# Swapping Core Residues in Homologous Proteins Swaps Folding Mechanism†

Paula M. Dalessio, Joshua A. Boyer, Jessica L. McGettigan, and Ira J. Ropson*

*Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033*

*Received August 31, 2004; Revised Manuscript Received December 10, 2004*

ABSTRACT: Rat intestinal fatty acid binding protein (IFABP) displays an intermediate with little if any secondary structure during unfolding, while the structurally homologous rat ileal lipid binding protein (ILBP) displays an intermediate during unfolding with nativelike secondary structure. Double-jump experiments indicate that these intermediates are on the folding path for each protein. To test the hypothesis that differences in the number of buried hydrophobic atoms in a folding initiating site are responsible for the different types of intermediates observed for these proteins, two mutations (F68C-IFABP and C69F-ILBP) were made that swapped a more hydrophobic residue for a more hydrophilic residue in the respective cores of these two proteins. F68C-IFABP followed an unfolding path identical to that of WT-ILBP with an intermediate that showed nativelike secondary structure, whereas C69F-ILBP followed an unfolding path that was identical to that of WT-IFABP with an intermediate that lacked secondary structure. Further, a hydrophilic residue was introduced at an identical hydrophobic structural position in both proteins (F93S-IFABP and F94S-ILBP). Replacement of phenylalanine with serine at this site led to the appearance of an intermediate during refolding that lacked secondary structure for both proteins that was not detected for either parental protein. Altering the chemical characteristics and/or size of residues within an initiating core of hydrophobic interactions is critical to the types of intermediates that are observed during the folding of these proteins.

One of the major questions in protein folding is how proteins with very different amino acid sequences achieve similar final structures. For proteins smaller than 100 residues that follow a two-state folding mechanism, it appears that the nature of the transition state is well conserved for proteins with similar final structures, even in cases of low sequence homology (reviewed in ref *1*). However, the number and types of intermediates observed for the folding of proteins greater than 100 residues varies considerably, even for proteins with high structural similarities (*1*). One such family of proteins with apparently divergent folding mechanisms is the intracellular lipid binding proteins.

These predominantly $\beta$-sheet proteins have very similar structures despite considerable sequence heterogeneity (*2*, *3*). For example, the structures of rat ileal lipid binding protein (ILBP)[1] and rat intestinal fatty acid binding protein (IFABP) are shown in Figure 1. These particular proteins share 23% sequence identity and 39% sequence similarity (*4*). However, the structures of rat IFABP (*5*) and pig ILBP (*6*) show a 1.98-Å RMSD for 97% of the backbone atoms (*4*), similar to the differences observed for the structures of IFABP determined by NMR (*7*) or X-ray crystallography (*5*). The only major structural differences between the various family members are due to insertions and deletions in turns (Figure 1).

Studies of the folding of these proteins have shown that two different types of intermediates are observed for different family members (*4*, *8*, *9*). In the case of IFABP, an intermediate with little if any secondary structure is found (*10*, *11*), whereas ILBP displays an intermediate with nativelike secondary structure (*4*). Two other members of the family, cellular retinol binding protein II (*8*, *9*) and heart fatty acid binding protein (data not shown) have intermediates similar to that of IFABP, while two other proteins, cellular retinoic acid binding protein I and cellular retinoic acid binding protein II, have intermediates similar to that of ILBP (*8*, *9*, *12*).

We have proposed that differences in the number of buried hydrophobic atoms in a putative initiating site for folding are responsible for the types of intermediates observed in these proteins (*4*). Proteins with larger numbers of hydrophobic atoms in this vicinity display intermediates with little if any secondary structure. Conversely, proteins with hydrophilic substitutions in this core have intermediates with nativelike secondary structure. To test this hypothesis, two mutations were made (F68C-IFABP and C69F-ILBP) that essentially swapped more hydrophobic and more hydrophilic residues at structurally equivalent positions in these two proteins. As a further test of this hypothesis, a hydrophilic replacement was introduced at an identical hydrophobic structural position in both proteins (F93S-IFABP and F94S-ILBP).

† This research was supported by NIH Grant GM-57906 to I.J.R.

* To whom correspondence should be addressed. 500 University Dr., Hershey, PA 17033. Tel.: 717-531-4064. Fax: 717-531-7072. E-mail: iropson@psu.edu.

[1] Abbreviations: IFABP, rat intestinal fatty acid binding protein; ILBP, rat ileal lipid binding protein; F68C-IFABP, F93S-IFABP, mutants of rat IFABP.; C69F-ILBP, F94S-ILBP mutants of rat ILBP; CD, circular dichroism.

10.1021/bi048125u CCC: $30.25 
Published on Web 02/03/2005

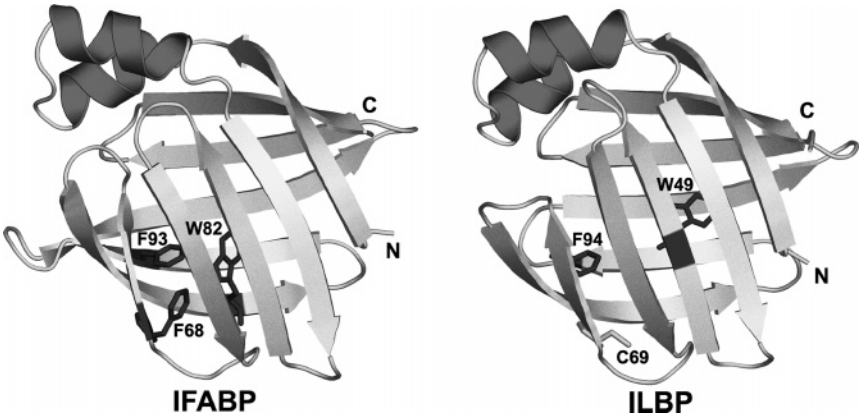


FIGURE 1: Ribbon drawing showing the backbone tracing and the location of the tryptophans and mutated residues for IFABP (PDB ID 1ifc) and ILBP (based on PDB IDs 1o1u and 1eal).

## MATERIALS AND METHODS

*Protein Source and Purification.* Expression plasmids for IFABP and ILBP have been previously described (*13*, *14*). Mutant expression vectors for rat F68C-IFABP, F93S-IFABP, C69F-ILBP, and F94S-ILBP were constructed using the QuikChange kit (Stratagene). The plasmid DNA for each construct was isolated using a QIA mini-prep kit (Qiagen) and sent to the macromolecular core facility (Penn State College of Medicine) for sequencing to verify the presence of the specified mutation and the lack of any other mutations in the gene sequence. Expression was induced in *E. coli* strain MG1655.

The proteins were extracted from *E. coli* using the freeze−thaw method (*15*). Following lysis of the cells, IFABP, ILBP, F68C-IFABP, and C69F-ILBP were purified from the cell supernatant as previously described (*4*, *13*). F93S-IFABP and F94S-ILBP were isolated as inclusion bodies and purified as described by Kim and Frieden (*16*). During the purification of ILBP, F68C-IFABP, and F94S-ILBP, 0.1 mM DTT was added to all buffers to prevent cysteine oxidation. Protein purity was demonstrated by the presence of a single band on a 20% SDS polyacrylamide gel. The following extinction coefficients at 280 nm were calculated for the various proteins: 1.13 $mg^{-1}$ $cm^{-1}$ for IFABP and its mutant proteins and 0.87 $mg^{-1}$ $cm^{-1}$ for ILBP and its mutant proteins (*17*).

*Reagents.* Urea stock solutions (approximately 10 M, ultrapure, ICN) were prepared and stored at −20 °C as previously described (*11*). A working solution of 9 M urea was prepared on the day of an experiment by adding the buffer components to the freshly thawed urea stock solution. All buffers contained 25 mM $NaPO_4$, 75 mM NaCl, and 0.1 mM EDTA. The pH values of working buffer and urea solutions were adjusted to pH 7 (IFABP, F68C-IFABP, F93S-IFABP) or to pH 8 (ILBP, C69F-ILBP, F94S-ILBP). Just before use, 0.1 mM DTT was added to solutions used for ILBP, F94S-ILBP, and F68C-IFABP. Exact denaturant concentrations were determined by refractive index measurements with a Milton Roy Abbe-3 refractometer at 25 °C (*18*). All experimental buffers were filtered through a 0.2 mm Whatman nylon membrane. All chemicals were reagent grade unless otherwise denoted. There is no difference in the folding mechanism at pH 7 and pH 8 for any of these proteins at concentrations below 0.5 mg/mL. However, at pH 7, the higher concentrations of protein needed for the stopped-flow CD and spectral studies described below caused a detectable aggregation for ILBP and mutants of ILBP. As such, all experiments on the ILBP derivatives were done at pH 8.

*Equilibrium Studies.* Unfolding transitions were monitored as a function of denaturant concentration by CD and fluorescence. A Jasco J-710 spectropolarimeter was employed to follow changes in secondary structure in the far-UV portion of the CD spectrum using a thermostated 0.1 mm cell at 25 °C. For F68C-IFABP, fluorescence changes were detected with an Aminco-Bowman Series 2 luminescence spectrometer with excitation at 290 nm (2-nm band-pass) and emission spectra were collected from 300 to 400 nm (8-nm band-pass) in a 1-cm path length cell at 25 °C. For the remaining proteins, fluorescence studies were carried out with a PTI QuantaMaster luminescence spectrometer with excitation at 290 nm (2-nm band-pass), and emission spectra were collected from 300 to 400 nm (4-nm band-pass) in a 0.3-cm path length cell at 25 °C.

Approximately 1 h before the start of an experiment, an identical amount of protein (0.15 mg/mL, final concentration) was added to the working buffer and 9 M urea solutions. Individual samples ranging in denaturant concentrations from 0 to 8.5 M urea were prepared by mixing specific volumes of protein in buffer and protein in 9 M urea using a Hamilton Microlab 40C titrator (*19*). The data were corrected for the background signal of the buffer and urea solutions. For purposes of comparison of denaturation curves from all of the proteins by all methods, the data was normalized from 0 to 1, where 0 and 1 were assigned to the lowest and highest observed intensity, respectively (*10*).

Nonlinear least-squares fit to the equilibrium data were generated using KaleidaGraph (Synergy Software) in conjunction with an equation adapted from Santoro and Bolen (*20*) as previously described (*10*). All fits were to a minimum of two independent data sets. The parameters and errors shown in Table 1 resulted from simultaneous fits to multiple data sets. The figures show one data set for each protein for ease of viewing.

*Integrated Stopped-Flow Fluorescence Kinetic Studies.* The kinetics of unfolding and refolding were monitored using an Applied Photophysics DX-17MV stopped-flow spectrophotometer. Excitation was at 290 nm (2-nm band-pass) using a 0.2-cm path length cell at 25 °C. Emission intensity was monitored above 305 nm at 90° through a WG305 Schott glass filter (Oriel). Two drive syringes (2.5 mL and 0.5 mL) were used to mix five parts of denaturant solution with one

Table 1: Summary of Fluorescence and Stopped-Flow Fluorescence Data

| protein | $\Delta G_{H_2O}$ (kcal mol$^{-1}$) | $^{m}$G (kcal mol$^{-1}$M$^{-1}$) | midpoint (M) |
|---|---|---|---|
| WT-IFABP | | | |
| urea ($F_{340}$) | 4.79 ± 0.22 | −1.13 ± 0.07 | 4.21 ± 0.04 |
| urea (FL-ST-FL) | 4.76 ± 0.27 | −1.19 ± 0.05 | 4.05 ± 0.10 |
| F68C-IFABP | | | |
| urea ($F_{340}$) | 4.84 ± 0.29 | −1.24 ± 0.07 | 3.90 + 0.03 |
| urea (FL-ST-FL) | 4.59 ± 0.62 | −1.17 ± 0.15 | 3.93 ± 0.07 |
| F93S-IFABP | | | |
| urea ($F_{340}$) | 3.40 ± 0.22 | −1.87 ± 0.10 | 1.82 ± 0.02 |
| urea (FL-ST-FL) | 2.80 ± 0.36 | −1.60 ± 0.17 | 1.74 ± 0.05 |
| WT-ILBP | | | |
| urea ($F_{340}$) | 4.84 ± 0.36 | −1.61 ± 0.12 | 3.01 ± 0.03 |
| urea (FL-ST-FL) | 5.49 ± 0.60 | −1.83 ± 0.20 | 3.00 ± 0.05 |
| C69F-ILBP | | | |
| urea ($F_{340}$) | 4.01 ± 0.30 | −1.80 ± 0.12 | 2.23 ± 0.03 |
| urea (FL-ST-FL) | 4.20 ± 0.53 | −1.82 ± 0.20 | 2.31 ± 0.06 |
| F94S-ILBP | | | |
| urea ($F_{340}$) | 4.00 ± 0.33 | −1.79 ± 0.06 | 2.24 ± 0.02 |
| urea (FL-ST-FL) | 3.99 ± 0.68 | −1.75 ± 0.26 | 2.27 ± 0.08 |

part of protein solution (0.26 mg/mL, final concentration). Five−seven kinetic traces were averaged for each denaturant concentration. The dead time of this instrument is 5−10 ms, depending on the final denaturant concentration (*11*). Data collected during the dead time were not included in the analysis.

The nonlinear least-squares fit of the integrated fluorescence kinetic data to monophasic, monophasic plus steady state, biphasic, and triphasic models was accomplished with the program supplied by Applied Photophysics. The goodness of the fit to the various models was determined using published criteria (*21*, *22*).

*Wavelength-Dependent Stopped-Flow Fluorescence Studies*. To study time-dependent fluorescence spectral changes during refolding, the WG305 cutoff filter was replaced with a high-efficiency monochromator in the Applied Photophysics DX-17MV stopped-flow instrument. Kinetic time courses (2000 data points) depicting the change in intensity during refolding were collected for each 10-nm interval (9-nm band-pass) over a wavelength range of 310−400 nm. The excitation wavelength and band-pass were 290 and 9 nm, respectively. As described above, 5 parts of denaturant was mixed with one part of protein solution in buffer. The final urea concentrations were 0.8 and 1.3 M, and the final protein concentration was 0.52 mg/mL.

The spectral data sets were analyzed using the GLINT software package provided by Applied Photophysics. This program allows the entry of models for potential reaction mechanisms and globally determines the kinetic rates and amplitudes that best describe the transition at each wavelength by nonlinear least-squares regression. The program fits the spectra and concentration of the reactants, intermediates, and products that participate in the reaction. The fit parameters and residual plots at all wavelengths and times were examined to assess the accuracy of the model (*21*, *22*).

*CD Kinetic Studies.* A Jasco J-710 spectropolarimeter in conjunction with a RX1000 stopped-flow apparatus (Applied Photophysics) was used to monitor the kinetics of unfolding and refolding of all of the proteins at 218 nm (2-nm band-pass) in a 0.1-cm cell at 25 °C. For all experiments, five parts of denaturant were mixed with one part of protein for a final protein concentration of 0.52 mg/mL. Five−seven transitions were averaged at each denaturant concentration. Data points collected during the dead time of the instrument (15−30 ms) were not included in the analysis (*11*). Fitting of the data was performed as described above for stopped-flow fluorescence studies.

*Structure Modeling*. The Swiss protein data bank (PDB) server (*23*) was used to create a structural model for rat ILBP based on the NMR structures of pig ILBP (1eal, *6*) and human ILBP (1o1u, *24*). Rat ILBP is 69.3 and 76.6% identical and 82.7 and 86.7% similar to pig ILBP and human ILBP, respectively. Mutagenesis and structural alignments were performed in silico using the Swiss PDB viewer program (*23*). For each mutation, all of the available side chain rotamers were evaluated for their compatibility with the overall structure, and the rotamer configurations most similar to those in the known structures were used. The ASC program (*25*) was used to determine the solvent exposure of the side chains.

## RESULTS

*Spectral Comparison.* The far-UV CD spectra of the native state of all of the wild-type and mutant proteins were very similar, reflecting the extensive $\beta$-sheet structure of these proteins (data not shown). The CD spectra of the unfolded proteins were identical, and these mutations did not cause large changes in secondary structure propensities.

Differences in the native-state fluorescence spectra were observed for IFABP and ILBP and their respective mutants. Both F68 and F93 are within 4 Å of W82 in the native structure of IFABP (Figure 1). This tryptophan is responsible for most of the fluorescence signal in this protein (*26*,*27*). The increased polarity of these substitutions (F68C-IFABP, F93S-IFABP) near W82 caused a red-shift in the native-state fluorescence of these proteins (maximum at 337−338 nm) compared to the wild-type protein (maximum at 330 nm). Although the fluorescence intensities of IFABP and F93S-IFABP were similar, the F68C mutation reduced the intensity by about 30%, suggesting that the presence of the cysteine caused some quenching of the fluorescence of W82 in this mutant protein.

Mutations at the structurally equivalent sites in ILBP (C69F-ILBP and F94S-ILBP) are more than 10 Å from W49, the only tryptophan in ILBP (Figure 1). As such, these amino acid substitutions had minimal effects on the maximal wavelength of emission of these native proteins. However, native F94S-ILBP had roughly twice the fluorescence intensity of either ILBP or C69F-ILBP. Since the CD spectra of the native proteins were identical and the wavelength of maximal fluorescence did not change, relatively minor structural changes must have occurred in this protein that changed the quenching behavior of residues close to W49 in the structure without significantly changing the backbone conformation or the overall hydrophobicity of the W49 environment.

The wavelength of maximal fluorescence for all of the denatured wild-type and mutant proteins in 8 M urea was identical to *N*-acetyl-tryptophanamide in buffer. The intensity of fluorescence was proportional to the number of tryptophans in each protein.

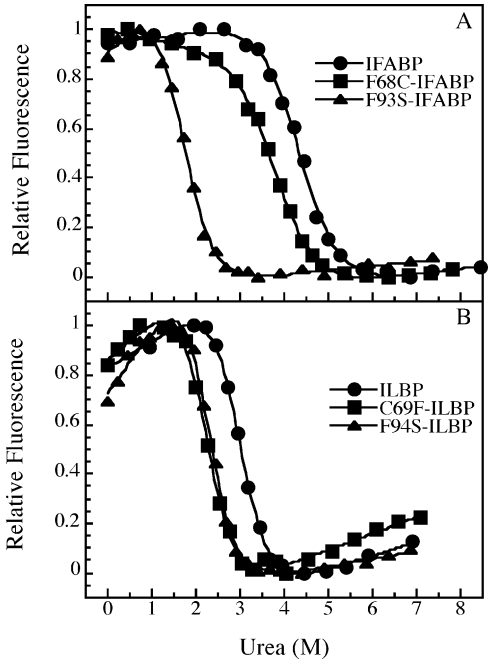


FIGURE 2: The urea denaturation of WT-IFABP, F68C-IFABP, and F93S-IFABP (A) and WT-ILBP, C69F-ILBP, and F94S-ILBP (B) were monitored by fluorescence at 340 nm.

*Equilibrium Studies.* Equilibrium unfolding transitions are shown in Figure 2. Folding was completely reversible and independent of protein concentration for all proteins. Each transition fit well to a two-state model for unfolding for every protein, showing that significant concentrations of only the native and unfolded states were present at equilibrium. The parameters for these fits are shown in Table 1.

The F68C mutation modestly decreased the stability of IFABP, while the F93S mutation produced a more drastic decrease in stability. The dependence of the free energy of the transition on denaturant concentration ($m_G$) was unusually low for IFABP ($1.13 \pm 0.07$) and F68C-IFABP ($1.24 \pm 0.07$) compared to other proteins of this size (*28*) or in this family (*4*, *8*, *9*, Table 1). Although the cause of this behavior is not known, this discrepancy was absent for the F93S mutation ($m_G$ of $1.87 \pm 0.10$), suggesting a role for this site in this phenomenon. It was not surprising that these mutations were destabilizing, since large hydrophobic amino acids were replaced with smaller more hydrophilic ones.

Mutations at the same sites in the ILBP background modestly decreased the stability of the protein. In the case of C69F-ILBP, a smaller, more hydrophilic residue was replaced by a larger more hydrophobic one. The other mutation, F94S, placed a hydrophilic residue in a hydrophobic pocket and thus destabilized the protein, although the cost of this mutation in ILBP was significantly less than at the structurally equivalent position in IFABP.

*Unfolding of WT-IFABP and Mutant IFABPs.* WT-IFABP displays two phases during unfolding (*4*, *8*, *10*, *11*). The first phase shows about 75% of the total expected amplitude change by stopped-flow fluorescence and all of the expected amplitude change by stopped-flow CD, followed by the remaining fluorescence change. The dependence of these rates on denaturant concentration is shown in Figure 3. The simplest model that fits the data requires an intermediate with little if any secondary structure. The fluorescence spectrum

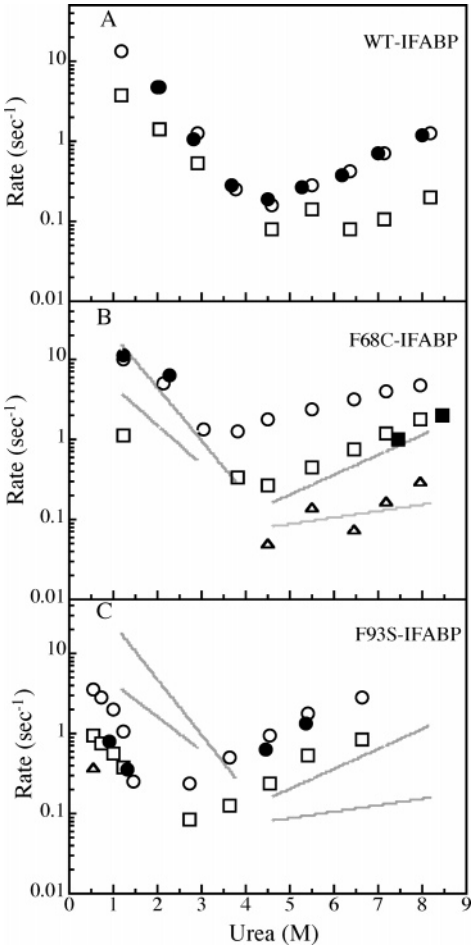


FIGURE 3: The folding and unfolding rates of WT-IFABP (A), F68C-IFABP (B), and F93S-IFABP (C) were monitored by stopped-flow fluorescence (open symbols) and stopped-flow CD (closed symbols). The folding and unfolding rates for the wild-type protein are indicated by gray lines on panels B and C.

of the intermediate state has a wavelength maximum similar to that of the unfolded state (*11*), suggesting considerable solvent exposure of the tryptophans in this intermediate. Mutant IFABP proteins containing only a single tryptophan have shown that W82 participates in this intermediate, whereas W6 does not (*27*).

Unlike WT-IFABP, the F68C-IFABP mutant protein displayed three phases during unfolding by stopped-flow fluorescence (Figure 3B). The fastest, second, and slowest unfolding phases accounted for approximately 20, 75, and 5% of the total expected amplitude change, respectively. Only one rate was observed by stopped-flow CD, which was coincident with the second rate observed by stopped-flow fluorescence and accounted for the entire expected amplitude change. As such, the first observed unfolding intermediate had the spectral properties of a molten globule, followed by a second intermediate with spectral properties more similar to that observed during the unfolding of IFABP. Therefore, the unfolding path of F68C-IFABP had become similar to that of ILBP (described below). The unfolding of the F93S-IFABP mutant was identical to that of WT-IFABP (Figure 3C) in both the number and relative amplitudes of the observed phases, suggesting that this hydrophilic substitution in the core had no effect on the unfolding path.

*Unfolding of WT-ILBP and Mutant ILBPs.* Previously published data has shown that WT-ILBP also displays two

phases during unfolding (*4*). However, unlike IFABP, the first phase displays 15−20% of the expected amplitude change by fluorescence, with the second slower phase showing the remainder of the fluorescence change and all of the expected amplitude change by CD. The fluorescence spectrum of the intermediate state has a wavelength maximum like that of native state, suggesting that the single tryptophan of ILBP remains buried in this intermediate state (*4*). This intermediate has the characteristics of a molten globule, with nativelike secondary structure but modified tertiary interactions (*29*). As such, the spectral properties and the structure of the intermediate observed during the unfolding of ILBP is very different from that observed for IFABP (*4*). Since that report, our stopped-flow fluorescence instrumentation has been upgraded, and a third slower small amplitude phase is described here for the unfolding of WT-ILBP (approximately 5% of the total amplitude change, Figure 4A). Thus, during the unfolding of ILBP two intermediates were observed, one molten-globule-like and one with spectral characteristics similar to that observed during the unfolding of IFABP.

Unlike WT-ILBP, C69F-ILBP displayed only two phases during unfolding by stopped-flow fluorescence (Figure 4B). Approximately 90% of the expected amplitude was associated with the first phase with the remaining amplitude in the second phase. Only one phase was observed by stopped-flow CD, which accounted for the entire expected amplitude and was coincident with the first phase observed by stopped-flow fluorescence. The fastest phase, which characterized the transition from the native state to the molten-globule state for WT-ILBP, was missing in this mutant protein. Hence, C69F-ILBP followed an unfolding path that was very similar to that of IFABP, showing only one intermediate with little if any secondary structure. The unfolding of F94S-ILBP was identical to that of ILBP, in both the number and relative amplitude change of each phase during the transition (Figure 4C).

*Refolding of WT-IFABP and Mutant IFABPs.* The folding of IFABP is complex, especially for a small single domain protein. Initial experiments showed a burst phase by both CD and fluorescence whose amplitude increases at lower final denaturant concentration, followed by an observed phase that results in native spectral properties (*11*) and tight ligand binding (*30*). This burst phase has been resolved into two distinct parts using a continuous-flow instrument, one with a rate > 10 000 $s^{-1}$ (faster than the deadtime of that instrument) and an observable rate around 1 500 $s^{-1}$ (*31*). Stopped-flow double-jump experiments have shown that the product of the burst phase reaction is not stable, because it unfolds in the deadtime of these experiments (*4*). The double-jump experiments also show that an intermediate with fluorescence spectral properties similar to those observed during the unfolding of IFABP (wavelength maxima similar to unfolded protein) accumulates early in the folding process, although it cannot be detected directly due to the masking effects of the burst phase. Recent experiments with improved instrumentation have added another slower observed refolding rate by fluorescence, which accounts for about 5% of the total expected amplitude change (*27*). Mutant IFABP proteins containing only a single tryptophan have shown that W6 participates in this transition, whereas W82 has achieved its native conformation prior to this phase (*27*). There are

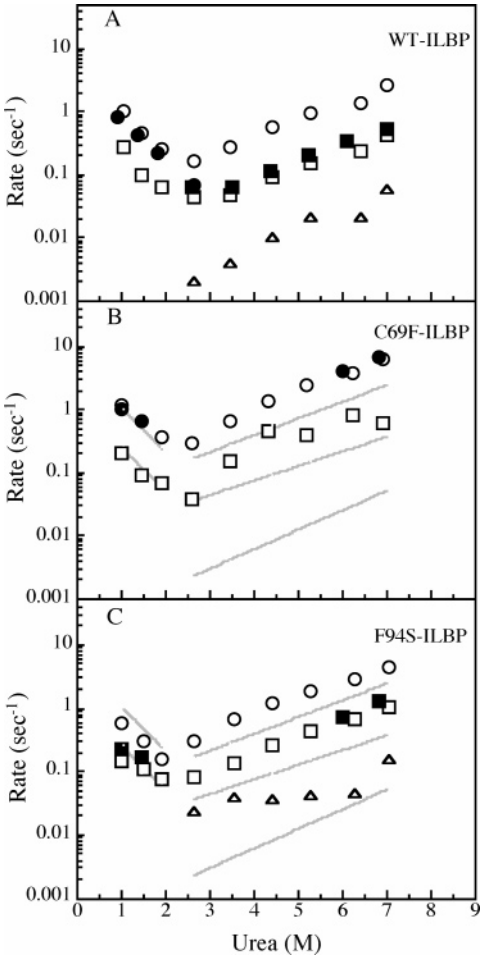


FIGURE 4: The folding and unfolding rates of WT-ILBP (A), C69F-ILBP (B), and F94S-ILBP (C) were monitored by stopped-flow fluorescence (open symbols) and stopped-flow CD (closed symbols). The folding and unfolding rates for the wild-type protein are indicated by lines on panels B and C.

no prolines in IFABP, eliminating proline isomerization as a cause for this small amplitude slow phase. Since ligand binding is associated with the faster of the two observed fluorescence phases (*30*) and double-jump experiments indicate that the product of the faster phase has nativelike spectral properties and stability (*4*), this final folding phase is probably a minor tertiary structure rearrangement at the periphery of the structure that locally effects the environment around W6 (*27*).

The F68C-IFABP mutant showed refolding behavior similar to that of WT-IFABP, including a burst phase by both CD and fluorescence, a fast phase that accounted for all of the remaining expected CD intensity and most of the remaining expected fluorescence intensity. Unlike WT-IFABP, the slowest refolding phase was only observed at the lowest concentration of denaturant (Figure 3). Although this mutation had significant effects on protein stability and the unfolding mechanism, few differences were observed during refolding.

Unexpectedly, the F93S mutation altered the refolding mechanism significantly. Unlike WT-IFABP, little if any burst phase fluorescence intensity change was observed (Figure 5). However, these experiments measured the total fluorescence intensity above 305 nm and could not determine if a change in the wavelength of maximal fluorescence had occurred. Wavelength-dependent stopped-flow kinetic studies

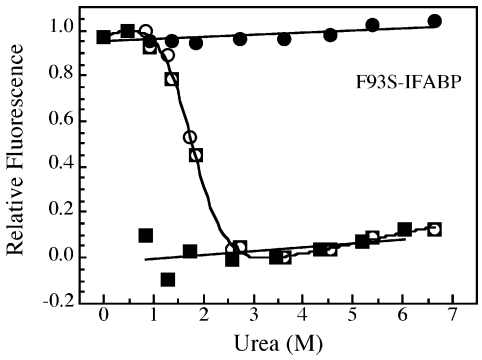


FIGURE 5: The fluorescence equilibrium endpoints obtained from folding (○) and unfolding (□) stopped-flow experiments, illustrating the lack of a refolding burst phase for F93S-IFABP. Closed symbols are the initial fluorescence signal at the start of the kinetic transition for folding (●) and unfolding (■). Solid lines indicated the expected initial fluorescence based on the extrapolated native and denatured baselines.

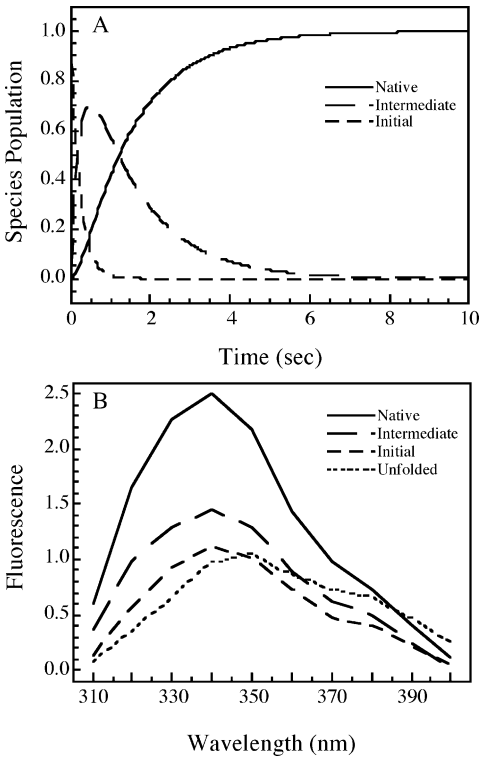


FIGURE 6: (A) The time-dependent concentration of the native, intermediate, and unfolded states from the fit of the data collected from wavelength-dependent stopped-flow fluorescence to the model U → I → N for F93S-IFABP. (B) The fluorescence spectra of the native, intermediate, and initial state of F93S-IFABP. The unfolded spectrum of F93S-IFABP in 8 M urea is also shown. The final urea concentration is 0.84 M and the final protein concentration is 0.52 mg/mL.

were performed on this protein in order to determine if the burst phase was still present. The emission fluorescence intensities were measured at individual wavelengths and were globally fit to a three-state sequential model. The data fit well to this model, showing a prominent lag phase in the population of the native state (Figure 6A). However, the spectra of the initial observed species was not identical to that expected for the unfolded state (Figure 6B). The wavelength of maximal emission was shifted from the 352 nm expected for the unfolded state to 340 nm for the first observed species after the dead time of mixing. There was little if any overall intensity change associated with this wavelength shift. As such, the stopped-flow integrated fluorescence experiments failed to detect a burst phase, even though one was present. F93S-IFABP also differed from WT-IFABP in that three phases were observed during refolding. The fastest phase was observed only by fluorescence, represented an increase in intensity without any change in secondary structure or the wavelength of maximal fluorescence (Figure 6), and was not observed for WT-IFABP (Figure 3). Given the large differences in the amplitude of the burst phase for F93S-IFABP compared to WT-IFABP, this phase may correspond to the hidden phase of WT-IFABP that could only be observed in double-jump experiments. The second phase showed all of the expected amplitude change by CD and most of the remaining fluorescence change. This phase corresponds to the first phase observed for WT-IFABP. Finally, but only at the lowest final concentration of denaturant (<0.6 M urea), a slower small amplitude fluorescence change was observed, similar to that observed for WT-IFABP. The refolding behavior of F93S-IFABP was more similar to that observed for WT-IFABP at pH 10 (*32*) rather than that of WT-IFABP at pH 7. In summary, the refolding of this mutant protein had a folding pathway that included a burst phase event, an intermediate with little if any secondary structure, and a molten-globule-like intermediate.

*Refolding of WT-ILBP and Mutant ILBPs.* The stopped-flow refolding of ILBP is described by two phases (*4*). The first phase accounts for 80% of the fluorescence signal change and all of the CD signal change, followed by a slower phase that was observed only by fluorescence. Unlike IFABP, no significant burst phase is observed at any final concentration of denaturant (*4*). Double-jump experiments indicate that the intermediate observed during refolding has similar stability and spectral properties to those observed for the major unfolding intermediate and is likely to be on the reversible path for structure formation in this protein (*4*). As such, the refolding path of ILBP has a single intermediate with molten-globule-like properties.

The refolding behavior of C69F-ILBP was identical to that of WT-ILBP (Figure 4). As such, although significant differences were observed during the unfolding of these two proteins, the refolding reaction was unaffected by this mutation.

Overall, the refolding of F94S-ILBP was very different from WT-ILBP (Figure 4). Similar to WT-ILBP, there is no burst phase during refolding. However, unlike WT-ILBP, the first observed phase had a small amplitude and was only detected by fluorescence. The second phase was observed by both fluorescence and CD. As such, the intermediate observed during the folding of F94S-ILBP was not a molten globule, since little if any secondary structure was detected.

## DISCUSSION

This study tests the hypothesis that altering the chemistry of critical residues at specific structural locations can change the number and types of intermediates observed during the folding of proteins in this family. The importance of specific residues to the mechanism of folding of structurally related proteins has been addressed in a few experimental systems. Two proteins in the cks family appear to fold via different mechanisms due to differences in the number and characteristics of tertiary contacts in the core of the protein (*33*).

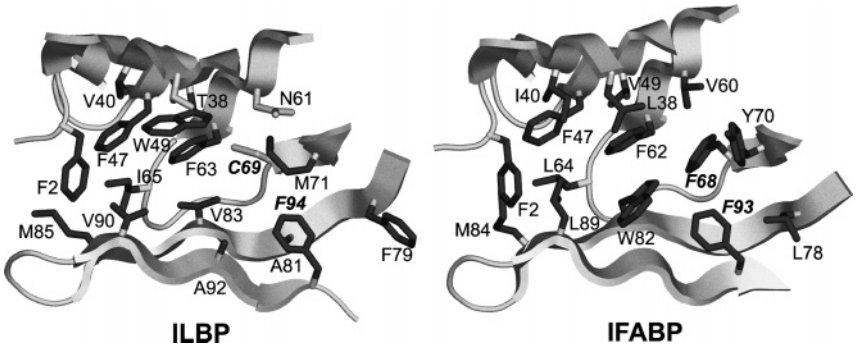


FIGURE 7: Hydrophobic clusters of IFABP and ILBP. Residues with side chains within 7.5 Å of W82 in IFABP and V83 in ILBP are shown and labeled. The view is looking between the sheets with the α-helices projecting toward the viewer (see Figure 1). The helices and parts of the sheets have been removed so that the core can be viewed.

An intermediate is observed in the folding of Im9, a four-helix bundle protein, whereas the homologous Im7 protein does not exhibit any intermediates during folding (*34*). Despite these apparent kinetic differences, Φ analysis has shown that Im7 and Im9 have similar transition states, suggesting that differences in observed kinetics do not preclude similarities in folding mechanism (*35*). Protein G and protein L initiate folding from different $\beta$-hairpins, despite having similar final structures (*36*, *37*), and the initiating site for folding in protein G has been changed to that of protein L by site-directed mutagenesis (*38*).

The mutations described here, F68C-IFABP, C69F-ILBP, F93S-IFABP, and F94S-ILBP are dramatic substitutions, swapping large hydrophobic residues and smaller more hydrophilic residues in very different sequence contexts. This degree of mutation can significantly disrupt structure, but these particular replacements are found in other members of the family, and do not appear to prevent the formation of the native structure.

*Structure*. The structures of IFABP and ILBP (Figure 1) consist of two orthogonal $\beta$-sheets enclosing a large solvent-filled pocket in the apoprotein (*2*). The mouth of the pocket is closed by a helix−turn−helix motif connecting the first two $\beta$-strands. Since a large amount of solvent is contained within the structure, the hydrophobic core of these proteins is small and located opposite the two helices (Figure 7). The interior solvent pocket in the apoproteins is immediately adjacent to the core. This hydrophobic core is poorly conserved in sequence, since these large hydrophobic residues are frequently replaced by smaller more hydrophilic residues. The two structural locations chosen for mutation in this study reflect this diversity.

The most frequently observed residue at position 68 in this family of proteins is phenylalanine (Figure 7). However, a cysteine residue is found at this position (C69) in rat ILBP, and alanine, serine, and threonine are also observed. In the NMR structure of pig ILBP (*6*), the sulfhydryl group of C69 is oriented toward the external solvent (Figure 7). In human ILBP (*24*), this residue is a serine, and the hydroxyl group is also oriented toward the external solvent. Both the $C_\beta$ and the terminal atom in these side chains have significant solvent exposure. However, in sequences where phenylalanine is found at this position, the side chain is rotated into the hydrophobic interior of the protein, although the $C_\beta$ still has some exposure to solvent (*5*). The structural models predict that the substituted phenylalanine could be buried in C69F-ILBP, with minimal movements of other side chains, principally M71 (Figure 7). Similarly, the cysteine side chain is predicted to project into solvent for F68C-IFABP, by a simple rotation of the $C_\alpha$ to $C_\beta$ bond. Structures of these mutant proteins are needed to confirm these predictions.

The most frequent amino acid in the sequence alignment at position 93 is glutamine (about 70% of the sequences), which has some exposure to the internal solvent pocket. Phenylalanine is observed at this position in 15% of the sequences, including both proteins used in this study. One edge of the phenylalanine side chain is slightly exposed to the internal solvent pocket. Eight other amino acids are found at this site as well. The NMR structures of pig ILBP have a serine at this structural location, which is not involved in any obvious hydrogen-bonding interactions, and some of the model structures direct the hydroxyl group toward the internal solvent pocket (*6*). As such, serine was substituted in IFABP and rat ILBP to test the effect of a hydrophilic substitution at this structural location on the folding mechanism.

The structural models and the spectral data suggest that these mutations had minimal effects on the native structure. However, these mutations had significant effects on the apparent folding mechanism, changing the number and types of the intermediates observed.

*The Folding of Proteins in this Family*. The model shown here provides a conceptual framework for understanding the folding of these proteins

$$N \leftrightarrow I_{MG} \leftrightarrow I_{TC} \leftrightarrow I_{NMR} \leftrightarrow (U_E \leftrightarrow U_B)$$

In this model, N represents the native state. $I_{MG}$ is a state observed by stopped-flow kinetics that has molten-globule-like properties (nativelike secondary structure and perturbed tertiary contacts). This state can bind ligand (*30*), and the wavelength of maximal emission indicates that the tryptophan environments are similar in hydrophobicity to the native state. However, the fluorescence intensity is not identical to the native state. The $I_{TC}$ state is observed only by stopped-flow fluorescence. At least one tryptophan must participate in $I_{TC}$ in order for it to be observed. $I_{TC}$ has little if any secondary structure by stopped-flow CD, and the tryptophan environments are hydrophilic, with an emission maximum similar to that of the unfolded state. The fluorescence intensity of this state is perturbed by some tertiary contacts (thus $I_{TC}$) to other residues. $I_{NMR}$ is an equilibrium intermediate that lacks secondary structure by CD criteria. The fluorescence signal of $I_{NMR}$ appears to be identical to that of the unfolded state. However, both amide proton (*39*) and fluorine NMR experi-

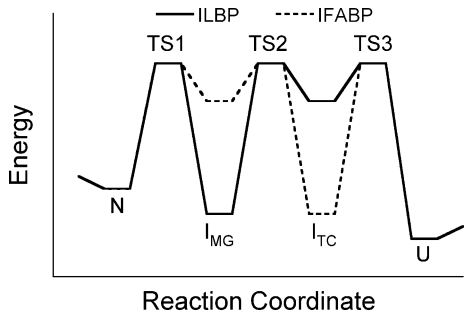


FIGURE 8: Simplified energy diagram for the unfolding of IFABP (solid line) and ILBP (dashed line). N, $I_{MG}$, $I_{TC}$, and U are potential or actual structures on the reaction paths. TS1, TS2, and TS3 are the transition states between these structures.

ments (*26*) show that it is in chemical exchange with the unfolded state ensemble for IFABP and other proteins (unpublished data). Although $I_{NMR}$ has some properties similar to those of $I_{TC}$ (no secondary structure, solvent exposed tryptophans), the states are not identical. Stopped-flow kinetic experiments show that the rate of exchange between $I_{TC}$ and the unfolded state is on the seconds time scale, whereas the rate of exchange between $I_{NMR}$ and the unfolded state must be greater than 100 $s^{-1}$ (*26*). The residues that participate in $I_{NMR}$ probably form the initiating site for folding, bringing together hydrophobic side chains that are not local in the sequence that interact with each other in the native state. The structural states within the parentheses, $U_E$ and $U_B$, are in rapid equilibrium with each other. $U_E$ represents the extended unfolded state at high concentrations of denaturant, whereas $U_B$ represents the structures associated with the burst phase, which appears to be a nonspecific hydrophobic collapse of the unfolded state at low final concentrations of denaturant (*40*). The spectral characteristics of $U_B$ are not similar to those of $I_{MG}$, which is much more nativelike.

Each of these states have been observed during the folding and unfolding of several of these proteins, although most proteins do not show all of these states. The observation of a particular intermediate depends on two major factors. First, the intermediate state must be of low enough energy to be significantly populated. As an example, consider the simple unfolding free energy diagram shown in Figure 8. In both cases, every protein molecule passes through both intermediate states during unfolding. In the case of the unfolding of ILBP, only $I_{MG}$ accumulates and is detected. For IFABP, $I_{TC}$ is stable enough to accumulate and be detected. Second, there must be a signal change accompanying that transition. For example, $I_{TC}$ has only been detected by tryptophan fluorescence. If a particular sequence does not have a tryptophan that participates in $I_{TC}$, no intermediate would be detected even if $I_{TC}$ is highly populated. This hypothesis describes a simplified energy landscape for folding where the proteins in this family follow the same path, but different intermediate states accumulate and are detected due to differences in their stability. Many proteins that are thought to follow a two-state model for the folding transition may have intermediate states that are not easily detected because of these two factors (*41*, *42*).

It is impossible to prove that each sequence follows this identical folding path through all of these states, since only those states that accumulate and can be detected are observed. However, all of these states have been observed for sequences that encode this fold, and single-site mutations can change the number and type of intermediates observed for any particular protein. This scheme for the folding of these proteins is the simplest model that fits the data, but other explanations are possible. For the sake of further discussion, polar substitutions are assumed to be less likely to interact with hydrophobic surfaces in intermediates, whereas hydrophobic substitutions are more likely to interact with those same surfaces.

During the stopped-flow unfolding of WT-IFABP, only $I_{TC}$ is observed (*4*, *10*, *11*, *27*). F68C-IFABP displayed both intermediates $I_{MG}$ and $I_{TC}$, while F93S-IFABP unfolded like WT-IFABP. As such, the F68C mutation may have stabilized an additional intermediate state that is not highly populated for WT-IFABP at neutral pH, although it is detected for WT-IFABP at pH 10 (*32*).

$U_B$ and $I_{MG}$ are directly observed during the refolding of WT-IFABP (*4*, *10*, *11*, *27*). $I_{TC}$ can only be detected indirectly for WT-IFABP by double-jump experiments, because the large amplitude change associated with the burst phase masks its appearance. The F68C mutation did not affect the kinetic folding pathway. However, $I_{TC}$ was detected for the F93S mutant protein, because the amplitude of the burst phase was so much smaller in F93S-IFABP compared to WT-IFABP. The presence of serine at position 93 had dramatic effects on the fluorescence of the burst phase, implying that long-range interactions (between S93 and W82) were involved in these states. These data suggest that the hairpin connecting positions 93 and 82 may be present very early in the folding process. Preliminary data on the native-state H/D exchange kinetics of IFABP suggest that the amides in this hairpin were among the slowest to exchange (unpublished results), supporting this hypothesis.

WT-ILBP displays both $I_{MG}$ and $I_{TC}$ during unfolding (*4*). However, only $I_{TC}$ was detected during the unfolding of C69F-ILBP, suggesting that this mutation prevented the accumulation of the molten-globule intermediate normally observed during unfolding. The reciprocal mutation in IFABP (F68C-IFABP) caused the appearance of the $I_2$ intermediate state. The F94S mutation did not change the unfolding mechanism from that of WT-ILBP.

WT-ILBP displays only $I_{MG}$ during refolding. No burst phase is observed for this protein, and it appears that the stability or lifetime of $I_{TC}$ is not sufficient for it to be detected (*4*). There was no detectable change in the folding mechanism for C69F-ILBP. On the other hand, F94S-ILBP exhibited only $I_{TC}$ during refolding, completely changing the nature of the observed intermediate. It is interesting that substitution of a serine for phenylalanine at this site allowed $I_{TC}$ to be observed for both IFABP and ILBP. This suggests a role for this residue in either the stability or lifetime of $I_{TC}$ during the refolding of proteins in this family. Unfortunately, the structural basis for the appearance of this intermediate is not clear. This issue is currently being pursued.

In conclusion, these data show that single-site mutations in the cores of these proteins can significantly perturb the apparent folding and unfolding paths, stabilizing some intermediate states and/or destabilizing others. The size and stability of an initiating core of hydrophobic interactions appears to be crucial to the types of intermediates that are observed, and thus to the apparent folding path.

## REFERENCES


1. Gunasekaran, K., Eyles, S. J., Hagler, A. T., and Gierasch, L. M. (2001) Keeping it in the family: folding studies of related proteins. *Curr. Opin. Struct. Biol. 11*, 83−93.
2. Banaszak, L., Winter, N., Xu, Z., Bernlohr, D. A., Cowan, S., and Jones, T. A. (1994) Lipid-binding proteins: A family of fatty acid and retinoid transport proteins. *Adv. Protein Chem. 45*, 89−151.
3. Hertzel, A. V., and Bernlohr, D. A. (2000) The mammalian fatty acid-binding protein multigene family: Molecular and genetic insights into function. *Trends Endocrinol. Metab. 11*, 175−180.
4. Dalessio, P. M., and Ropson, I. J. (2000) $\beta$-sheet proteins with nearly identical structures have different folding intermediates. *Biochemistry 39*, 860−871.
5. Scapin, G., Gordon, J. I., and Sacchettini, J. C. (1992) Refinement of the structure of recombinant rat intestinal fatty acid-binding apoprotein at 1.2 Å resolution. *J. Biol. Chem. 267*, 4253−4269.
6. Lucke, C., Zhang, F., Ruterjans, H., Hamilton, J. A., and Sacchettini, J. C. (1996) Flexibility is a likely determinant of binding specificity in the case of ileal lipid binding protein. *Structure 4*, 785−800.
7. Hodsdon, M. E., and Cistola, D. P. (1997) Discrete backbone disorder in the nuclear magnetic resonance structure of apo intestinal fatty acid-binding protein: implications for the mechanism of ligand entry. *Biochemistry 36*, 1450−1460.
8. Burns, L. L., Dalessio, P. M., and Ropson, I. J. (1998) Folding mechanism of three structurally similar $\beta$-sheet proteins. *Proteins 33*, 107−118.
9. Burns, L. L., and Ropson, I. J. (2001) Folding of intracellular retinol and retinoic acid binding proteins. *Proteins 43*, 292−302.
10. Ropson, I. J., Gordon, J. I., and Frieden, C. (1990) Folding of a predominately $\beta$-structure protein: rat intestinal fatty acid binding protein. *Biochemistry 29*, 9591−9599.
11. Ropson, I. J. and Dalessio, P. M. (1997) Fluorescence spectral changes during the folding of intestinal fatty acid binding protein. *Biochemistry 36*, 8594−8601.
12. Clark, P. L., Weston, B. F., and Gierasch, L. M. (1996) Intrinsic tryptophans of CRABP I as probes of structure and folding. *Protein Sci. 5*, 1108−1117.
13. Sacchettini, J. C., Banaszak, L. J., and Gordon, J. I. (1990) Expression of rat intestinal fatty acid binding protein in *E. coli* and its subsequent structural analysis: A model for studying the molecular details of fatty acid−protein interaction. *Mol. Cell. Biochem. 98*, 81−93.
14. Tochtrop, G. P., Richter, K., Tang, C., Toner, J. J., Covey, D. F., and Cistola, D. P. (2002) Energetics by NMR: Site specific binding in a positively cooperative system. *Proc. Natl. Acad. Sci. U.S.A. 99*, 184−1852.
15. Johnson, B. H., and Hecht, M. H. (1994) Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology 12*, 1357−1360.
16. Kim, K., and Frieden, C. (1998) Turn scanning by site-directed mutagenesis: Application to the protein folding problem using the intestinal fatty acid binding protein. *Protein Sci. 7*, 1821−1828.
17. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci. 4*, 2411−2423.
18. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol. 131*, 266−280.
19. Gualfetti, P. J., Bilsel, O, and Matthews, C. R. (1999) The progressive development of structure and stability during the equilibrium unfolding of the α-subunit of tryptophan synthase from *Escherichia coli*. *Protein Sci. 8*, 1623−1635.
20. Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method: Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry 27*, 8063−8068.
21. Mannervik, B. (1982) Regression analysis, experimental error, and statistical criteria in the design and analysis of experiments for discrimination between rival kinetic models. *Methods Enzymol. 87*, 370−390.
22. Motulsky, H. J., and Ransnas, L. A. (1987) Fitting curves to data using nonlinear regression: A practical and nonmathematical review. *FASEB J. 1*, 365−374.
23. Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis 18*, 2714−2723.
24. Kurz, M., Brachvogel, V., Matter, H., Stengelin, S., Thuring, H., and Kramer, W. (2003) Insights into the bile acid transportation system: The human ileal lipid-binding protein-cholyltaurine complex and its comparison with homologous structures. *Proteins 50*, 312−328.
25. Eisenhaber, F., and Argos, P. (1993) Improved Strategy in Analytic Surface Calculation for Molecular Systems: Handling of Singularities and Computational Efficiency. *J. Comput. Chem. 14*, 1272−1280.
26. Ropson, I. J. and Frieden, C. (1992) Dynamic NMR spectral analysis and protein folding: Identification of a highly populated folding intermediate of rat intestinal fatty acid binding protein. *Proc. Natl. Acad. Sci. U.S.A. 89*, 7222−7226.
27. Dalessio, P. M., Fromholt, S. E., and Ropson, I. J. The role of Trp-82 in the folding of intestinal fatty acid binding protein. *Proteins*, in press.
28. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci. 4*, 2138−2148.
29. Arai, M., and Kuwajima, K. (2000) Role of the molten globule state in protein folding. *Adv. Protein Chem. 53*, 209−282.
30. Chattopadhyay, K., Zhong, S., Yeh, S.-R., Rousseau, D. L., and Frieden, C. (2002) The intestinal fatty acid binding protein: the role of turns in fast and slow folding processes. *Biochemistry 41*, 4040−4047.
31. Yeh, S.-R., Ropson, I. J. and Rousseau, D. L. (2001) Hierarchical folding of intestinal fatty acid binding protein. *Biochemistry 40*, 4205−4210.
32. Dalessio, P. M., and Ropson, I. J. (1998) pH dependence of the folding of fatty acid binding protein. *Arch. Biochem. Biophys. 359*, 199−208.
33. Seeliger, M. A., Breward, S. E., and Itzhaki, L. S. (2003) Weak cooperativity in the core causes a switch in folding mechanism between two proteins of the cks family. *J. Mol. Biol. 325*, 189−199.
34. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C., and Radford, S. E. (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol. 286*, 1597−1608.
35. Friel, C. T., Capaldi, A. P., and Radford, S. E. (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and differences in the folding of homologous proteins. *J. Mol. Biol. 326*, 293−305.
36. McCallister, E. L., Alm, E., and Baker, D. (2000) Critical role of $\beta$-hairpin formation in protein G folding. *Nat. Struct. Biol. 7*, 669−673.
37. 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.
38. Nauli, S., Kuhlman, B., and Baker, D. (2001) Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol. 8*, 602−605.
39. Hodsdon, M. E., and Frieden, C. 2001. Intestinal fatty acid binding protein: The folding mechanism as determined by NMR studies. *Biochemistry 40*, 732−742.
40. Sosnick, T. R., Shtilerman, M. D., Mayne, L., and Englander, W. S. (1997) Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci. U.S.A. 94*, 8545−8550.
41. Bai, Y. (2003) Hidden intermediates and Levinthal paradox in the folding of small proteins. *Biochem. Biophys. Res. Comm. 305*, 785−788.
42. Sanchez, I. E., and Kiefhaber, T. (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol. 325*, 367−376.


BI048125U