residue *i* and the NH group of residue $i + 4$. In contrast, β-sheet proteins are stabilized by extensive interstrand hydrogen bonds between residues that are not close in sequence. It has been suggested that the rate-limiting step for folding of β-sheet proteins involves the formation of β-turns that connect β-strands (*3*−*5*). However, the folding of most β-sheet proteins typically takes more than a few seconds to complete (*1*, *2*), although only ∼10 μs is required for a model peptide to form the correct β-turn structure (*5*).

Members of the lipid binding protein family bind ligands such as fatty acids, retinoids, and bile salts in their central cavity (*6*). They have low levels of sequence homology but virtually identical three-dimensional structures, typified by the structure of IFABP[1] shown in Figure 1 (*7*). Prior folding studies on the lipid binding protein family have revealed some remarkable differences between the various members. Cellular retinol binding protein II (CRBP II) and IFABP both form intermediates in which there appears to be little if any secondary structure (*8*, *9*). On the other hand, cellular retinoic acid binding protein I (CRABP I) and ileal lipid binding protein (ILBP) both display intermediates with native-like secondary structure (*10*−*13*). The cause of the diverse folding kinetics in this protein family is unknown, in part due to the fact that the folding reactions of most of these proteins display large unresolved burst phases (*4*, *13*, *14*).

As a member of this class, IFABP is an attractive model for studying folding of a β-sheet protein as it is relatively small (15.1 kDa) and contains no proline or cysteine residues that might lead to misfolded structures. During the folding reaction of IFABP, the unresolved burst phase is followed by the formation of the native structure with monoexponential kinetics that requires ∼1 s to complete (*8*, *9*). The wild-type protein contains only two tryptophan residues at positions 6 and 82. Trp-82 has been proposed to be involved in forming an early hydrophobic nucleation site that initiates folding (*9*, *10*, *15*). To obtain new insights into the early folding events of IFABP and the role of the putative hydrophobic core, we have constructed the W6Y mutant and studied its folding kinetics by following the fluorescence from Trp-82 in a home-built continuous flow solution mixer with a dead time of ∼100 μs (*16*). We discovered two new phases, thereby resolving the folding of this protein into three distinct phases: a $>10000\ s^{-1}$ burst phase, a $1500\ s^{-1}$ rapid phase, and a $5\ s^{-1}$ slow phase.

MATERIALS AND METHODS

The W6Y mutant of IFABP was prepared as described previously (*15*). The unfolded protein (40 μM) was incubated with the desired concentration of urea (Aldrich) a few hours prior to the experiments. For the kinetics studies, the folding reaction was initiated by a 6-fold dilution of the unfolded

† D.L.R. and S.R.Y. are supported by National Institutes of Health Research Grants GM-54806 and GM-54812, and I.J.R. is supported by National Institutes of Health Research Grant GM-57906.

* To whom correspondence should be addressed: Department of Physiology and Biophysics, Albert Einstein College of Medicine, Morris Park Ave., Bronx, NY 10461. Telephone: (718) 430-4264. Fax: (718) 430-8808. E-mail: rousseau@aecom.yu.edu.

‡ Albert Einstein College of Medicine.

§ The Pennsylvania State University College of Medicine.

[1] Abbreviations: IFABP, intestinal fatty acid binding protein; CRBP, cellular retinol binding protein; ILBP, ideal lipid binding protein.

10.1021/bi0155044 CCC: $20.00 © 2001 American Chemical Society
Published on Web 03/14/2001

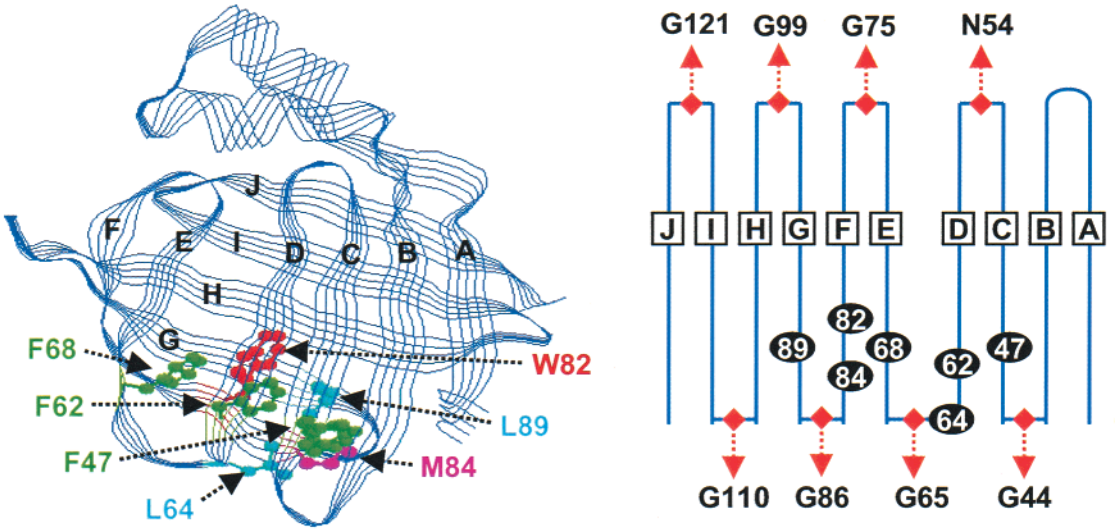

FIGURE 1: (Left) Crystal structure of intestinal fatty acid binding protein (PDB entry 1ifc; *7*). The postulated hydrophobic core is shown as a stick-and-ball presentation. (Right) Dissected smoothed representation of intestinal fatty acid binding protein with β-strands, hydrophobic core residues, and important turn residues labeled.

protein with the appropriate buffer solution in a rapid solution mixer with a mixing dead time of ~100 μs. In this rapid solution mixer, the two parent solutions were pressurized by a syringe pump and introduced into the T-shaped mixing nozzle, which is in direct contact with a quartz flow cell (250 μm × 250 μm). The detailed characterization of this mixer was described previously (*16*). The final protein concentration was 6.6 μM.

The output at 264 nm from a frequency-doubled argon ion laser (Coherent) was focused to a ~30 μm spot on the continuously flowing sample in the observation cell. The tryptophan fluorescence was collected and focused on the entrance slit of a 0.27 m polychromator (Spex) where it was dispersed and then detected by a charge-coupled device camera (Photometrics). The resulting spectra were not corrected with the instrumental response function, and thus are red-shifted about 10 nm compared to that obtained from a spectrally corrected steady-state fluorometer based on the measurements for *N*-acetyltryptophanamide (Aldrich). The time-resolved tryptophan fluorescence spectra were obtained by moving the quartz cell relative to the laser focusing point along the flow direction. The position of the flow cell was controlled by a micrometer on a translational stage in such a way that an accurate time domain could be calculated from the solution flow rate.

The stopped flow data were obtained with an Applied Photophysics SF17MV stopped flow system. The excitation wavelength for the stopped flow tryptophan fluorescence measurements was 290 nm, and the intensity was measured at 10 nm intervals between 310 and 400 nm (*9*, *10*).

For the measurement of the unfolding curve, the protein in the presence of various amounts of urea was prepared a few hours prior to the experiments. The resulting protein solution was loaded into two separate syringes, pressurized by a syringe pump for introduction into the T-shaped mixing nozzle and subsequent passage into the quartz observation cell. The tryptophan fluorescence from the protein was recorded as described above. As a comparison, the steady-state fluorescence measurements were recorded with an SLM-Aminco Bowman AB2 fluorimeter. For the latter measurements, the excitation wavelength was 290 nm with an excitation slit of 2 nm and an emission slit of 8 nm (*9*, *10*).

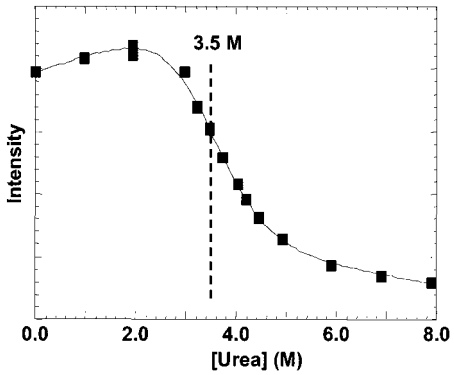

FIGURE 2: Urea titration curve of W6Y-IFABP monitored by tryptophan fluorescence.

RESULTS

As a first step, the equilibrium unfolding curve for the W6Y mutant was determined by following the tryptophan fluorescence in our sub-millisecond rapid mixing apparatus. These data, shown in Figure 2, give a midpoint of 3.5 M, indicating that the mutant is only slightly less stable than the wild-type protein in which the equilibrium midpoint concentration is 4.1 M (*9*, *10*). The unfolding curve obtained with our apparatus is identical to that obtained from steady-state measurements in a conventional fluorometer for this mutant, demonstrating the reliability of this apparatus.

The first kinetic measurement was carried out in our rapid mixing apparatus by a 6-fold dilution of the unfolded protein (in 6.0 M urea) into a buffer solution such that the final concentration of the denaturant was 1.0 M. The kinetic trace obtained at 330 nm is presented in Figure 3B. These data can be fitted by a single-exponential growth function with a rate constant of ~1500 s^{-1}. The initial value obtained by extrapolating the kinetic trace back to time zero does not match that expected for the unfolded protein at this final concentration of denaturant, indicating that there is a missing

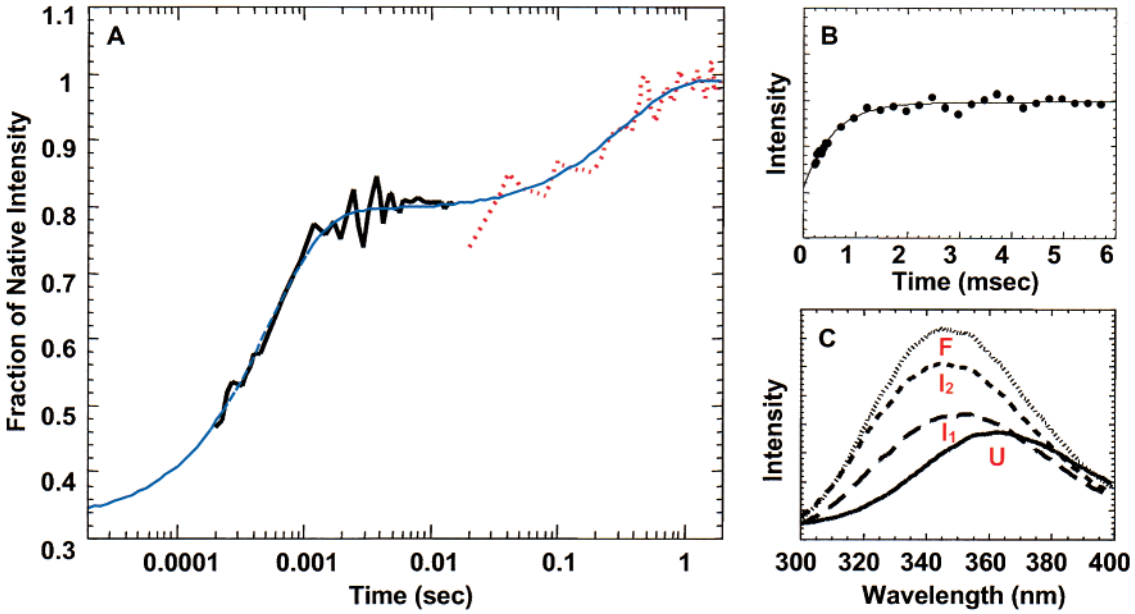

FIGURE 3: Refolding kinetics of W6Y-IFABP. (A) Plot of the full kinetic trace for the folding of W6Y-IFABP. The fraction of native intensity is calculated at 330 nm based on the ratio of $(I - I_U)/(I_F - I_U)$, where I_U and I_F are the fluorescence intensities of the unfolded and folded protein, respectively. The kinetic trace for the early time regime (<6 ms, thick solid line) was obtained from the home-built rapid solution mixer described in the text. The slower kinetic trace (dotted line) was obtained from conventional stopped flow instrumentation. The combined data were fit to a biexponential equation ($k_a = 1500\ s^{-1}$, $k_b = 5\ s^{-1}$, thin solid line). (B) Change in the fluorescence intensity at 330 nm as a function of time in the early time regime plotted in a linear scale. The initial and final urea concentrations are 6.0 and 1.0 M, respectively. (C) Tryptophan fluorescence spectra of the native state (F), the first folding intermediate (I_1), and the second folding intermediate (I_2) and the calculated spectrum of the unfolded state (U) at 1 M urea. All the spectra were obtained with the sub-millisecond mixing system and exhibit 10 nm red-shifts compared to those measured in commercial instruments due to the instrumental response function.

phase within the 100 μs dead time of the apparatus. The final kinetic trace (6 ms) matches that of the burst phase intermediate obtained from the stopped flow instrument. The recovery of the native state fluorescence from this burst phase intermediate in the stopped flow can be fitted with a single-exponential function with a rate of $\sim 5\ s^{-1}$ (Figure 3A). By combining these results, we concluded that that there are three distinct phases during folding: a burst phase appearing in less than 100 μs ($k_1 > 10\,000\ s^{-1}$) and leading to the first intermediate (I_1), a rapid phase ($k_2 \sim 1500\ s^{-1}$) resulting in the second intermediate (I_2), and a slow phase ($k_3 \sim 5\ s^{-1}$) terminating in the native state.

With a global fit of the fluorescence spectra at various delay times, the spectrum of I_1 was obtained by extrapolating the kinetic traces back to time zero as shown in Figure 3C. The spectrum of I_1 is significantly different from that expected for the unfolded protein (U), confirming that there is a significant burst phase unresolved in the sub-millisecond rapid mixer. The change in amplitude of the fluorescence intensity in this missing phase is small (Figure 3C) and insensitive to the urea concentration in the refolding solution experiments described below. The spectrum of I_2 shown in Figure 3C was obtained at the plateau region of the kinetics (see panels A and B of Figure 3). It is again significantly different from that of the native protein (F), indicating that the reaction is not yet complete at the end of the kinetic trace shown in Figure 3B. The spectrum of I_2 is the same as that of the extrapolated initial spectrum obtained from the stopped flow instrument, confirming that the end product from the second kinetic phase is the same as the burst phase intermediate observed in the stopped flow apparatus (with a dead time of ~10 ms).

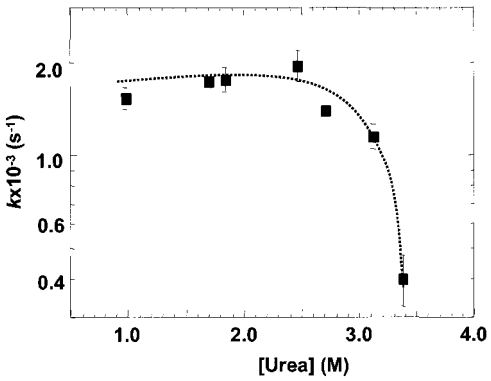

FIGURE 4: Refolding rate constants of W6Y-IFABP as a function of final urea concentration.

To further dissect the molecular mechanism that underlies the $I_1 \rightarrow I_2$ transition, a series of kinetic measurements were made in the sub-millisecond rapid mixer, such that the final concentrations of the denaturant varied from 1.0 to 3.4 M. The amplitudes at the end of the $I_1 \rightarrow I_2$ transitions (6 ms) for each final urea concentration matched that of the burst phase intermediate (10 ms) obtained in the stopped flow experiments (*9*). The kinetic traces fit well to a single-exponential function, and the resulting rate constants are plotted as a function of the final urea concentration in Figure 4. Unlike a typical chevron plot, the rate constants are not sensitive to the final urea concentration up to 2.5 M, suggesting that the rate-limiting step associated with the $I_1 \rightarrow I_2$ transition is not related to the burial of the urea sensitive surfaces. Although the rates of this transition were not affected by the final urea concentration, as the concentration of the urea was increased, the wavelengths of maximal

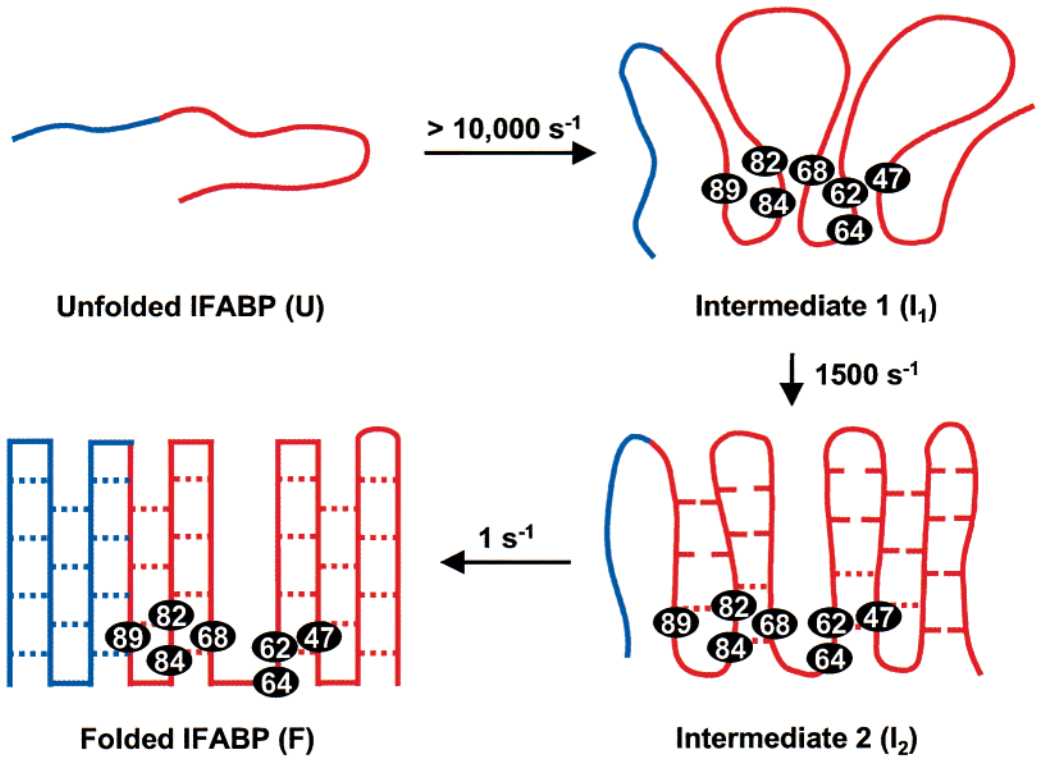

FIGURE 5: Schematic representation of the formation of the native conformation of W6Y-IFABP during the course of folding. The structure of IFABP is illustrated in a dissected smoothed representation for clarity. The U → I_1 transition involves hydrophobic collapse around residues 47, 62, 64, 68, 82, 84, and 89. In the subsequent phase (I_1 → I_2), these hydrophobic residues rearrange and the hydrophobic core becomes more compact. It serves as a nucleation center that facilitates the propagation of β-strands B−G toward the opposite end of the hydrophobic core. During the last phase of folding (I_2 → F), β-strands H−J are formed and the native hydrogen bonding network is established (indicated as the dotted lines).

fluorescence intensity of both the I_1 and I_2 intermediates were red-shifted and the amplitude for the I_1 → I_2 transition diminished (data not shown). These observations are attributed to the changes in the thermodynamic stability of I_1 and I_2 relative to the unfolded state in the presence of urea.

DISCUSSION

Both wild-type IFABP (*8*−*10*) and W6Y-IFABP (data not shown) fold completely reversibly at equilibrium, although W6Y-IFABP is slightly less stable. Very similar rates and relative amplitudes were observed during the folding and unfolding of these proteins by stopped flow fluorescence, and the intermediate states had very similar spectral properties. Previous studies have shown that the region around Trp-82 may serve as an initiating site for folding (*9*, *10*, *15*). As such, Trp fluorescence measurements in the W6Y mutant protein were used to selectively probe the structural changes in this region. On the basis of the results, the folding kinetics of IFABP are described as three sequential phases as follows:

$$\mathrm{U\ (365\ nm)} \xrightarrow{k_1} \mathrm{I_1\ (355\ nm)} \xrightarrow{k_2} \mathrm{I_2\ (345\ nm)} \xrightarrow{k_3} \mathrm{F\ (345\ nm)}$$

The numbers in parentheses indicate the uncorrected maximum of the tryptophan fluorescence peak in the submillisecond mixing system (Figure 3C). Although there may be alternative interpretations of these data (*17*, *18*), the simple on-pathway model presented here fully accounts for the observed results. A sequential model has been proposed in the past for another member of the family, CRABP I, although the details are very different as the rate constants are much slower and the protein displays considerable secondary structure immediately upon its collapse (*13*, *19*, *20*).

The fluorescence emission spectrum of the unfolded protein (U) shows a red-shifted maximum at approximately 365 nm, with respect to the value of 345 nm in the folded protein (F). It suggests that Trp-82 in the unfolded protein is in an aqueous environment, in contrast to the hydrophobic environment in the folded protein (*21*). The spectrum of the first folding intermediate (I_1) exhibits a maximum at approximately 355 nm and is increased in intensity with respect to that of the unfolded protein (Figure 3C). These data suggest that the initial phase of folding (U → I_1), which occurs in less than 100 μs, involves the burial of the Trp-82 residue in a more hydrophobic environment. Because the local sequence around Trp-82 (TGTWTME) is not very hydrophobic, this initial collapse must contain residues that are not local in sequence space. This is consistent with the previous assumption that the early folding event of IFABP involves hydrophobic collapse surrounding the hydrophobic core containing Trp-82, Phe-47, Phe-62, Leu-64, Phe-68, Met-84, and Leu-89 (*9*, *10*, *15*). Although this core in the I_1 intermediate is not as hydrophobic as that in the native protein, as is evident by the position of the fluorescence maximum, it provides a nucleation center for the propagation of the β-strands during the subsequent kinetic phases. A similar nascent phase involving hydrophobic collapse has

also been proposed for CRABP I (*13*, *19*, *20*), as well as many other protein systems (*16*, *22*−*29*). Thus, hydrophobic collapse may be a universal mechanism by which the unfolded polypeptide reduces its conformational space such that the subsequent folding reaction can be achieved efficiently and precisely (*24*).

In the second phase of folding ($I_1 \rightarrow I_2$), the emission maximum shifts further toward blue from 355 to 345 nm and is associated with an additional increase in intensity. The rate of this reaction is 1500 s^{-1}, and it is insensitive to the final urea concentration up to 2.5 M as shown in Figure 4. The extended blue shift in the peak maximum during this phase suggests that the hydrophobic residues surrounding Trp-82 rearrange such that the packing is tighter and more native-like. The product of this reaction, I_2, has the same fluorescence spectroscopic properties as the burst phase intermediate observed in a much slower stopped flow apparatus. This intermediate has considerable β-sheet secondary structure by CD criteria (*11*), although NMR amide protection studies suggest that this state does not contain a stable H-bonding network (I. J. Ropson, unpublished results). Thus, we postulate that during this kinetic phase the hydrophobic core becomes more compact and native-like, and β-strands B−G (defined in Figure 1) propagate from this hydrophobic core outward toward the opposite end of the β-sheet as illustrated in Figure 5. Due to weak side chain interactions and the likelihood that considerable amounts of water are still trapped within the structure, the hydrogen bonds between the established β-strands are only loosely formed and are not protected from amide exchange. The rearrangement of the hydrophobic core is the rate-limiting step during this folding phase; its kinetics are thus not sensitive to the final urea concentration as shown in Figure 4.

During the last phase of folding ($I_2 \rightarrow F$), the fluorescence peak maximum is unchanged but the intensity is gradually increased with a rate constant of $\sim$5 s^{-1}. The kinetics occurring during this phase are extremely sensitive to the mutations in the glycines (Gly-99, Gly-110, and Gly-121) located in the last three β-turns on the bottom half of the clam shell (*4*). In contrast, mutations in any of the three glycines (Gly-44, -65, and -86) located near the hydrophobic core do not affect the folding kinetics (*4*). This observation suggests that the topology of β-strands B−G connected by Gly-44, -65, and -86 is already established in intermediate I_2. The kinetics during this phase are also insensitive to the mutations in the other two turn residues, Gly-75 and Asn-54, that connect β-strands C and D and β-strands E and F at the opposite side of the β-sheet. The establishment of β-strands B−G in intermediate I_2 is consistent with the proposal that the $I_1 \rightarrow I_2$ transition is associated with the propagation of β-strands B−G from the hydrophobic core. These results suggest that the structure of β-strands B−G, which make up most of the top half of the clam shell and part of the bottom shell, serves as a template for the formation of the last three strands (H−J) on the bottom half of the shell. Furthermore, the formation of the native structure must rely on the coordinated interaction of residues from both halves of the structure. This is consistent with the observation that peptides of IFABP with only the top or bottom half of the clam shell do not form β-sheet structure (I. J. Ropson, unpublished results). On the basis of these results, we propose that the last phase of the folding reaction involves the formation of the last three β-strands on the bottom half of the clam shell. The rate-limiting step during this phase is the formation of the correct β-turns connecting these β-strands.

In conclusion, we postulate that the folding reaction of IFABP is hierarchical as illustrated in Figure 5. The unfolded polypeptide first collapses into a semicompact structure surrounding a hydrophobic core made up of Phe-47, Phe-62, Leu-64, Phe-68, Trp-82, Met-84, and Leu-89. This dramatically reduces the conformational space that needs to be searched such that the proceeding folding reactions can be achieved with efficiency and fidelity. In the subsequent step, β-strands B−G propagate out of the hydrophobic core, analogous to zippering in interlacing zippers. These zippers are not tightly matched due to side chain interactions, but they establish the native topology on the top half of the clam shell as well as part of the bottom shell. During the last phase of the folding, this native topology serves as a template for the folding of the last three β-strands in the bottom half of the clam shell, and at the same time, the native hydrogen bonding network is established through energy minimization of the side chain interactions.

REFERENCES

1. Capaldi, A. P., and Radford, S. E. (1998) *Curr. Opin. Struct. Biol. 8*, 86−92.
2. Carlsson, U., and Jonsson, B. H. (1995) *Curr. Opin. Struct. Biol. 5*, 482−487.
3. Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) *Adv. Protein Chem. 37*, 1−109.
4. Kim, K., and Frieden, C. (1998) *Protein Sci. 7*, 1821−1828.
5. Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) *Nature 390*, 196−199.
6. Banaszak, L., Winter, N., Xu, Z., Bernlohr, D. A., Cowan, S., and Jones, T. A. (1994) *Adv. Protein Chem. 45*, 89−151.
7. Scapin, G., Gordon, J. I., and Sacchettini, J. C. (1992) *J. Biol. Chem. 267*, 4253−4269.
8. Ropson, I. J., Gordon, J. I., and Frieden, C. (1990) *Biochemistry 29*, 9591−9599.
9. Ropson, I. J., and Dalessio, P. M. (1997) *Biochemistry 36*, 8594−8601.
10. Dalessio, P. M., and Ropson, I. J. (2000) *Biochemistry 39*, 860−871.
11. Burns, L. L., Dalessio, P. M., and Ropson, I. J. (1998) *Proteins: Struct., Funct., Genet. 33*, 107−118.
12. Burns, L. L., and Ropson, I. J. (2001) *Proteins: Struct., Funct., Genet.* (in press).
13. Clark, P. L., Weston, B. F., and Gierasch, L. M. (1998) *Folding Des. 3*, 401−412.
14. Kim, K., Ramanathan, R., and Frieden, C. (1997) *Protein Sci. 6*, 364−372.
15. Ropson, I. J., and Frieden, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7222−7226.
16. . Takahashi, S., Yeh, S.-R., Das, T. K., Chan, C. K., Gottfried, D. S., and Rousseau, D. L. (1997) *Nat. Struct. Biol. 4*, 44−50.
17. Baldwin, R. L. (1996) *Folding Des. 1*, R1−R8.
18. Sosnick, T. R., Shtilerman, M. D., Mayne, L., and Englander, S. W. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 8545−8550.
19. Clark, P. L., Liu, Z.-P., Zhang, J. A., and Gierasch, L. M. (1996) *Protein Sci. 5*, 1108−1117.
20. Clark, P. L., Liu, Z.-P., Rizo, J., and Gierasch, L. M. (1997) *Nat. Struct. Biol. 4*, 883−886.
21. Callis, P. R. (1997) *Methods Enzymol. 278*, 113−150.

22. Yeh, S.-R., Takahashi, S., Fan, B., and Rousseau, D. L. (1997) *Nat. Struct. Biol. 4*, 51−56.
23. Yeh, S.-R., and Rousseau, D. L. (1998) *Nat. Struct. Biol. 5*, 222−228.
24. Yeh, S.-R., Han, S., and Rousseau, D. L. (1998) *Acc. Chem. Res. 31*, 727−736.
25. Agashe, V. R., Shastry, M. C. R., and Udgaonkar, J. B. (1995) *Nature 377*, 754−757.
26. Miranker, A., and Dobson, C. M. (1996) *Curr. Opin. Struct. Biol. 6*, 31−42.
27. Forge, V., Wijesinha, R. T., Balbach, J., Brew, K., Robinson, C. V., Redfield, C., and Dobson, C. M. (1999) *J. Mol. Biol. 288*, 673−688.
28. Gilmanshin, R., Callender, R. H., and Dyer, R. B. (1998) *Nat. Struct. Biol. 5*, 363−365.
29. Hagen, S. J., and Eaton, W. A. (2000) *J. Mol. Biol. 297*, 781−789.

BI0155044