Intestinal Fatty Acid Binding Protein: The Folding Mechanism As Determined by NMR Studies†

Michael E. Hodsdon‡ and Carl Frieden*

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received July 5, 2000; Revised Manuscript Received October 20, 2000

ABSTRACT: The intestinal fatty acid binding protein is composed of two β-sheets surrounding a large interior cavity. There is a small helical domain associated with the portal for entry of the ligand into the cavity. Denaturation of the protein has been monitored in a residue-specific manner by collecting a series of two-dimensional $^{1}H-^{15}N$ heteronuclear single-quantum coherence (HSQC) NMR spectra from 0 to 6.5 M urea under equilibrium conditions. In addition, rates for hydrogen−deuterium exchange have been measured as a function of denaturant concentration. Residual, native-like structure persists around hydrophobic clusters at very high urea concentrations. This residual structure (reflecting only about 2−7% persistence of native-like structure) involves the turns between β-strands and between the two short helices. If this persistence is assumed to reflect transient native-like structure in these regions of the polypeptide chain, these sites may serve as nucleation sites for folding. The data obtained at different urea concentrations are then analyzed on the basis of peak intensities relative to the intensities in the absence of urea reflecting the extent of secondary structure formation. At urea concentrations somewhat below 6.5 M, specific hydrophobic residues in the C-terminal β-sheet interact and two strands, the D and E strands in the N-terminal β-sheet, are stabilized. These latter strands surround one of the turns showing residual structure. With decreasing urea concentrations, the remaining strands are stabilized in a specific order. The early strand stabilization appears to trigger the formation of the remainder of the C-terminal β-sheet. At low urea concentrations, hydrogen bonds are formed. A pathway is proposed on the basis of the data describing the early, intermediate, and late folding steps for this almost all β-sheet protein. The data also show that there are regions of the protein which appear to act in a concerted manner at intermediate steps in refolding.

The rat intestinal fatty acid binding protein (IFABP)[1] is one member of a large class of proteins that bind a diverse set of ligands (fatty acids, retinoids, and bile salts) into a large cavity located in the interior of the protein. IFABP is a small (15 kDa) protein consisting mostly of antiparallel β-strands having two tryptophan but no proline or cysteine residues. It and other members of the superfamily have been extensively used as models for the folding of β-sheet systems (*1*−*9*). In an early study, Ropson et al. (*2*) examined the ^{19}F NMR spectrum of the intestinal protein containing 6-^{19}F-labeled tryptophan, as a function of urea concentration. The data were consistent with the presence of an intermediate form in equilibrium with the native and unfolded protein. The maximum concentration of this intermediate occurred near the midpoint of the denaturation curve. More recently, we (*3*, *4*) have attempted to define the properties of such intermediates by performing a series of site-directed mutagenesis experiments. One turn between two antiparallel β-strands, termed the D−E turn, is unusual in that there is a relatively large gap between adjacent strands with hydrophobic interactions between residues in the D and E strands and no backbone−backbone hydrogen bonds. While there is no clearly defined hydrophobic core in IFABP, the region including this turn represents a hydrophobic cluster of residues. Kim et al. (*4*) showed that a leucine residue (Leu64) in the D−E turn was essential for stabilizing the final folded form of the protein. These authors proposed that this residue could play a role in both the early (a hydrophobic collapse) and late (structure stabilization) steps in the folding process. Replacing this residue with glycine leads to a much less stable protein (*3*) and considerable destabilization in the structure (M. Rajabzadeh and C. Frieden, unpublished data). These results suggested that it would be useful to examine the NMR properties of IFABP as a function of denaturant concentration.

Cistola and co-workers have determined the NMR structure of both apo (*10*, *11*) and ligand-bound (*12*) rat IFABP. The holo structure agreed with that determined by X-ray crystallography (*13*). In contrast to the X-ray structure of the apo structure (*14*), the NMR data showed local disorder and exhibited conformational differences in the regions of the D−E turn, the C−D turn, and the C-terminal half of the second α-helix (*10*, *11*). We have now used the NMR assignments obtained by Hodsdon and Cistola to examine persistence of structure in apoIFABP as a function of

† Supported by NIH Grant DK13332 to C.F. and Washington University Digestive Diseases Research Core Center Grant P30 DK52574.

* To whom correspondence should be addressed. Phone: (314) 362-3344. Fax: (314) 362-7183. E-mail: frieden@biochem.wustl.edu.

‡ Current address: Department of Pharmacology, P.O. Box 208066, Yale University School of Medicine, New Haven, CT 06520.

[1] Abbreviations: IFABP, apo intestinal fatty acid binding protein; EDTA, ethylenediaminetetraacetic acid; HSQC, heteronuclear single-quantum coherence.

10.1021/bi001518i CCC: $20.00
Published on Web 12/29/2000

denaturant concentration. The denaturation of IFABP was monitored in a residue-specific manner by collection of a series of two-dimensional $^1H-^{15}N$ HSQC NMR experiments at increasing urea concentrations under equilibrium conditions. In addition, we have measured the rate constants for hydrogen−deuterium exchange as the denaturant concentration is increased. Both measurements were used to define a mechanism for the folding process. In this mechanism, hydrophobic clusters are formed early, followed by strand formation and finally the stabilization of hydrogen bonds. The initial step in the folding process is implied from the observation that even at very high denaturant concentrations, the protein senses some native-like structure as determined by the identification of specific resonances in the NMR spectrum.

MATERIALS AND METHODS

Protein Preparation. The uniformly labeled ^{15}N-enriched protein was expressed and purified as described by Hodsdon and Cistola (*12*). The final amount of protein obtained after the lipid was removed was 530 mg with 96% incorporation of ^{15}N.

Materials. Ultrapure urea was purchased from United States Biochemical (Cleveland, OH), and stocks were freshly prepared. The concentration of urea was determined by the refractive index at 25 °C (*14*). All other chemicals were reagent grade.

NMR Spectroscopy. Measurements were performed at 25 °C on a Varian Unity 600 MHz spectrometer equipped with a Nalorac triple-resonance gradient probe. Two-dimensional $^1H-^{15}N$ HSQC spectra were collected with 3.5 and 7.7 kHz spectral widths and 256 and 512 complex points along the F1(^{15}N) and F2(^1H) frequency dimensions, respectively, using the gradient- and sensitivity-enhanced pulse sequence of Kay and co-workers (*15*, *16*). A delay value of 2.5 ms ($\leq {}^1/_4 J_{NH}$) was employed during the INEPT subsequences (*17*). A 1.9 ms low-power, rectangular pulse was used to selectively excite the water resonance and return its magnetization to an equilibrium state (no water presaturation was utilized). ^{15}N decoupling during acquisition was accomplished using a GARP composite pulse sequence (*18*) with a field strength of 1.1 kHz. Each FID consisted of 16 signal-averaged scans separated by a 1.0 s relaxation delay.

Equilibrium NMR Spectroscopy of [^{15}N]IFABP. A series of 600 μL NMR samples were prepared, each containing 1 mM [^{15}N]IFABP (diluted from a 4 mM stock), 20 mM potassium phosphate buffer at pH 6.5, 0.25 mM EDTA, 10% D_2O by volume, and increasing concentrations of urea (0.0, 1.0, 2.0, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, and 6.5 M). The samples were prepared by a final addition of the protein stock to premixed solutions of urea and buffer. The precise concentration of the urea stock solution was determined using the refractive index at 25 °C. NMR samples were allowed to equilibrate during an overnight incubation at 4 °C.

At each denaturant concentration that was used, the NMR probe was tuned and matched, the magnetic field was shimmed, and the 90° pulse widths were optimized for the sample (no deviation in the optimal ^1H or ^{15}N 90° pulse width was required for any of the urea concentrations). A $^1H-^{15}N$ HSQC spectrum was recorded for each sample with a total experimental time of 155 min. An additional HSQC spectrum for the sample containing no urea was collected with a widened F1(^{15}N) spectral width of 5 kHz to identify any aliased resonances. NMR spectra were processed with the software package NMRPipe (*19*), zero-filled to 1024 and 512 total points in the F2 and F1 dimensions, respectively, and apodized with a cosine window function along F2 and a squared-cosine function along the F1 dimension along with a first point multiplier of 0.5 for F1 only. ^1H chemical shifts were referenced versus external sodium 3-(trimethylsilyl)-propionate-2,2,3,3-d_4 (TSP), and ^{15}N chemical shifts were referenced indirectly using the protocol of Wishart et al. (*20*). Processed spectra were imported for display and analysis in Sparky (Goddard, T. D., and Kneller, D. G. *SPARKY 3*, University of California, San Francisco, CA).

To confirm assignment of peaks observed in the $^1H-^{15}N$ HSQC spectra during the urea-induced transition zone of unfolding, three-dimensional $^1H-^{15}N$ NOESY-HSQC and TOCSY-HSQC spectra were collected for the samples containing 3.5 and 2.0 M urea, respectively. Again, gradient- and sensitivity-enhanced pulse sequences of Kay and co-workers (*16*) were utilized. Experimental parameters were the same as the analogous parameters described above in detail for $^1H-^{15}N$ HSQC spectroscopy, except as stated below. Spectral widths were reduced to 7.2, 2.0, and 7.2 kHz with 128, 40, and 256 complex points in the F1(^1H), F2(^{15}N), and F3(^1H) frequency dimensions, respectively. Each FID consisted of 16 and 28 signal-averaged scans for the NOESY-HSQC and TOCSY-HSQC spectra, respectively. An isotropic mixing time of 150 ms was allowed for the NOESY-HSQC experiment. In the TOCSY-HSQC experiment, a 40 ms DIPSI-3 composite pulse subsequence for spin locking of the transverse magnetization was used (*21*).

Measurement of Hydrogen−Deuterium Exchange Rates at Varying Urea Concentrations. A saturated urea solution was prepared and the concentration determined by the refractive index at 25 °C. To replace the exchangeable protons on the urea with deuterons, the solution was taken to dryness and resuspended in 99.8% D_2O (Cambridge Isotope Laboratories, Andover, MA). This procedure was then repeated to remove residual protons resulting from the exchange of urea amide protons with solvent. The extent of deuteration was checked with a one-dimensional ^1H NMR spectrum by comparison of the residual solvent peaks to added 0.1 mM TSP. Finally, the urea concentration was confirmed, again by the refractive index. Separate solutions were then prepared containing 20 mM potassium phosphate buffer at pH 6.5, 0.25 mM EDTA, and increasing concentrations of urea (0.0, 1.0, and 2.0 M), all dissolved in 99.8% D_2O. These solutions were used to saturate Sephadex G10 beads (3 h incubation) which were then extensively washed with the remaining solution. Separately, stock solutions of 1 mM [^{15}N]IFABP in the identical series of urea solutions described above, except for dissolution in 100% H_2O, were prepared and allowed to equilibrate for 3 h.

To monitor hydrogen−deuterium exchange, each slurry of Sephadex G10 beads in "D_2O" urea solution was poured into a separate spin column. Excess liquid was removed from the column by gentle centrifugation. Immediately prior to the NMR experiments, 600 μL of the 1 mM [^{15}N]IFABP "H_2O" solution with the corresponding urea concentration was added to the top of the column and then the column spun for 30 s. Approximately 500 μL of the filtrate was

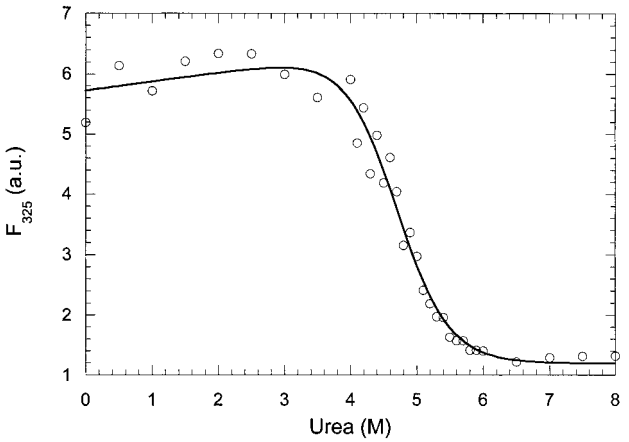

FIGURE 1: Equilibrium unfolding of native IFABP as a function of urea concentration as monitored by fluorescence (given in arbitrary units). The experiment was performed in 0.02 M phosphate buffer and 0.25 mM EDTA at pH 6.6 and 25 °C using an excitation wavelength of 290 nm and an emission wavelength of 325 nm. The solid line is a fit to the data as described by Ropson et al. (*1*).

collected, transferred to a clean and well-dried NMR sample tube, and immediately placed in the spectrometer, and 2 min was allowed for temperature equilibration (25 °C). After tuning and matching of the NMR probe and shimming of the magnetic field, acquisition of the series of NMR experiments was initiated, approximately 10 min after the solvent exchange. Two-dimensional $^1H-^{15}N$ HSQC NMR spectra were collected serially using a pulse sequence and acquisition parameters identical to those described above for the equilibrium NMR experiments. A one-dimensional ^1H NMR spectrum was collected initially and repeated after every six $^1H-^{15}N$ HSQC experiments to monitor any loss of signal due to protein aggregation using the nonexchangeable aliphatic protons. NMR spectra were collected serially until hydrogen−deuterium exchange was observed to be complete. Processing and analysis of the spectra were performed in a manner identical to that described above for the equilibrium HSQC experiments. Peak heights were fit to a simple two-parameter exponential function [$I(t) = I(t_0)\exp(-kt)$] to determine the rate constants for amide hydrogen exchange using a protocol equivalent to that described by Hodsdon and Cistola (*11*).

Fluorescence Measurements. These were obtained using a PTI Alphascan fluorometer (Photon Technology International) with an excitation wavelength of 290 nm and an emission wavelength of 325 nm. Denaturation curves were obtained using 1 mL protein samples at different urea concentrations. The experiment was performed in 0.02 M potassium phosphate buffer and 0.25 mM EDTA at pH 6.6 and 25 °C. Samples were incubated with the appropriate urea concentrations overnight to correlate with the NMR experiments.

RESULTS

Fluorescence Measurements. Changes in intrinsic fluorescence were used as a measure of the level of denaturation of IFABP. Figure 1 shows fluorescence data as a function of urea concentration obtained under similar conditions used in subsequent NMR experiments (0.02 M phosphate at pH 6.6 and 25 °C) except at much lower protein concentrations. The data were fit as described by Ropson et al. (*1*) using the equation of Santoro and Bolen (*22*). The midpoint of the curve occurs at 4.68 M urea. The data indicate that the protein is essentially completely denatured at ≥6 M urea. It should be noted that there is a slight increase in the fluorescence between 0 and 3 M urea. This increase may reflect the formation of intermediate form(s) of IFABP as described by the NMR results which are discussed below.

NMR Studies. The denaturation of IFABP was monitored in a residue-specific manner by collection of a series of two-dimensional $^1H-^{15}N$ HSQC NMR experiments at increasing urea concentrations under equilibrium conditions. Of the 11 urea concentrations that were investigated, HSQC spectra at 0.0, 2.0, 3.5, 4.0, 5.0, and 6.0 M urea are presented in Figure 2. Beginning with the previously established $^1H-^{15}N$ HSQC assignments under native conditions (*11*), we monitored changes in the chemical shifts of the backbone amides during the urea titration. Ambiguities in the assignment of the folded resonances at intermediate urea concentrations were eliminated by inspection of three-dimensional $^1H-^{15}N$ TOCSY-HSQC and NOESY-HSQC spectra at 2.0 and 3.5 M urea, respectively.

Similar to the results reported by van Mierlo et al. (*23*) for apoflavodoxin, a two-step transition (a minimum of three states) is observed for the unfolding of IFABP as monitored by the $^1H-^{15}N$ HSQC NMR spectra. A majority of the peaks, which correspond to native IFABP, display a loss of peak intensity between approximately 1.0 and 3.5 M urea. Resonances corresponding to the fully denatured protein begin to appear in this same range of urea concentrations and become maximal between 5.0 and 6.0 M urea. Evidence for the presence of a third, intermediate protein conformation present between 1.0 and 4.5 M urea is twofold. First, additional peaks appear coincident with the disappearance of the native resonances. These additional peaks do not correspond to any of the peaks seen for the native or fully denatured protein (>6.0 M urea). These intermediate resonances appear initially around 1.0 M urea, become maximal at about 2.0−3.5 M urea, and then disappear completely by 5.5 M urea. Second, a significant amount of total intensity appears to be missing in the range of 3.5−4.0 M urea. Inspection of the $^1H-^{15}N$ HSQC spectrum at 3.5 M urea in Figure 2 is illustrative. Although virtually all of the native state intensity has been lost, only a fraction of the intensity for the fully denatured protein has appeared. The missing NMR intensity can be explained by the presence of an intermediate protein conformation whose amide resonances have been broadened by exchange between multiple states. The time scale of these exchange processes is on the order of the population-weighted difference in chemical shift (about 1 ms). This intermediate may consist of either many highly populated and exchanging states or a single state in exchange with the fully denatured protein. In contrast, the line widths of the resonances corresponding to the fully native state are not broadened at any point during the transition. Therefore, native IFABP remains in slow exchange with both the intermediate and unfolded states of the protein.

Peak intensities were quantified by measurement of maximal peak heights using the Sparky software package and are available as Supporting Information (Table 1). Correlation of NMR peak heights to the fractional occupancy of conformational states relies on a number of assumptions which we have attempted to address. First, a gradient-

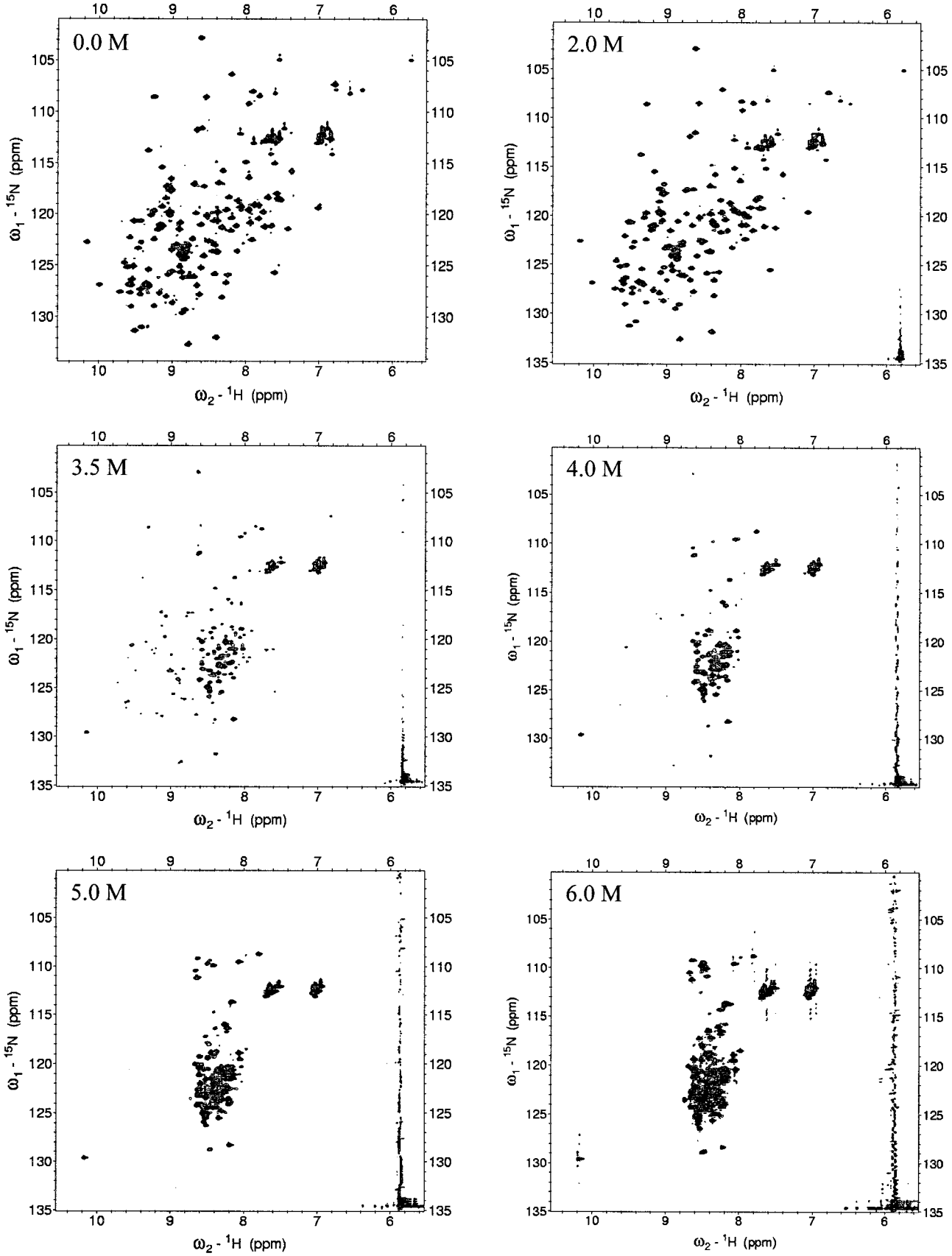

FIGURE 2: Gradient and sensitivity-enhanced two-dimensional $^1H-^{15}N$ HSQC spectra collected as described in the text are displayed for urea concentrations of 0.0, 2.0, 3.5, 4.0, 5.0, and 6.0 M. All spectra are displayed with an identical threshold. A vertical stripe of "T1 noise" can be seen at about 5.85 ppm (^1H chemical shift) corresponding to the solvent urea resonance.

enhanced $^1H-^{15}N$ HSQC pulse sequence was utilized which incorporated water flip-back pulses, thus minimizing decreases in intensity from water saturation transfer. Second, careful inspection of the peak cross sections revealed that the line widths of the resonances corresponding to the fully native protein did not vary significantly during the titration,

indicating that relative peak heights are a reliable indicator of relative signal intensity. Third, the method of sample preparation avoided the development of a significant protein aggregate, which is often observed with the incremental addition of a concentrated denaturant stock solution to the protein sample (*23*). A minimal aggregate (<10%, visually estimated) was observed only for the 2.0 M urea sample, which did not result in a significant loss of methyl resonance intensity in a one-dimensional pulse-acquire (presaturation of water) NMR experiment. Last, the use of urea, a neutral denaturant, appears to have minimized any loss of signal to noise due to a decrease in the probe "*Q*-factor", which has been reported for experiments using guanidine hydrochloride as a denaturant (*24*). Notably, the optimal 90° pulse widths of the ^{1}H and ^{15}N resonances were identical for all urea concentrations. Lengthening of the 90° pulse width is indicative of an inductive loss in the intensity of the NMR signal, i.e., decrease in the probe "*Q*-factor", with added salt or guanidine hydrochloride.

For many of the well-resolved resonances, peak volumes were also quantified and compared to peak heights. Peak volumes were determined both by fitting of the HSQC peaks to a Lorentzian peak shape and by summation of the net spectral intensity within an ellipsoid centered around the peak. Qualitatively, the three methods displayed the same general trends for loss of spectral intensity with increasing urea concentration. However, in more crowded regions, the results often differed. On a case-by-case basis, the fitting parameters could often be adjusted to account for some of the errors in integration, but no single method of determining peak volumes was optimal for all regions of the spectra. For this reason, we decided to utilize maximal peak heights instead of volumes for estimation of intensities as they could be objectively determined in a uniform manner. We were reassured in this decision by the qualitative agreement between the methods for well-resolved resonances. A selection of peak volumes for these residues is available as Supporting Information (Table 3).

Figure 3 plots the change in the NMR peak intensities for selected residues across the range of urea concentrations that were studied. A majority of the residues in IFABP followed a pattern similar to those residues shown in the upper panel of the figure. These residues had minimal loss of intensity at 1.0 M urea, followed by a sharp decrease around 2.0−3.5 M. For most residues, there was a complete loss of intensity by 4.0−4.5 M urea. A smaller subset of residues displayed the pattern shown in the middle panel of Figure 3, decreasing linearly to zero at around 4.0 M urea. This pattern of sensitivity at low urea concentrations localized into distinct regions of the protein sequence and tertiary structure. Diagrammed in Figure 4 is a backbone trace of the X-ray crystallographic structure of IFABP (*13*). Regions of secondary structure with increased sensitivity to urea, and by inference reduced stability, are shaded dark. They were chosen because the majority of residues within the segments had the lowest relative peak intensities at 3−4 M urea. These segments of low stability are grouped into two regions of the protein: the first being about half of the N-terminal β-sheet and the second being the second helical region. The latter is not surprising since Hodsdon and Cistola have shown that this region is disordered in the apoprotein in the absence of urea (*11*). While individual residues in the N-terminal

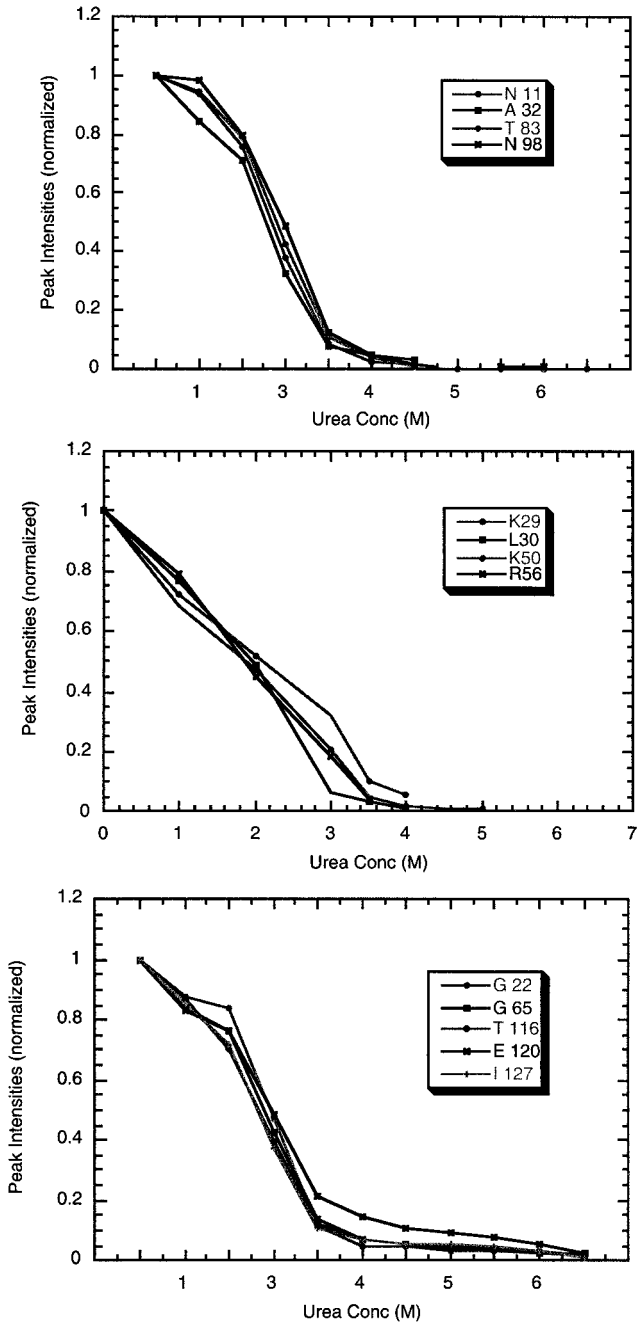

FIGURE 3: Peak intensities for the equilibrium urea titration of IFABP as monitored by $^{1}H-^{15}N$ HSQC NMR spectroscopy are shown as a function of urea concentration. In the upper panel, amide resonances of residues Asn11, Ala32, Thr83, and Asn98 are displayed. A majority of the residues in the protein follow a similar pattern. In the middle panel, the peak intensities of residues Lys29, Leu30, Lys50, and Arg56 are shown. The residues are members of the subset that displayed increased sensitivity to urea denaturation. In the bottom panel, amide resonances of residues Gly22, Gly65, Thr116, Glu120, and Ile127 are shown. These five residues are examples of residues found to display persistent native-like structure at high urea concentrations based upon the presence of residual $^{1}H-^{15}N$ HSQC peaks.

β-sheet region may respond differently to changes in urea, the data suggest that a portion of the sheet may move dynamically in a concerted fashion.

The residues displayed in the lower panel of Figure 3 and the spectra shown in Figure 5 illustrate a third pattern of sensitivity to urea denaturation with differences evident at high urea concentrations. Whereas a majority of amide

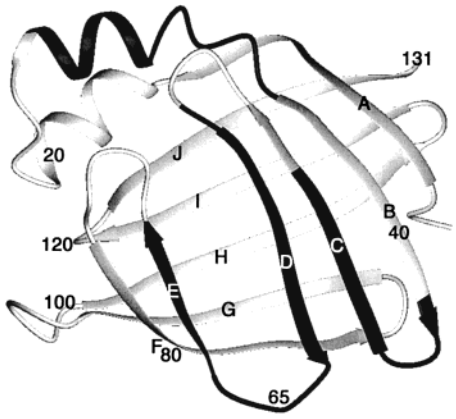

FIGURE 4: Molmol (*45*) structure of IFABP with those portions of the protein that are particularly sensitive to urea concentrations between 3 and 4 M shaded dark. The strands are labeled A−J, and the positions of some residue positions are shown.

resonances corresponding to the fully native protein are completely lost by 4.0−4.5 M urea, a small subset of $^{1}H-^{15}N$ HSQC peaks, which can be traced back to the folded state, are still detectable at urea concentrations as high as 5.5−6.5 M (up to 7% intensity relative to 0.0 M urea). The persistence of a residual intensity is confirmed by the determination of peak volumes described above. For most native resonances, the summed spectral intensity within a fixed-sized ellipsoid becomes vanishingly small (<0.1% of the intensity at 0.0 M urea) as the peak disappears into the noise at urea concentrations between 4.0 and 5.5 M urea. However, for the residues with apparent residual intensity at high urea concentrations, peak volumes determined in the same fashion were greater than 1% of the original volume at 0.0 M urea [selected examples of volume calculations can be found as Supporting Information (Table 3)]. Residues displaying residual intensity at high urea concentrations did not differ significantly from other residues with respect to their initial spectral intensities or peak line widths under native conditions.

The location in the tertiary structure of IFABP of all residues with observable $^{1}H-^{15}N$ HSQC peak intensities at >5.0 M urea involves regions near both the D and E strands and the last two strands (I and J). At these urea concentra-

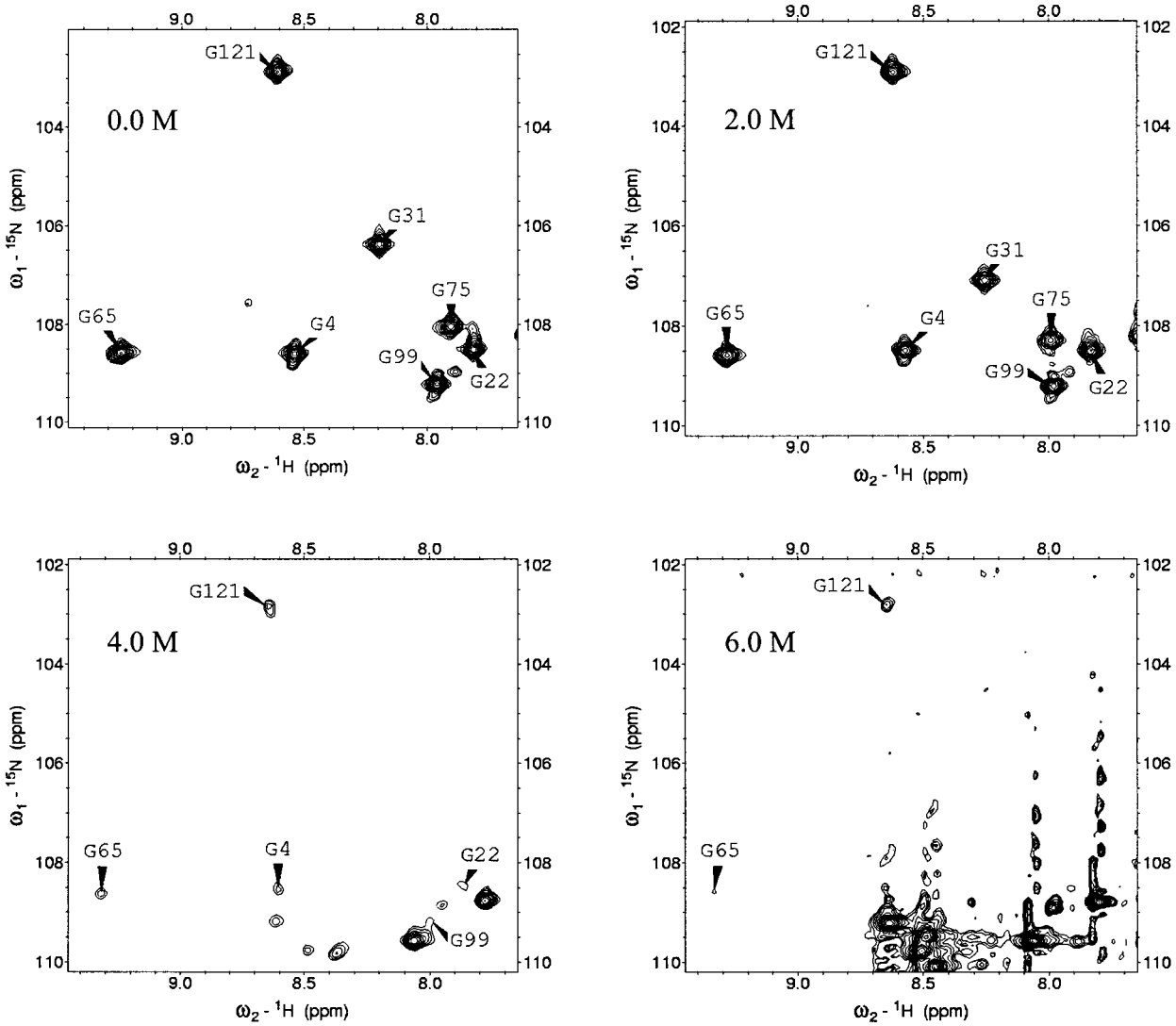

FIGURE 5: Subsections of the $^{1}H-^{15}N$ HSQC spectra displayed in Figure 2 are plotted with a number of residues labeled. The spectra that are shown correspond to urea concentrations of 0.0, 2.0, 4.0, and 6.0 M. Display thresholds are identical in all spectra except at 6.0 M urea, where the threshold has been reduced by a factor of 2 to demonstrate the residual resonances corresponding to residues Gly65 and Gly121. A significant amount of distortion is seen resulting from the truncation artifact of the very slowly relaxing unfolded Asn and Gln side chain amides.

tions, the native tertiary structure of the protein is completely destabilized by any measure and is not populated to any significant extent. Yet, the amide ^{1}H and ^{15}N chemical shifts of these residual peaks are similar to their values in the absence of urea and are well-resolved from the crowded region of the HSQC spectrum corresponding to the unfolded state. We interpret the presence of these amide resonances as indicative of some residual, localized structure in the conformation of the denatured polypeptide. Thus, at high urea concentrations, a few short sections of the polypeptide backbone appear to persist with conformations which, by chemical shift, are similar to their structure in the native protein. The importance of the turn between the D and E strands in the folding process has been discussed previously (*2*−*4*).

At low urea concentrations, the residues which appear to retain a small degree of residual native-like structure in the denatured state follow a peak intensity pattern similar to that of all other residues in IFABP. As can be seen in Figure 3, the behavior of residues in the lower panel is qualitatively similar to that of the residues shown in the other two panels, within the range of 0.0−4.0 M urea. At first, this similarity may seem to contradict the interpretation of residual structure for these residues, and a shift of the denaturation curve toward higher urea concentrations might have been expected. The explanation for the observed result is based on the known high cooperativity for the folding of the global tertiary structure of a protein. Although chemical denaturants have been shown to have a minimal effect on the conformations adopted by short peptides in solution (*25*), for large proteins these denaturants act by stabilizing the unfolded tertiary structure relative to the folded state, either by binding to exposed sites on the unfolded protein or by reducing the activity of bulk water. A suggested model for IFABP is that at high urea concentrations there are local sections of the polypeptide sequence that adopt a native-like structure, presumably transiently. Successful progress along the folding pathway would then be initiated from these structural elements.

Hydrogen−*Deuterium Exchange.* Rates of amide hydrogen−deuterium exchange were measured for the subset of backbone amides in IFABP that exchange slowly by monitoring the decrease in the $^{1}H-^{15}N$ HSQC peak intensities after exchange into a fully deuterated solution using a spin column technique (see Materials and Methods). A table of the exchange rates determined at 0.0, 1.0, and 2.0 M urea is available as Supporting Information (Table 2). Amide hydrogen exchange rates were significantly increased by the presence of low concentrations of urea such that by 2.0 M urea only a small fraction of the originally observable exchange rates are detectable after the 10 min dead time of the experimental setup (the time required for setting up the sample, tuning and matching the probe, and shimming the magnet). Above 2 M urea, all backbone amides exchanged too rapidly to be measured. Figure 6 shows the location of these remaining slowly exchanging amide hydrogens (colored dark) in the tertiary structure of IFABP. Strikingly, these residues are mostly localized to one-half of the C-terminal β-sheet. Amide hydrogen exchange rates of note include Thr83 and Val90 whose amide hydrogen exchange rates in the absence of urea are too slow to be measured within the current experimental framework. These have been observed

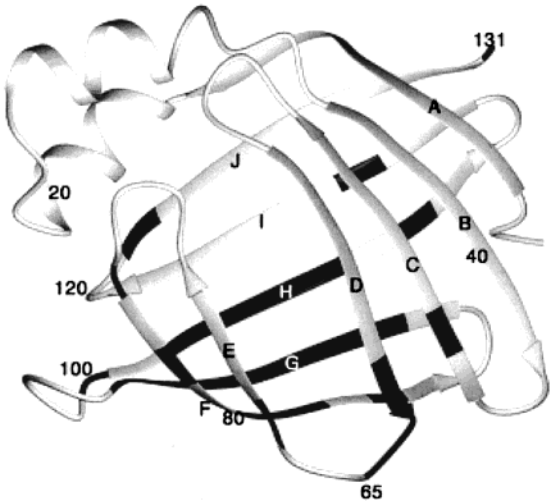

FIGURE 6: Structure of IFABP with those portions of the protein that exhibit slow hydrogen−deuterium exchange at 2 M urea shaded dark. Strands are labeled as in Figure 4, and the positions of some residues are shown.

to be preserved for weeks after dissolution of the protein in D_2O. Interestingly, of the residues believed to retain some residual structure in the denatured state (see Figure 7a), only residues 17, 63−66, and 131 have amide hydrogen exchange rates which are slow enough to be measured; that is, amide resonances corresponding to residues that appear to show persistent structure may be completely exchanged within the 10 min dead time in the absence of urea. The implications of these notable amide hydrogen exchange rates will be discussed further below.

DISCUSSION

Determining the mechanism of folding for any protein is a challenge. IFABP is of particular interest because it is composed of two β-sheets surrounding a large interior cavity filled with solvent. Consequently, the protein has no large hydrophobic core, but rather small clusters of hydrophobic residues with most side chains being solvent accessible. As noted earlier, it is an excellent model for protein folding studies. There have been a few other studies in which investigators have followed denaturation in a residue-specific manner using NMR (*23*, *26*, *27*) but none of a protein in the superfamily represented by IFABP. Earlier equilibrium and kinetic data suggested, using site-directed mutagenesis (*3*, *4*) or ^{19}F-labeled tryptophan (*2*), an important role for residues in or near the D−E turn for stability and the folding process. Thus, mutations in the D−E turn lead to destabilization of the protein, while both the ^{19}F-labeled protein and mutation data suggest the presence of intermediates during folding. The current NMR data confirm the conclusion of the importance of the D−E turn and also implicate residues in or near the I−J turn as well as those in the turn between the two small helices. In addition, in the range of 2.0−3.5 M urea, there is a major decrease in the $^{1}H-^{15}N$ HSQC NMR peak intensities without an equivalent increase in the intensities of the denatured peaks (Figure 2). This change in the backbone amide resonances for the entire protein indicates a cooperative transition between the rigid, well-formed secondary structure of the native state to one or more intermediate forms with either poorly formed secondary structure or a significantly more dynamic polypeptide

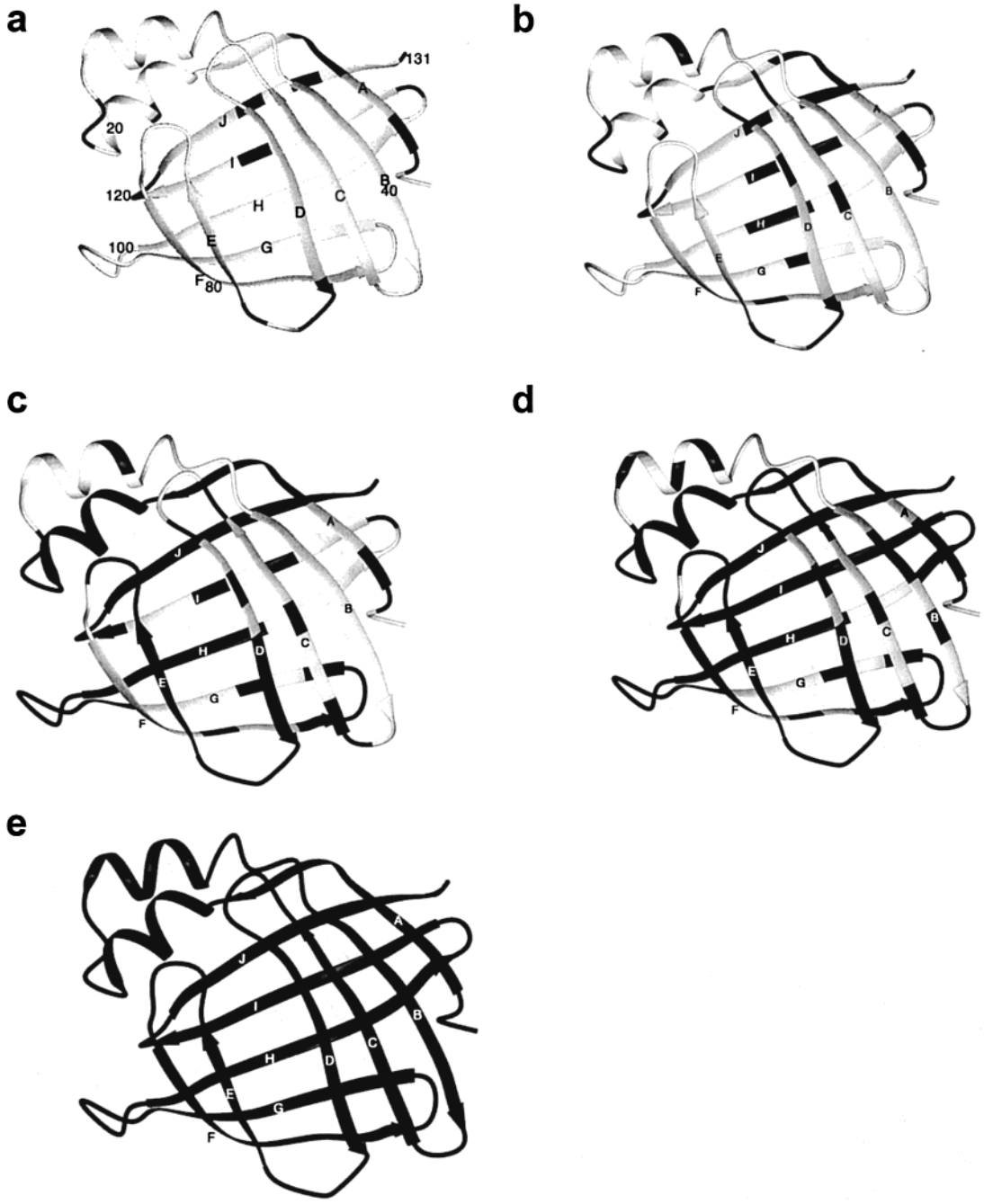

FIGURE 7: Proposed folding pathway of IFABP as described in the text. In this figure, the shaded regions reflect structure formation during refolding. The nonshaded regions would be disordered rather than structured as shown. (a) The initial hydrophobic collapse. (b−d) Subsequent intermediate steps in the folding process, including stabilization of strands A, B, D, and E as well as that of the C-terminal β-sheet as discussed in the text. (e) Final step in folding reflecting stabilization of the whole protein.

backbone. As described in detail in the Results, the spectra suggest that these non-native protein conformations are in intermediate exchange (on a millisecond time scale) while the native state is in slow exchange with both the intermediate and unfolded states. As can be seen from the data shown in Figure 4, there are some distinct subsections of the protein structure that may show concerted motions since they appear to be more susceptible to intermediate urea concentrations. These regions involve the C, D, and E stands as well as the second helical segment.

Folding Pathway. A pathway for the folding process of IFABP is proposed which is consistent with the NMR data collected under equilibrium conditions. The interpretation of equilibrium denaturation studies in defining a folding pathway is always controversial. Our justification for this procedure is that the major features of the kinetic mechanism have been previously defined by mutagenesis and kinetic studies (reviewed in the introductory section and references therein). The current NMR data contribute to this model by providing a description of the intrinsic structural properties for each residue during the equilibrium denaturation. The important assumption is that changes in relative peak intensities for each residue can be interpreted as the preferential formation (or disappearance) of structure. It is always possible that some intermediate structures present only at very low levels will not be detected by this method. It should also be clear that there is some subjectivity to this method in the sense of deciding at what point NMR

resonance intensities at a given urea concentration (relative to those of the native protein) form a sequential group of residues that reflect formation of a segment of secondary structure. With those caveats, we can then delineate the folding pathway into distinct steps.

Initial Folding Steps. Native-like structure persists at a level between 2 and 7% under strongly denaturing conditions in some regions of the protein (Figures 2, 3, and 5). We interpret these local regions of the polypeptide sequence to have an intrinsic and transient tendency toward native-like structure independent of the global fold of the protein. We propose that the initial step of refolding involves these residues with persistent structure in the unfolded state. Inspection of the tertiary structure of IFABP reveals that these residues with persistent structure form hydrophobic clusters which center around the turn between the helices, around the turn between β-strands D and E, and in and near the turn between the last two (I and J) strands (Figure 7a). Hydrophobic residues in the D−E turn [also implicated in previous studies (*2*, *4*)] are Val60, Val61, Leu64, and Leu89; those near the I−J turn are Tyr119 and Val122, and those near the helical region are Phe17, Ile23, and Val122.

From other studies, there is strong evidence that β-hairpin structures can form rapidly (*28*), and residual structure, sometimes including hydrophobic clustering, has been observed for several other proteins (*29*−*32*). Persistent structure in the turn between the I and J strands was surprising, although we had previously shown that the Gly121Val mutation in the I−J turn resulted in a decrease (2−3 orders of magnitude) in the rate of folding (*3*).

The importance of residues in the D−E turn for both early and late steps in the folding process has been previously discussed in the introductory section. Kim et al. (*4*) and Kim and Frieden (*3*) mutated residues in this turn and showed that Leu64 was critical for the overall stability of the protein. For some mutants, in which the three residues of the turn were mutated, kinetic refolding studies exhibited only a burst phase with no slower refolding phase being observed (*4*). Although these mutants were very unstable compared to the wild type, they showed high β-sheet content as measured by circular dichroism (*4*). It was concluded that in these triple mutants there was a loss of defined structure in the region of the D−E turn and that the interaction of Leu64 with other nearby hydrophobic residues was responsible for both (1) an early hydrophobic collapse constituting the slow step of refolding and (2) a later, final stabilization of the structure. These conclusions are supported by recent NMR data (M. Rajabzadeh and C. Frieden, unpublished results) indicating that the Leu64Gly mutation of the D−E turn results in destabilization of structure not only around the D−E turn but also in the C-terminal β-sheet.

Most unexpected was the appearance of a residual peak corresponding to Gly22. The turn between the two small helices (Met21, Gly22, Ile23, and Asn24) represents a Schellman motif, which is considered to be quite stable and has been predicted to serve as a common nucleation site for folding (*33*−*35*). The amide hydrogens of Gly22 and Ile23, which display persistent structure and are predicted to fold early, participate in the two hydrogen bonds known to stabilize this motif. Sukumar and Gierasch (*36*) have examined a small peptide from the equivalent region of the cellular retinoic acid binding protein and noted the importance of hydrophobic interactions of nearby residues in stabilizing the Schellman motif. However, the overall role of the helical region is not clear since that region can be removed entirely without much change in the protein structure (*37*), stability, or rate of refolding (*38*). Perhaps, however, any nucleation at this site is only slowly propagated through the structure. Taken together, the NMR results presented here suggest an early role of three distinct structural regions in the early steps of refolding of IFABP.

Intermediate Folding Steps. We propose that after the initial nucleating steps described above, intermediate steps of refolding occur involving hydrophobic interactions in the C-terminal β-sheet and between the D and E strands (see panels b and c of Figure 7), which together constitute a majority of the slower refolding of native IFABP. During the equilibrium renaturation pathway suggested by the current data, interactions first occur between hydrophobic residues in strands G−J (shaded dark in Figure 7b). Strands D, E, and J and the first helical region then become stabilized (Figure 7c). Stabilization of the whole C-terminal β-sheet, the A strand, and a portion of the second helical region follows (Figure 7d). Many of the residues that are stabilized after the initial collapse are hydrophobic (Figure 7b), while a higher percentage of those that form in the next step (Figure 7c) are polar. In previous experiments, we concluded that there was cooperativity in the formation of the C-terminal β-sheet since mutating any of the glycine residues in the turns between the last four β-strands slowed the refolding (*3*). The data presented here are consistent with that observation, although stabilization of the I and J strands appears to be critical. It is interesting that I. Ropson (personal communication) has shown that, as measured by circular dichroism, there is no structure in the peptide of residues from 83 to 131, i.e., the C-terminal β-sheet. Thus, stabilization of regions around the D−E turn may be critical for forming the C-terminal β-sheet.

Late Folding Steps. At $\leq$2 M urea, the remaining structure forms (Figure 7e). Hydrogen bond formation may be a late step since, as described in detail in the Results, amide hydrogen exchange at $>$2 M urea becomes too fast to assess in the dead time of the experimental procedure (10 min). Clark et al. (*8*) have proposed that hydrogen bonds form after formation of the internal cavity required for ligand binding. Our experiments agree with that but also suggest that there is some preference for the stabilization of hydrogen bonds between backbone carbonyls and amide nitrogens. Those that form early are primarily in the D−E turn and strands F−H, the latter three strands representing about half of the C-terminal β-sheet (Figure 6). Those that are stabilized later exist primarily between the last two (I and J) strands as well as between the B−D strands of the N-terminal β-sheet. It is interesting that many amide hydrogen exchange rates do not correlate with the regions of persistent structure seen at very high urea concentrations. For example, Thr83 and Val90 exchange extremely slowly in the native protein, suggesting amide hydrogen exchange should be protected in the globally unfolded state. However, their native structure does not persist at high urea concentrations according to the results presented here. Further, many amide hydrogens with evidence for persistent structure under highly denaturing conditions have very fast exchange rates in the absence of urea (e.g., Gly22, Ile23, Glu120, and Gly121). There are a number

of hydrogen bonds that are present in the native structure but exchange rapidly (less than 10 min), and these are found to be distributed rather randomly throughout the structure (Table 2 of the Supporting Information). One interpretation of these results is that slow hydrogen−deuterium exchange in the native state does not reflect regions of stability in the urea denatured state as has been discussed (*39*, *40*). Nevertheless, native state amide hydrogen exchange remains a very useful tool for monitoring local backbone stability under specific conditions.

Formation of the native structure most likely involves side chain stabilization as a late step. In this regard, data obtained with dihydrofolate reductase may be pertinent. For this protein, we have shown (*41*, *42*) that stabilization of tryptophan side chains (located near the surface) is a late step in folding. A similar conclusion has recently been published by Staniforth et al. for cystatin (*43*), and the same may be true for IFABP since the β-sheet structure is such that most residues are solvent-exposed.

The above description of the folding pathway for equilibrium renaturation represents our interpretation of the NMR data. Many of the features of this model are consistent with previous kinetic studies and mutagenesis experiments; hence, we hypothesize on a general agreement of the equilibrium and kinetic pathway. In this regard, the current results provide a structural description of the events along the folding pathway. There is some subjectivity in the use of this NMR methodology to monitor protein folding. Some of the backbone amide resonances were not observable at low urea concentrations, and others became obscured by the emerging peaks for the unfolded protein. The interpretations of peak intensities are intended to be qualitative rather than specifically quantitative. Hydrogen−deuterium exchange rates involved a 10 min dead time before measurement and are not corrected for intrinsic effects (*44*). On the other hand, the overall interpretation given here is probably a reasonable picture of the refolding pathway for a β-sheet protein and emphasizes the importance of hydrophobic interactions in early steps.

ACKNOWLEDGMENT

We thank Dr. Greg DeKoster for figures of the IFABP structure, Dr. Ira Ropson for communicating unpublished results, Dr. Keehyuk Kim for preparation of the protein, and Drs. David Cistola and Ira Ropson for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Relative peak intensities for all residues and rate constants for hydrogen−deuterium exchange as a function of urea concentration (Tables 1 and 2) and peak volumes determined by two different methods for selected residues as described in the Results (Table 3). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Ropson, I. J., Gordon, J. I., and Frieden, C. (1990) *Biochemistry 29*, 9591−9599.
2. Ropson, I. J., and Frieden, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7222−7226.
3. Kim, K., and Frieden, C. (1998) *Protein Sci. 7*, 1821−1828.
4. Kim, K., Ramanathan, R., and Frieden, C. (1997) *Protein Sci. 6*, 364−372.
5. Ropson, I. J., and Dalessio, P. M. (1997) *Biochemistry 36*, 8594−8601.
6. Burns, L. L., Dalessio, P. M., and Ropson, I. J. (1998) *Proteins 33*, 107−118.
7. Clark, P. L., Liu, Z. P., Zhang, J. H., and Gierasch, L. M. (1996) *Protein Sci. 5*, 1108−1117.
8. Clark, P. L., Liu, Z. P., Rizo, J., and Gierasch, L. M. (1997) *Nat. Struct. Biol. 4*, 883−886.
9. Dalessio, P. M., and Ropson, I. J. (1998) *Arch. Biochem. Biophys. 359*, 199−208.
10. Hodsdon, M. E., Ponder, J. W., and Cistola, D. P. (1996) *J. Mol. Biol. 264*, 585−602.
11. Hodsdon, M. E., and Cistola, D. P. (1997) *Biochemistry 36*, 1450−1460.
12. Hodsdon, M. E., and Cistola, D. P. (1997) *Biochemistry 36*, 2278−2290.
13. Scapin, G., Gordon, J. I., and Sacchettini, J. C. (1992) *J. Biol. Chem. 267*, 4253−4269.
14. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
15. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson., Ser. B 103*, 203−216.
16. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
17. Messerle, B. A., Wider, G., Otting, G., Weber, C., and Wuthrich, K. (1989) *J. Magn. Reson. 85*, 608−613.
18. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
19. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
20. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
21. Shaka, A. J., Lee, C. J., and Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
22. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry 27*, 8063−8068.
23. van Mierlo, C. P., van den Oever, J. M., and Steensma, E. (2000) *Protein Sci. 9*, 145−157.
24. Kugel, H. (1991) *J. Magn. Reson. 91*, 179−185.
25. Plaxco, K. W., Morton, C. J., Grimshaw, S. B., Jones, J. A., Pitkeathly, M., Campbell, I. D., and Dobson, C. M. (1997) *J. Biomol. NMR 10*, 221−230.
26. Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H., and Clore, G. M. (1993) *Science 260*, 1110−1113.
27. Schulman, B. A., Kim, P. S., Dobson, C. M., and Redfield, C. (1997) *Nat. Struct. Biol. 4*, 630−634.
28. Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) *Nature 390*, 196−199.
29. Neri, D., Billeter, M., Wider, G., and Wuthrich, K. (1992) *Science 257*, 1559−1563.
30. Evans, P. A., Topping, K. D., Woolfson, D. N., and Dobson, C. M. (1991) *Proteins 9*, 248−266.
31. Mok, Y. K., Alonso, L. G., Lima, L. M., Bycroft, M., and de Prat-Gay, G. (2000) *Protein Sci. 9*, 799−811.
32. Wong, K. B., Freund, S. M. V., and Fersht, A. R. (1996) *J. Mol. Biol. 259*, 805−818.
33. Aurora, R., and Rose, G. D. (1998) *Protein Sci. 7*, 21−38.
34. Aurora, R., Srinivasan, R., and Rose, G. D. (1994) *Science 264*, 1126−1130.
35. Aurora, R., Creamer, T. P., Srinivasan, R., and Rose, G. D. (1997) *J. Biol. Chem. 272*, 1413−1416.
36. Sukumar, N., and Gierasch, L. M. (1997) *Folding Des. 2*, 211−222.
37. Steele, R. A., Emmert, D. A., Kao, J., Hodsdon, M. E., Frieden, C., and Cistola, D. P. (1998) *Protein Sci. 7*, 1332−1339.
38. Kim, K., Cistola, D. P., and Frieden, C. (1996) *Biochemistry 35*, 7553−7558.
39. Finucane, M. D., and Jardetzky, O. (1996) *Protein Sci. 5*, 653−662.
40. Clarke, J., Itzhaki, L. S., and Fersht, A. R. (1997) *Trends Biochem. Sci. 22*, 284−287.

41. Hoeltzli, S. D., and Frieden, C. (1996) *Biochemistry 35*, 16843−16851.
42. Hoeltzli, S. D., and Frieden, C. (1998) *Biochemistry 37*, 387−398.
43. Staniforth, R. A., Dean, J. L., Zhong, Q., Zerovnik, E., Clarke, A. R., and Waltho, J. P. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5790−5795.
44. Milne, J. S., Mayne, L., Roder, H., Wand, A. J., and Englander, S. W. (1998) *Protein Sci. 7*, 739−745.
45. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.

BI001518I