BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

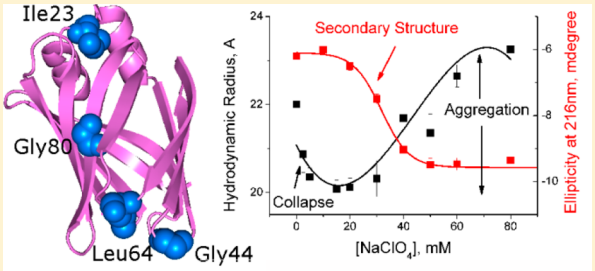
Studies of Early Events of Folding of a Predominately β-Sheet Protein Using Fluorescence Correlation Spectroscopy and Other Biophysical Methods

Suparna Sarkar and Krishnananda Chattopadhyay*

Protein Folding and Dynamics Laboratory, Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, 4, Raja S. C. Mullick Road, Kolkata, India

Supporting Information

ABSTRACT: The interplay between the early collapse of the unfolded state and the formation of the secondary structure has been the subject of extensive research in protein chemistry. In this study, we used the intestinal fatty acid binding protein (IFABP), a small model protein with predominately β-sheet structure, to study the early events, including the early chain collapse and the formation of the secondary structure. We used a combination of fluorescence correlation spectroscopy and far-UV circular dichroism (CD) to understand how these early processes influence the late folding events like the stabilization of the secondary structure and aggregation. Acid-induced unfolded IFABP was found to collapse in the presence of low concentrations of added salt and aggregate at higher concentrations. Both the formation of the collapsed state and aggregation were conveniently probed by fluorescence correlation spectroscopy, a sensitive fluorescence technique with single-molecule resolution. In contrast, the formation of the secondary structure was monitored by far-UV CD. The results suggested that backbone hydrogen bond formation, not only the overall hydrophobicity of IFABP, may play crucial roles in the early collapse. Two mutant proteins positioned at a crucial nucleating site, namely, G80V and L64G, although being opposite in their overall hydrophobicity, collapsed relatively rapidly compared to the wild-type protein. The interconnection among the early collapse, the formation of the secondary structure, and aggregation was similar for these two mutants. Another mutant, G44V, which was identical in its overall hydrophobicity to G80V but situated in a region distant from the hydrophobic core, was found to be very different from G80V and L64G.



During folding, a protein needs to search different possible conformations to reach into its native functional state. A comprehensive understanding of the early stages of protein folding remains elusive to date and is a subject of extensive research efforts. The major bottlenecks to studying the early events arise from the lack of computational and experimental techniques for studying the unfolded and initial intermediate states. The early events of protein folding, which are expected to take place on the microsecond time scale, are not readily accessible to ensemble biophysical techniques like nuclear magnetic resonance (NMR) (with a time resolution typically of milliseconds) or time-resolved fluorescence (typically nanoseconds or picoseconds) methods. The use of molecular dynamics simulation is also challenging because large amounts of computational resources are required for these calculations on a microsecond time scale. Fluorescence correlation spectroscopy (FCS) has been emerging as an important technique with single-molecule resolution for studying the diffusional and conformational dynamics of the early events of protein folding.[1,2]

Experimental results obtained by different laboratories, including ours, have shown the presence at least two conformers in the unfolded states of different proteins. One of these conformers is extended, while the other is relatively compact.[3,4] The fluctuations occurring between the extended and collapsed conformers in the unfolded states have also been studied extensively.[5] Although it has been suggested that the formation of the collapsed state occurs because of hydrophobic interactions between nonpolar residues, exciting debates with many unanswered questions exist. In a recent publication, we have shown using cytochrome *c* from *Saccharomyces cerevisiae* that the interconnections between the early collapse and the formation of the secondary structure depend on solution conditions.[4] For cytochrome *c* and in aqueous solution, the hydrophobic collapse has been shown to occur first, which is followed by the formation of the secondary structure. In contrast, in the presence of urea, these events take place simultaneously. In an exciting recent development, it has been suggested that backbone-induced hydrogen bond formation can also facilitate an early collapse.[6] In this model, both the backbone and side chains of a protein play a combined role in its search for the native folded state.

Received: October 31, 2013
Revised: February 19, 2014
Published: February 20, 2014

dx.doi.org/10.1021/bi4014837 | *Biochemistry* 2014, 53, 1393−1402

In this paper, we used the rat intestinal fatty acid binding protein (IFABP) as a model system. IFABP belongs to the intracellular lipid binding protein (iLBP) family. Fatty acid binding proteins are transport proteins delivering fatty acids (FAs) from the cellular periphery to their subcellular destinations.[7] They have been shown to play important roles in lipid metabolism and organelle dynamics.[8,9] Some members of the FABP family are also known to behave like transcription factors.[10] The structure of IFABP was determined by X-ray crystallography.[11] Because of its low molecular mass (15000 Da), its predominately β-sheet structure (Figure 1), and the lack of cysteine or proline residues, IFABP has been an excellent model system for protein folding studies using both equilibrium and kinetic techniques.[12] For example, Ropson et al. have studied the unfolded state of IFABP using the ^{19}F NMR method.[13] A combination of stopped flow and continuous flow measurements on the wild type (WT)[14] and several mutants of IFABP[15] suggested the formation and stabilization of secondary structures of IFABP occurring in two kinetic events separated by a relatively large kinetic barrier. It has also been shown that the binding of fatty acids occurs only with the folded protein.[15]

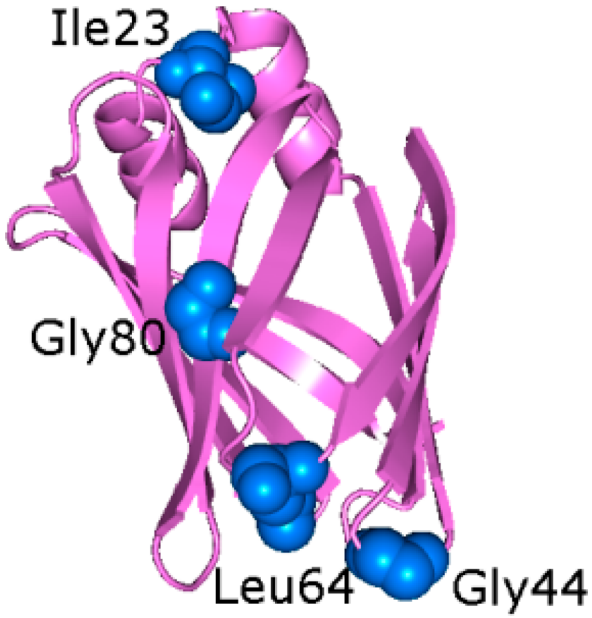


Figure 1. Crystal structure of IFABP. The positions of Gly44, Gly80, and Leu64 are highlighted and labeled. Residue Ile23, which was mutated to a Cys for the purpose of labeling, is also shown.

In this work, we studied the initial collapse and secondary structure formation of WT and a number of different mutants of IFABP (Figure 1 shows the structure of the protein) using fluorescence correlation spectroscopy (FCS) and far-UV circular dichroism (CD). It is interesting to note that the β-sheet proteins have been predicted to result in more pronounced collapse.[16] Nevertheless, experimental studies of the collapse of β-sheet proteins are limited. One of the main objectives of this study is to experimentally study the early collapse of IFABP to decipher its implications in the folding of a β-sheet protein in general and IFABP in particular. Because β-sheet intermediates are proposed to be a common factor in the aggregation of multiple proteins, we would also aim to determine if the early events (early collapse and associated formation of the secondary structure) have any relevance in the late stages of protein folding, including folding defects like aggregation. Although this connection has been largely unexplored in the case of globular proteins, the transition between the extended and collapsed conformers and its implication for protein aggregation have been studied extensively with α-synuclein, a natively unfolded protein that is relevant to Parkinson's disease (PD).

Our results showed that the addition of sodium perchlorate ($NaClO_4$) at pH 2 induced a large decrease in the hydrodynamic radius (r_H) of the IFABP proteins. As a result, the proteins regained partial secondary structure. The nature of the mutants played profound roles in the relative occurrence of chain collapse and secondary structure formation. WT IFABP showed chain collapse, which was followed by the formation of the secondary structure. G44V, a mutant with an increased hydrophobicity but situated distant from the hydrophobic core of the protein, also showed this behavior. Another mutant, L64G, located in the hydrophobic core, showed very different behavior. In this mutant, chain collapse took place simultaneously with the formation of secondary structure. A mutant with greater hydrophobicity, G80V, located at the hydrophobic core, showed properties similar to those of L64G and was found to be different from the WT. We found that G80V and L64G while forming the collapsed state more rapidly reached the final folded state relatively slowly compared to the WT and G44V. G80V and L64G also showed aggregation propensities stronger than those of the WT and G44V. The results of this work indicated a prominent role of the backbone hydrogen bond in the early collapse and implicated a complex interplay among the chain collapse, the formation of the secondary structure, and aggregation.

MATERIALS AND METHODS

Chemicals and Reagents. Primers for subcloning and mutagenesis were obtained from Integrated DNA Technologies (Coralville, IA). All chemicals for *in vitro* biophysical studies were purchased from Sigma Chemical Co. (St. Louis, MO). Alexa488Maleimide (Alexa488) was purchased from Molecular Probes (Eugene, OR).

Subcloning of Recombinant Rat IFABP and Its Expression and Purification. Cloned plasmid pQE-80L containing recombinant rat IFABP was a kind gift from the laboratory of C. Frieden (Washington University School of Medicine, St. Louis, MO). The DNA was subcloned into a pET21c (Novagen, Madison, WI) plasmid vector containing six His residues at the end for affinity column purification of the IFABP protein through a Ni-NTA column. Subcloning was conducted using NheI and XhoI (New England BioLabs, Ipswich, MA) as the 5′ and 3′ restriction enzymes, respectively, following a standard molecular biology protocol. Primers used

Table 1. Primers Used for the Subcloning of IFABP into the pET21c Vector

construct	forward primer (5′ → 3′)	reverse primer (5′ → 3′)
pET21c-IFABP	GGCTAGCATGGCATTTGATGGC	CGCTCGAGTTCCTTCTTAAAGATCCG

dx.doi.org/10.1021/bi4014837 | *Biochemistry* 2014, 53, 1393−1402

Table 2. Primers Used for Site-Directed Mutagenesis

mutant	forward and reverse primers (5′ → 3′)
I23C	forward primer (5′ → 3′)
	GTTCATGGAGAAAATGGGCTGTAACGTGGTGAAGAGGAAG
	reverse primer (5′ → 3′)
	CTTCCTCTTCACCACGTTACAGCCCATTTTCTCCATGAAC
G44V	forward primer (5′ → 3′)
	GAAACTGACGATCACACAGGAAGTAAATAAATTCACAGTCAAAGAAT
	reverse primer (5′ → 3′)
	ATTCTTTGACTGTGAATTTATTTACTTCCTGTGTGATCGTCAGTTTC
G80V	forward primer (5′ → 3′)
	GGAACAGAACTCACTGTGACCTGGACCATGGAG
	reverse primer (5′ → 3′)
	CTCCATGGTCCAGGTCACAGTGAGTTCTGTTCC
L64G	forward primer (5′ → 3′)
	CCGAAACATTGATGTTGTGTTTGAAGGCGGCGTCGACTTT
	reverse primer (5′ → 3′)
	AAAGTCGACGCCGCCTTCAAACACAACATCAATGTTTCGG

for this subcloning are listed in Table 1. The sequence of the resulting cloned IFABP construct was verified.

Site-Directed Mutagenesis. Site-directed mutagenesis was conducted using the QuickChange Lightning kit (Stratagene, La Jolla, CA) following the manufacturer's protocol. All mutations were confirmed by sequencing. Various mutants and the primers used for the mutations are listed in Table 2. It needs to be mentioned that, because our subcloned IFABP gene contained a Met residue at its 5′ end right after the NheI restriction site, the usual nomenclature of the residues was shifted one place to the right; i.e., the G44V residue was read as G45V, G80V as G81V, etc. In this work, we shall follow the usual nomenclature to avoid any difference in the identification of the mutants with respect to the literature.

The sequence-verified WT and mutant IFABP plasmids were grown and expressed in *Escherichia coli* strain Bl21-DE3 and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (AMRESCO, Solon, OH) using previously mentioned procedures[17] with the following slight modifications.

Purification of the Protein from the Supernatant (for the WT and G44V mutant). The supernatant was stirred at 4 °C while ammonium sulfate was being added until 60% saturation had been reached for 2 h. The resulting precipitate was removed by centrifugation at 13000 rpm for 30 min at 4 °C. The supernatant was transferred to dialysis tubing [molecular weight (MW) cutoff of 3500, Snakeskin (Thermo Scientific)] and dialyzed against 20 mM sodium phosphate buffer (pH 7.5) at 4 °C with multiple changes. The dialyzed protein was then allowed to bind to pre-equilibrated Ni-NTA column material [equilibration conducted using wash buffer containing 50 mM sodium dihydrogen phosphate, 300 mM sodium chloride, and 20 mM imidazole (pH 8.0)] at 4 °C for 1.5 h. Then the protein-bound column material was loaded onto an empty column and eluted with wash buffer containing 250 mM imidazole. The protein-containing fractions from elution were checked via sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE), and they were pooled down together in a dialysis tubing with a 3500 MW cutoff and subjected to dialysis against 20 mM sodium phosphate buffer (pH 7.5) at 4 °C with multiple changes. The dialyzed protein was collected and run through a Lipidex-1000 column [pre-equilibrated with 20 mM sodium phosphate buffer (pH 7.5)] to delipidify the protein. This final purified protein was homogeneous as checked by Coomassie Blue staining of the SDS−PAGE gel.

Purification of the Protein from Inclusion Bodies (for the L64G and G80V mutants). The pellet, containing inclusion bodies, was washed five times with buffer A [20 mM Tris-HCl (pH 8.0) and 0.25 mM EDTA]. The inclusion body was dissolved in 6 M guanidinium hydrochloride (Gdn) in buffer A, and the insoluble material was removed by centrifugation. The unfolded protein in the supernatant was refolded by 20-fold dilution in buffer A (with 2 mM DTT in the case of cysteine-containing mutants) while the mixture was being stirred constantly. Any additional aggregates were removed by centrifugation. The supernatant was dialyzed against 20 mM sodium phosphate buffer (pH 7.5). The homogeneity of the protein was verified by Coomassie Blue staining of the SDS−PAGE gel.

Far-UV CD Experiments. Far-UV CD experiments were conducted using a Jasco J720 spectropolarimeter (Japan Spectroscopic Ltd.). All experiments were conducted with a protein concentration of 10 μM using a cuvette with a path length of 1 mm at 25 °C. Far-UV CD measurements were recorded between 200 and 250 nm, and an average of 10 repeat scans was considered. Ellipticity at 216 nm was recorded for the unfolding data analyses.

Steady State Tryptophan Fluorescence Experiments. Steady state tryptophan fluorescence spectroscopy was conducted using a PTI fluorimeter (Photon Technology International). For all measurements, a 1 cm path length cuvette was used. For tryptophan fluorescence, an excitation wavelength of 295 nm was used to eliminate any contribution from the tyrosine residues. Emission spectra between 305 and 400 nm were recorded in triplicate for each measurement. The typical protein concentration used for the tryptophan fluorescence experiments was 3 μM. Necessary background corrections were made for each experiment.

Dynamic Light Scattering (DLS) Experiments. Dynamic light scattering experiments were conducted using a Nano-ZS (Malvern Instruments, Worcestershire, U.K.) at 25 °C (5 mW, He−Ne laser, λ = 632 nm). The operating procedure was programmed using the DTS software supplied with the instrument to record the average of 10 runs. Every run was collected for 30 s, and an equilibration time of 5 min was used.

dx.doi.org/10.1021/bi4014837 | *Biochemistry* 2014, 53, 1393−1402

Fluorescence Correlation Spectroscopy (FCS). Cysteine mutants (Ile23Cys) of the proteins (both the WT and the mutants) were labeled with Alexa488 using the published protocol.[18]

Fluorescence correlation spectroscopy (FCS) experiments were conducted using a Carl Zeiss 510 META Confocor3 LSM setup (Evotech, Jena, Germany) with a 40× water immersion objective (NA 1.2).[19] A 488 nm argon ion laser was used as the excitation source. The laser power was kept at a minimum (approximately 10−30 μW) to minimize photobleaching and other undesired artifacts. The pinhole diameter was properly adjusted at 70 μm, and the number of molecules was maintained between 5 and 10 in the confocal volume. For all FCS experiments, a final protein concentration of 75 nM was used. The experimental solution also contained 20 mM dithiothreitol (DTT) and 1 μM unlabeled protein. The addition of excess unlabeled protein was needed to inhibit adsorption of labeled protein on the surface.

FCS Data Analysis. *Discrete Component Analysis.* If we exclude the contributions of the triplet state, the auto correlation function corresponding to the free diffusion of single component can be defined by the following equation:

$$G(\tau) = 1 + \frac{1}{N} \times \frac{1}{1 + \frac{\tau}{\tau_D}} \times \frac{1}{\left(1 + S^2 \frac{\tau}{\tau_D}\right)^{1/2}} \quad (1)$$

where τ_D and N denote the diffusion time and the average number molecules in the observation volume, respectively. S is the structural parameter defining the ratio between the radius and the height. We have determined the value of S by FCS experiments using the free dye (Alexa488), whose value (typically around 5) was fixed for the analyses of the protein data. For the majority of the correlation function analyses, we did not include data faster than 10 μs to exclude the triplet contribution in the fitting. We have performed two control experiments to justify that. First, we have independently measured the value of the triplet state lifetime of the free dye, Alexa488, using FCS, and the value obtained (typically 2−3 μs) is <10 μs, the cutoff used in our data analyses. Second, we have conducted a few sample random experiments with different mutants, and the data were fit both by using the triplet component (using data beginning at 1 μs) and by excluding the triplet component (using data beginning at 10 μs). We observed similar values of τ_D using both methods.

The value of τ_D obtained by fitting the correlation function is related to the diffusion coefficient (D) of a molecule by the following equation:

$$\tau_D = \frac{\omega^2}{4D} \quad (2)$$

The value of ω, which defines the size of the observation volume, has been calculated using FCS measurements with Rhodamine 6G, whose value of D has been well established (D = 4.2 × 10^{-10} m^2 s^{-1}). The value of the hydrodynamic radius (r_H) can be obtained from D using the Stokes−Einstein formalism.

$$D = \frac{kT}{6\pi\eta r_H} \quad (3)$$

where η is the viscosity and k is the Boltzmann constant.

Maximal Entropy Method (MEM) Analysis. Because different components in discrete component analyses may change each other's contribution in a multicomponent fit, the correlation function data was further analyzed by the maximal entropy method. This method has been recently applied to FCS using the MEMFCS algorithm.[20] In this method, the multicomponent correlation function can be represented by the following equation:

$$G(\tau) = \sum_{i=1}^{n} a_i \left(\frac{1}{1 + \frac{\tau}{\tau_{D_i}}}\right) \times \left(\frac{1}{1 + S^2 \frac{\tau}{\tau_{D_i}}}\right)^{0.5} \quad (4)$$

where n is the number of noninteracting fluorescent species, each of which can have a diffusion time between 0.001 and 500 ms. MEMFCS optimizes the fit by maximizing an entropic quantity $S = \sum_i p_i \ln p_i$.

RESULTS

An analysis of the IFABP sequence using the Expasy ProtParam tool (http://web.expasy.org/protparam/) suggested that the theoretical pK_a of the protein would be 6.62 (Figure 1S, Supporting Information). At pH 2, the protein becomes positively charged (+22.8). Repulsive interactions between the positively charged residues lead to unfolding at pH 2. The addition of salt at pH 2 leads to an overall increase in the dielectric constant of the medium. The charge screening effect of the added salt at pH 2 leads to the contraction of the protein chain. The addition of salt at pH 2 has been shown to result in a decrease in the hydrodynamic radius (r_H) and partial recovery of the secondary structure of different proteins.

For the measurements of the secondary structure, we used far-UV CD. The far-UV CD of WT IFABP showed a minimum at 216 nm, which was characteristic of a β-sheet protein (Figure 2a). Decreasing the pH of the protein to 2 decreased the ellipticity at 216 nm, suggesting significant unfolding of the protein (Figure 2a). The addition of 100 mM sodium perchlorate (pH 2) resulted in an increase in the ellipticity at 216 nm (Figure 2a).

For the measurements of r_H, we conducted FCS experiments using Alexa488-labeled protein. For this purpose, we inserted a solvent-exposed cysteine residue (I23C), which was labeled with Alexa488. Neither the additional single-site mutation nor the labeling by Alexa488 resulted in any significant effect on the secondary structure of the mutants as judged by far-UV CD experiments (Figure 2S, Supporting Information). The correlation functions obtained by FCS experiments with the Alexa488-labeled I23C mutant of IFABP (Alexa488IFABP) at pH 2 in the absence and presence of sodium perchlorate are shown in Figure 2b. The correlation function data were fit successfully to a model, which assumed the diffusion of a single component. The goodness of the correlation function fit was verified further using the randomness of the residual distribution (Figure 2b). The addition of a second diffusion component did not improve the quality of the fit.

Because the data presented in this paper correspond to different events, collapse, secondary structure formation, and aggregation (see below), all these events may contribute to the observed values of τ_D. Because each component in the discrete component analyses used in FCS may affect other components, we have used the maximum entropy method (MEM) as an additional way of analyzing the FCS data. Because the MEM is essentially a model free method, it provides an alternative and potentially bias-minimized method for analyzing FCS data.

dx.doi.org/10.1021/bi4014837 | *Biochemistry* 2014, 53, 1393−1402

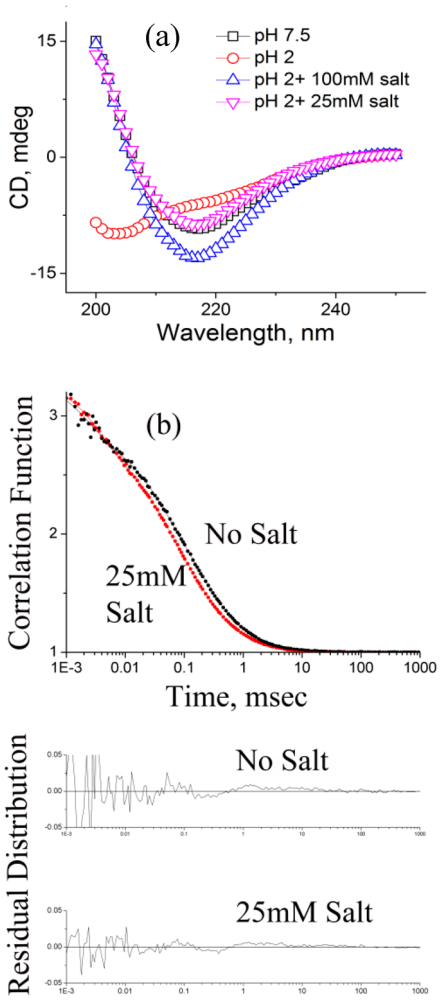


Figure 2. (a) Far-UV CD spectra observed with IFABP at pH 7.5, at pH 2, and in the presence of salt at pH 2. (b) Representative correlation functions obtained by FCS experiments using Alexa488-Maleimide-labeled WT IFABP at pH 2 in the absence (black) and presence (red) of salt. The correlation functions, which include data starting from 1 μs, were fit using a one-component diffusion model that included an additional triplet component. The residual distributions of the fit are also shown.

MEM analyses of Alexa488IFABP at pH 2 also showed a single component (shown later in Figure 5d,e for two different mutants of Alexa488IFABP).

Using the value of diffusion time (τ_D) observed with WT IFABP, the hydrodynamic radius of the protein was calculated at pH 2. Both discrete component and MEM analyses provided similar values of τ_D. A potential caveat of this calculation comes from the inherent assumption of the Stokes–Einstein formalism that the diffusing particle (or the protein unfolded at pH 2 in this case) is spherical. Nevertheless, the value of r_H obtained at pH 2 via our measurement (21 Å) matched well with previous experimental data (22 Å).[1] The addition of sodium perchlorate shifted the correlation function, resulting in a value of r_H decreased to 18 Å (Figure 2b). This was expected, because the addition of salt leads to the collapse and contraction of the unfolded chain (see above). The value of r_H observed with the folded protein in the presence of 20 mM sodium phosphate buffer (pH 7.5) is identical to the previously reported value of 16.5 Å.[1]

Figure 3a shows the salt dependence of the increase in the ellipticity at 216 nm and the change in r_H as observed with the

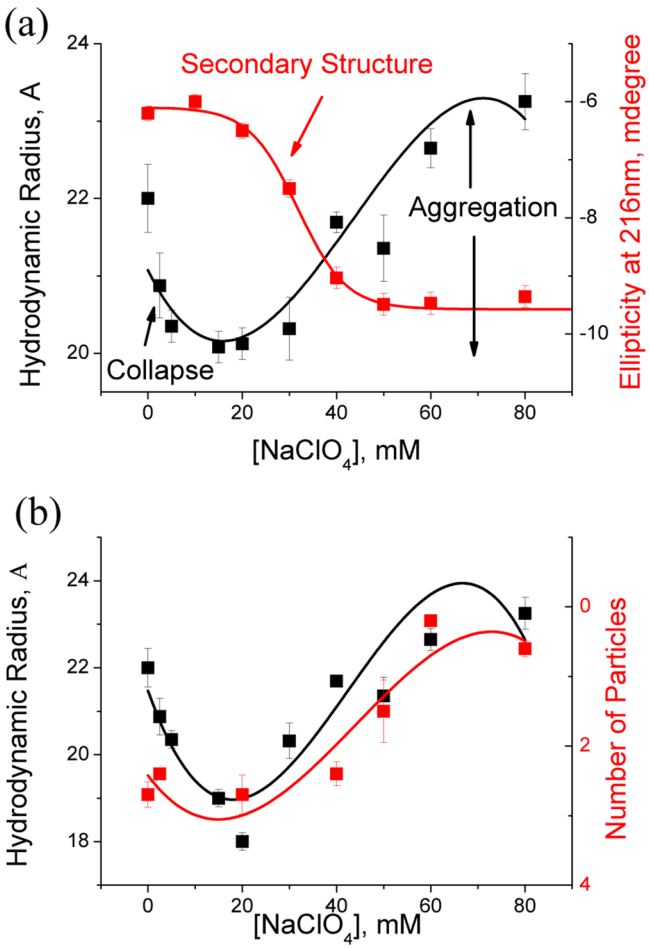


Figure 3. (a) Variation of r_H (black) and far-UV CD intensity at 216 nm (red) with the increase in salt concentration. These experiments were conducted with WT IFABP at pH 2. The variation in r_H showed an early collapse, which was followed by aggregation. (b) Variation in r_H and the number of particles with salt concentration. The lines drawn through the data are used to show the trend and do not correspond to any mathematical modeling.

FCS data at pH 2. The increase in the level of secondary structure (monitored by the ellipticity at 216 nm) occurred in a single sigmoidal step as the salt concentration was increased. The salt-induced change in r_H, in contrast, took place in multiple steps. At a low salt concentration, r_H decreased, indicating the compaction of the protein chain as a result of chain collapse. Figure 3a shows that the decrease in r_H became saturated at ~20 mM $NaClO_4$, and then r_H increased at higher salt concentrations. We verified that this increase in r_H that occurred at high $NaClO_4$ concentrations corresponded to an event of IFABP self-association. This is the case for three reasons. (a) The number of particles (N) decreased in the presence of $NaClO_4$, and the decrease in N was simultaneous with the increase in r_H (Figure 3b). It can be noted that the loss of protein particles due to adsorption at the glass surface was minimized using excess unlabeled protein in the experimental solution (see Materials and Methods). Protein association and aggregation have been shown to decrease N and increase τ_D.[21] (b) The intensity fluctuation data in the presence of very high salt concentrations indicated the presence of large spikes, which

dx.doi.org/10.1021/bi4014837 | Biochemistry 2014, 53, 1393–1402

corresponded to particles with many fluorophores attached (as a result of the formation of large aggregates) (Figure 3S, Supporting Information). It can be noted that the number of spikes was used as a diagnostic tool to detect the aggregated amyloid β in Alzheimer's patients.[22] (c) Independent dynamic light scattering experiments showed the formation of high-molecular weight species in the presence of salt (Figure 4S, Supporting Information). The difference in the r_H of the self-associated species observed in the FCS data (around 2−3 nm) and the large aggregates obtained by the dynamic light scattering experiments (~300 nm) was possibly caused by the use of a relatively high salt concentration (100 mM) in the dynamic light scattering measurements. The correlation functions observed by FCS in the presence of a high salt concentration (e.g., 100 mM) were found to be complex and showed the presence of long components, which were difficult to fit (see below). Figure 3a, which compares the far-UV CD and FCS data of WT IFABP, clearly shows that the chain collapse (monitored by the change in r_H) and the formation of the secondary structure of the WT protein did not occur concomitantly. In contrast, the data suggested early collapse of the unfolded chain that was followed by the formation of the secondary structure. This was identical to our previous observation of early collapse followed by the formation of secondary structure in the case of salt-induced collapse of cytochrome *c* from *S. cerevisiae* at pH 2.[4]

From the results shown in Figure 3a, we identified different events of WT IFABP folding, namely, chain collapse, secondary structure formation, and protein aggregation. The next question we wanted to address was whether these events (chain collapse, secondary structure formation, and aggregation) have any influence on each other. For this purpose, we created a number of single-site mutants. These mutants were chosen on the basis of two considerations: (1) their roles in IFABP structure and (2) their effects on IFABP folding based on the existing literature. From the crystal structure of IFABP, which is shown in Figure 1, a hydrophobic core could be identified (Figure 4a−c). Residue Gly80 (colored red in Figure 4a), which is the part of a critical half-turn, can be identified as the key residue in the early hydrophobic collapse (Figure 4a). Gly80 resides inside the cluster of many hydrophobic residues, including Val60, Phe62, Tyr70, Phe78, and Trp82 (colored green in Figure 4a). The G80V mutant would hence increase the hydrophobicity of this core further. The G80V mutation has been shown to have a slow refolding rate and rollover behavior in the folding arm of the chevron plot.[15,23] The second mutation, L64G, is also important (Figure 4b). With L64G, the late event of folding has been found to be absent.[17] It has been established that this late event is responsible for the formation of key hydrogen bonds and for the stabilization of secondary structure.[15] L64G was shown to affect the stability of the protein.[17,23] In contrast, the third mutation, G44V, is the replacement of a glycine residue (Gly44), which is quite distant from the hydrophobic core (Figure 4c). Gly44 is well conserved in the iLBP family,[24] and hence, the G44V mutation results in a protein with very weak fatty acid binding.[23] However, the thermodynamic stability of the G44V mutant is almost identical to that of the WT protein.[23] In our study, we used these three mutants to study the salt-induced chain collapse and the formation of secondary structure at pH 2.

FCS experiments examining the salt-induced collapse were conducted using the Alexa488-labeled G80V, G44V, and L64G mutants. Data were analyzed by fitting the autocorrelation

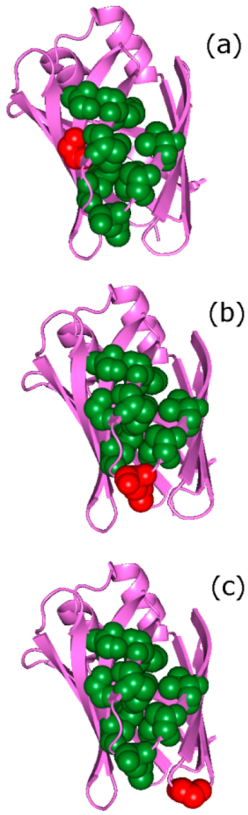


Figure 4. IFABP mutants (a) G80V, (b) L64G, and (c) G44V. The mutated residues are colored red. The identified hydrophobic clusters are colored green.

functions to a single-diffusion component model as mentioned before for all these proteins. In addition, the salt-induced formation of the secondary structure for these mutants was studied using far-UV CD at pH 2. Panels a−c of Figure 5 show the variations of r_H and the ellipticity at 216 nm with $NaClO_4$ concentration at pH 2 for the G44V, G80V, and L64G mutants, respectively. Panels d and e of Figure 5 show the change in MEM profiles at pH 2 in the presence of different salt concentrations for the G44V and L64G mutants, respectively, as representative examples.

The data for the G44V mutant (shown in Figure 5a), which represents a region distant from the hydrophobic core, also showed the occurrence of early collapse, which was followed by the formation of secondary structure (monitored by the ellipticity at 216 nm). The decrease in r_H (or τ_D) at low salt concentrations was also confirmed for G44V using MEM analysis (Figure 5d).

FCS and far-UV CD spectroscopy showed different behaviors with G80V (Figure 5b) and L64G (Figure 5c). This behavior is particularly prominent with L64G, which showed complete overlap between the collapse and the formation of secondary structure. Although complete overlap (as observed with L64G) was not observed with G80V, the gap between the collapse and the formation of secondary structure is smaller with this mutant than with the WT and G44V. In addition, for both G80V and L64G, the decrease in r_H (early collapse) occurred much earlier (at a lower salt concentration) than that observed with the WT and G44V. For the WT, the decrease in r_H became saturated at 30 mM salt, while for G44V, G80V, and L64G, the saturation concentrations were 50, 16, and 14 mM, respectively. Figure 5e shows the MEM profile of

dx.doi.org/10.1021/bi4014837 | *Biochemistry* 2014, 53, 1393−1402

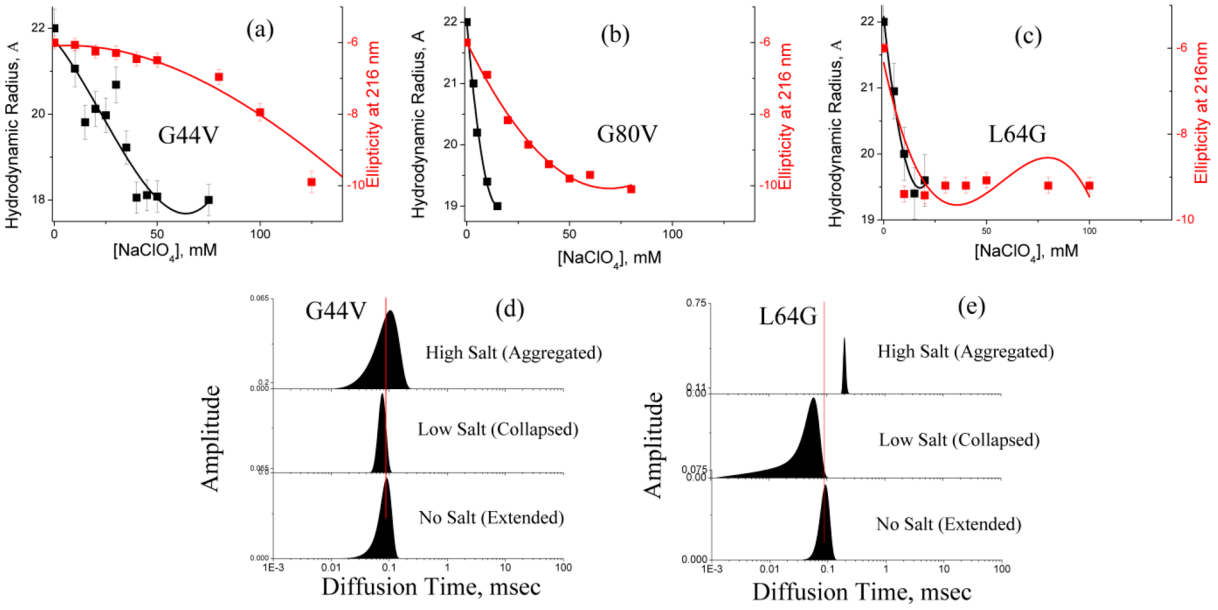


Figure 5. Relative occurrences of chain collapse (black) and secondary structure formation (red) for (a) G44V, (b) G80V, and (c) L64G. The lines drawn through the data are used to show the trend and do not correspond to any mathematical modeling. The changes in MEM profiles with salt concentration for (d) G44V and (e) L64G are shown. Red vertical lines are used in panels d and e to show the shifts in MEM profiles due to collapse and aggregation.

L64G, which also shows the decrease in τ_D in the presence of a low salt concentration.

As mentioned before, the value of r_H increased in the presence of a high $NaClO_4$ concentration, indicating aggregation of the protein. A comparison between the increase in r_H occurring at a high salt concentration clearly suggested that the extent of aggregation was significantly higher for mutants G80V and L64G than for the WT and G44V (Figure 6a). It may be noted that the correlation functions observed in the presence of high salt concentrations (≥55 mM) with the mutants with higher levels of aggregation (G80V and L64G) deviated from the single-diffusion fits. Figure 6b shows an example with G80V in the presence of 55 mM sodium perchlorate. Under these conditions, the data were fit to a two-component diffusion model with a fast (monomer diffusion) and a slow diffusion (the diffusion of the aggregated molecule). Although a more accurate model would be to use a combination of brightness-weighted components to account for the larger contribution of the aggregated molecules to the correlation functions,[25] the model presented here should be sufficient to suggest, at least qualitatively, that the extent of aggregation is significantly greater for G80V and L64G mutants. Panels d and e of Figure 5, which show the variations of MEM profiles of G44V and L64G, respectively, as a function of $NaClO_4$ concentration, also show that the value of τ_D is much larger for L64G in the presence of a high salt concentration than that observed with G44V. The larger extent of aggregation was further confirmed by a comparatively large number of fluorescent spikes in FCS data in the presence of high salt concentrations for L64G and G80V than for the WT and G44V (Figure 3cS, Supporting Information).

It may be important to consider these data in light of some of the recent kinetic measurements conducted with these mutants. While this paper deals with the simultaneous or sequential formation of the collapsed state and secondary structure, direct measurements of the early kinetics come from the continuous flow studies reported for different IFABP mutants, which showed that the rate of initial burial of W82 in the hydrophobic core is higher for G80V than for the WT and other turn mutants.[15] Besides, the kinetics of collapse and secondary structure formation were measured with another single-chain β-sheet protein, monellin, by the combined use of continuous flow small-angle X-ray and far-UV CD spectroscopy.[26] The results with monellin also suggested the early formation of the collapsed state that was followed by the formation of the secondary structure. The slow kinetics of the formation of the final folded states of WT and these mutants has been reported recently.[15,23] The rate of the slowest step, measured using the stopped flow fluorescence methods and reported as an extrapolation to a zero guanidinium hydrochloride concentration (k_f^0), was found to be 75 s^{-1}.[15] While the value of k_f^0 for G44V was similar (33 s^{-1}) to that of WT IFABP, the value of k_f^0 was found be significantly lower (0.7 s^{-1}) for G80V.[15] More interestingly, this step was found to be absent for L64G ($k_f^0 \sim 0$ s^{-1}).[15] For this mutant, key hydrogen bond contacts do not form and the protein does not complete its folding. As a result, the kinetics for its slow step of folding has been found to be absent.[17]

DISCUSSION

Although traditional protein folding studies generally assumed the unfolded state as an extended random coil, a large number of theoretical and experimental studies have changed that view. The presence of different conformers in the unfolded states has been shown using multiple experimental techniques, including X-ray scattering, fluorescence resonance energy transfer, and FCS.[27] It has also been found that the equilibrium distribution of the extended and collapsed isomers of the unfolded state can vary in the presence of protecting or denaturing osmolytes.[4] However, controversies regarding the nature of the compact isomer of the unfolded state exist: whether it is similar to the native state of the protein representing the molten globule intermediate or if it is a compact globule without any appreciable secondary structure. A question associated with

dx.doi.org/10.1021/bi4014837 | Biochemistry 2014, 53, 1393−1402

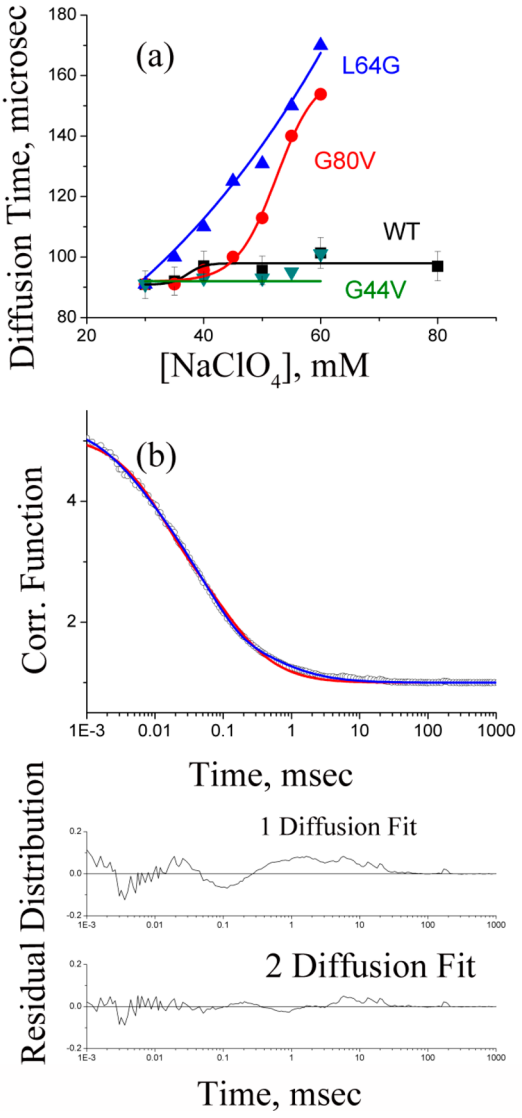


Figure 6. (a) Extent of aggregation observed at high salt concentrations for different IFABP mutants. As mentioned in the text, the increase in r_H was used as a measure of the level of aggregation in the presence of high salt concentrations. (b) Correlation function of G80V in the presence of 55 mM salt. The fit using the one-diffusion component model (red line) deviates from the data, which is confirmed by the residual distribution analyses (shown at the bottom). The data were fit well using the two-component diffusion model (blue line through the data), which is also apparent in the randomness of the residual distribution (shown in the figure).

this debate relates to the nature of the transition between the extended and compact unfolded state: if it is a nonspecific event driven only by the hydrophobic forces or if the amino acid sequence has a role to play. These results showed the different collapse behavior of G44V and G80V mutants, although the overall increase in hydrophobicity (replacing Gly with a Val) is identical for both proteins. In addition, the L64V mutant, in which the overall hydrophobicity actually decreases compared to those of the WT and other valine mutants, showed an earlier collapse. In fact, the overall hydrophobicity of the studied proteins increases in the following order: L64G < WT < G80V ∼ G44V. In contrast, via comparison of the salt concentration that represents the minimum of r_H, the following trend can be considered for the collapse transition of IFABP mutants: G44V < WT < G80V ∼ L64G.

To summarize, we did not find any direct correlation between overall hydrophobicity and collapse behavior in this study. Additionally, the literature showed the formation of the initial collapse for single-chain monellin (containing a lower overall hydrophobicity),[26] although a number of other proteins, including the IgG binding domain of protein L[28] and ubiquitin,[29] did not undergo hydrophobic collapse (although these later proteins were more hydrophobic). All these results rule out the possibility that the early collapse is entirely hydrophobic, suggesting a prominent role of hydrogen bonding and other factors related to protein sequence. It has recently been suggested that backbone hydrogen bonding can play a pivotal role in early collapse.[30,31] The direct experimental evidence supporting this hypothesis comes from a FCS study of N-methylated polyglycine chains.[6] Simulations with polyglycine chains have also demonstrated the formation of highly collapsed states induced by backbone hydrogen bonding.[32] Presumably, the residual structure of the unfolded state, which is sequence-dependent, may have a significant role to play in the early events. The presence of the strong and pronounced bias toward secondary structure has been shown in protein sequence.[33]

NMR experiments with IFABP showed the presence of transient but persistent residual structures around the D−E turn of IFABP, suggesting key roles of the amino acids around this region in the initiation of IFABP folding.[13,34] The dynamics of the unfolded state of IFABP in the presence of salt at pH 2 has been found to have a time constant of 8 μs,[35] which is comparable to the value of 3.5 μs observed in a laser temperature jump experiment to study β-hairpin formation of the 16-residue C-terminal fragment of protein GB1.[36] The time constant (1 μs) for the formation of a 10-residue loop has been calculated semiempirically.[37] All these observations can lead to speculation that the hydrophobic collapse and some specific structure formation at and around the D−E turn possibly happen with similar time constants. The initial formation of short-range contacts (at and around the time of hydrophobic collapse) can be followed by the formation of medium- and long-range contacts to form the overall secondary structure in a hierarchical manner. Any change in the specificity of this important initiation region (around the D−E turn) either with an increase in hydrophobicity (G80V) or via an effect on the loop (L64G) may affect both the early and late events. This is because of the presence of misfolded contacts in the unfolded and/or early intermediate states that are not productive for efficient folding. The possibility of the formation of incorrect contacts and wrong turns has been suggested by molecular dynamics simulations and by simulations using LINUS.[23,38] These contacts need to be broken in subsequent events before the right contacts can be established (requiring more time to fold). Alternatively, the protein needs to find different ways to stabilize these misfolded conformations (potentially resulting in aggregation). The delayed late kinetics and higher level of aggregation of these two mutants support this hypothesis.

A number of members of the iLBP family have been extensively studied as model systems for the folding and aggregation research. Although these proteins exhibit a high degree of structural similarity, they are quite diverse in their sequence. For example, two member proteins, namely, ileal lipid binding protein (ILBP) and intestinal fatty acid binding protein (IFABP), have identical structures, while their

dx.doi.org/10.1021/bi4014837 | Biochemistry 2014, 53, 1393−1402

sequences are only 23% identical.[39] The folding mechanism of these two proteins has been shown to be markedly different.[39] In contrast, the folding mechanism of cellular retinoic acid binding protein-1 (CRABP1), another protein of the iLBP family, and IFABP show a similar refolding mechanism. Both proteins form the native folded state (N) through the formation of two intermediate states (I_1 and I_2), although the time constants are different.[40] For CRABP1, intermediate I_1 has been shown to be a collapsed state while intermediate I_2 contains nativelike topology.[40] These FCS data are consistent with the findings described above. Folding kinetics measurements combined with single-point mutational analyses showed that the folding pathway of CRABP1 involves an early event of barrel closure.[40] The barrel closing event has been suggested to be a general feature in the folding of iLBP family, and a necessary strategy for minimizing aggregation risks. It is also noted that mild perturbations in proteins caused by low denaturant concentrations or point mutations may increase the risk of aggregate formation.[41] This is also consistent with our observations for G80V and L64G, in which a small but specific change leads to a large increase in the level of aggregation.

In an earlier study, we showed the simultaneous occurrence of early chain collapse and the formation of secondary structure under solution conditions that hinder the formation of specific contacts. There are extensive debates regarding early contact formation, collapse, and protein misfolding. Much of this research concerns natively unfolded proteins like α-synuclein because of its implications in neurodegeneration. The overall observations from this study show that the interplay among hydrophobic collapse, secondary structure formation, and aggregation is a crucial factor in determining the protein folding pathway. It is important to study all these events in concert at a single-molecule resolution to gain a complete understanding of the complexity of protein folding.

■ ASSOCIATED CONTENT

Supporting Information

Snapshot of the ExPASy ProtParam tool, far-UV CD of both unlabeled and Alexa488-labeled WT IFABP, fluorescence intensity fluctuations in the absence and presence of salt at pH 2, and DLS data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: 011913324995843. E-mail: krish@iicb.res.in.

Funding

S.S. is thankful for the fellowship provided by CSIR, Government of India. K.C. acknowledges the CSIR network project grant, MIND.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Carl Frieden of the Washington University School of Medicine for the IFABP plasmid. We thank two anonymous reviewers, whose suggestions were extremely helpful for the revision of the original manuscript. We thank Prof. Sudipta Maiti of the Tata Institute of Fundamental Research for the MEMFCS software. We also thank Professor Siddhartha Roy (Director, CSIR-Indian Institute of Chemical Biology) for his support.

■ REFERENCES

(1) Chattopadhyay, K., Saffarian, S., Elson, E. L., and Frieden, C. (2002) Measurement of microsecond dynamic motion in the intestinal fatty acid binding protein by using fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. U.S.A. 99*, 14171−14176.

(2) Haldar, S., Mitra, S., and Chattopadhyay, K. (2010) Role of protein stabilizers on the conformation of the unfolded state of cytochrome c and its early folding kinetics: Investigation at single molecular resolution. *J. Biol. Chem. 285*, 25314−25323.

(3) Haran, G. (2012) How, when and why proteins collapse: The relation to folding. *Curr. Opin. Struct. Biol. 22*, 14−20.

(4) Haldar, S., and Chattopadhyay, K. (2012) Interconnection of salt-induced hydrophobic compaction and secondary structure formation depends on solution conditions: Revisiting early events of protein folding at single molecule resolution. *J. Biol. Chem. 287*, 11546−11555.

(5) Ziv, G., Thirumalai, D., and Haran, G. (2009) Collapse transition in proteins. *Phys. Chem. Chem. Phys. 11*, 83−93.

(6) Teufel, D. P., Johnson, C. M., Lum, J. K., and Neuweiler, H. (2011) Backbone-driven collapse in unfolded protein chains. *J. Mol. Biol. 409*, 250−262.

(7) Montoudis, A., Delvin, E., Menard, D., Beaulieu, J. F., Jean, D., Tremblay, E., Bendayan, M., and Levy, E. (2006) Intestinal-fatty acid binding protein and lipid transport in human intestinal epithelial cells. *Biochem. Biophys. Res. Commun. 339*, 248−254.

(8) Darimont, C., Gradoux, N., Persohn, E., Cumin, F., and De Pover, A. (2000) Effects of intestinal fatty acid-binding protein overexpression on fatty acid metabolism in Caco-2 cells. *J. Lipid Res. 41*, 84−92.

(9) Wiederkehr, A., and Wollheim, C. B. (2009) Linking fatty acid stress to β-cell mitochondrial dynamics. *Diabetes 58*, 2185−2186.

(10) Tan, N. S., Shaw, N. S., Vinckenbosch, N., Liu, P., Yasmin, R., Desvergne, B., Wahli, W., and Noy, N. (2002) Selective cooperation between fatty acid binding proteins and peroxisome proliferator-activated receptors in regulating transcription. *Mol. Cell. Biol. 22*, 5114−5127.

(11) 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.

(12) Ropson, I. J., Gordon, J. I., and Frieden, C. (1990) Folding of a predominantly β-structure protein: Rat intestinal fatty acid binding protein. *Biochemistry 29*, 9591−9599.

(13) Ropson, I. J., Boyer, J. A., and Dalessio, P. M. (2006) A residual structure in unfolded intestinal fatty acid binding protein consists of amino acids that are neighbors in the native state. *Biochemistry 45*, 2608−2617.

(14) Yeh, S. R., Ropson, I. J., and Rousseau, D. L. (2001) Hierarchical folding of intestinal fatty acid binding protein. *Biochemistry 40*, 4205−4210.

(15) 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.

(16) Shea, J. E., Onuchic, J. N., and Brooks, C. L., III (2002) Probing the folding free energy landscape of the Src-SH3 protein domain. *Proc. Natl. Acad. Sci. U.S.A. 99*, 16064−16068.

(17) Kim, K., Ramanathan, R., and Frieden, C. (1997) Intestinal fatty acid binding protein: A specific residue in one turn appears to stabilize the native structure and be responsible for slow refolding. *Protein Sci. 6*, 364−372.

(18) Sharma, S., Sarkar, S., Paul, S. S., Roy, S., and Chattopadhyay, K. (2013) A small molecule chemical chaperone optimizes its unfolded state contraction and denaturant-like properties. *Sci. Rep. 3*, 3525.

(19) Ghosh, R., Sharma, S., and Chattopadhyay, K. (2009) Effect of arginine on protein aggregation studied by fluorescence correlation spectroscopy and other biophysical methods. *Biochemistry 48*, 1135−1143.

(20) Sengupta, P., Garai, K., Balaji, J., Periasamy, N., and Maiti, S. (2003) Measuring size distribution in highly heterogeneous systems with fluorescence correlation spectroscopy. *Biophys. J. 84*, 1977−1984.

(21) Nath, S., Meuvis, J., Hendrix, J., Carl, S. A., and Engelborghs, Y. (2010) Early aggregation steps in α-synuclein as measured by FCS and FRET: Evidence for a contagious conformational change. *Biophys. J. 98*, 1302−1311.
(22) Pitschke, M., Prior, R., Haupt, M., and Riesner, D. (1998) Detection of single amyloid β-protein aggregates in the cerebrospinal fluid of Alzheimer's patients by fluorescence correlation spectroscopy. *Nat. Med. 4*, 832−834.
(23) 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.
(24) Marcelino, A. M., Smock, R. G., and Gierasch, L. M. (2006) Evolutionary coupling of structural and functional sequence information in the intracellular lipid-binding protein family. *Proteins 63*, 373−384.
(25) Middleton, E. R., and Rhoades, E. (2010) Effects of curvature and composition on α-synuclein binding to lipid vesicles. *Biophys. J. 99*, 2279−2288.
(26) Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T., Akiyama, S., and Fujisawa, T. (2005) Specific collapse followed by slow hydrogen-bond formation of β-sheet in the folding of single-chain monellin. *Proc. Natl. Acad. Sci. U.S.A. 102*, 2748−2753.
(27) Ziv, G., and Haran, G. (2009) Protein folding, protein collapse, and Tanford's transfer model: Lessons from single-molecule FRET. *J. Am. Chem. Soc. 131*, 2942−2947.
(28) Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R., III, and Baker, D. (1997) Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry 36*, 3373−3382.
(29) Larios, E., Li, J. S., Schulten, K., Kihara, H., and Gruebele, M. (2004) Multiple probes reveal a native-like intermediate during low-temperature refolding of ubiquitin. *J. Mol. Biol. 340*, 115−125.
(30) Bowler, B. E. (2012) Residual structure in unfolded proteins. *Curr. Opin. Struct. Biol. 22*, 4−13.
(31) Bolen, D. W., and Rose, G. D. (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem. 77*, 339−362.
(32) Tran, H. T., Mao, A., and Pappu, R. V. (2008) Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc. 130*, 7380−7392.
(33) Srinivasan, R., and Rose, G. D. (1999) A physical basis for protein secondary structure. *Proc. Natl. Acad. Sci. U.S.A. 96*, 14258−14263.
(34) Hodsdon, M. E., and Frieden, C. (2001) Intestinal fatty acid binding protein: The folding mechanism as determined by NMR studies. *Biochemistry 40*, 732−742.
(35) Chattopadhyay, K., Elson, E. L., and Frieden, C. (2005) The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. *Proc. Natl. Acad. Sci. U.S.A. 102*, 2385−2389.
(36) Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of β-hairpin formation. *Nature 390*, 196−199.
(37) Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615−11617.
(38) Srinivasan, R., and Rose, G. D. (1995) LINUS: A hierarchic procedure to predict the fold of a protein. *Proteins 22*, 81−99.
(39) Dalessio, P. M., and Ropson, I. J. (2000) β-Sheet proteins with nearly identical structures have different folding intermediates. *Biochemistry 39*, 860−871.
(40) Budyak, I. L., Krishnan, B., Marcelino-Cruz, A. M., Ferrolino, M. C., Zhuravleva, A., and Gierasch, L. M. (2013) Early folding events protect aggregation-prone regions of a β-rich protein. *Structure 21*, 476−485.
(41) Ferrolino, M. C., Zhuravleva, A., Budyak, I. L., Krishnan, B., and Gierasch, L. M. (2013) Delicate Balance between Functionally Required Flexibility and Aggregation Risk in a β-Rich Protein. *Biochemistry 52*, 8843−8854.