Structure, Vol. 10, 1173–1185, September, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00804-3

An NMR View of the Folding Process of a CheY Mutant at the Residue Level

Pascal Garcia,[1] Luis Serrano,[2] Manuel Rico,[1] and Marta Bruix[1,3]
[1]Instituto de Química Física "Rocasolano"
CSIC
Serrano 119
28006 Madrid
Spain
[2]European Molecular Biology Laboratory
Meyerhofstrasse 1
Heidelberg D-69012
Germany

Summary

The folding of CheY mutant F14N/V83T was studied at 75 residues by NMR. Fluorescence, NMR, and sedimentation equilibrium studies at different urea and protein concentrations reveal that the urea-induced unfolding of this CheY mutant includes an on-pathway molten globule-like intermediate that can associate off-pathway. The populations of native and denatured forms have been quantified from a series of ^{15}N-^{1}H HSQC spectra recorded under increasing concentrations of urea. A thermodynamic analysis of these data provides a detailed picture of the mutant's unfolding at the residue level: (1) the transition from the native state to the molten globule-like intermediate is highly cooperative, and (2) the unfolding of this state is sequential and yields another intermediate showing a collapsed N-terminal domain and an unfolded C-terminal tail. This state presents a striking similarity to the kinetic transition state of the CheY folding pathway.

Introduction

Describing the various conformations adopted by a polypeptide chain between the unfolded and native states is a key issue in the understanding of the protein folding process. Several models have been proposed to illustrate how proteins fold [1–6]. These models are mainly based on the limited milestones existing in the folding pathway, such as the denatured, transition, intermediate, or native state.

An important insight into the first steps of this process has recently been obtained through the structural characterization of the denatured state of several proteins [7–11]. It has been shown that the denatured state does not necessarily correspond to a highly disordered random coil [12] but can contain transient structured elements that favor the setting up of longer-range interactions, funneling the reaction toward the transition state, as proposed by the nucleation-condensation mechanism [6]. Because of their extremely transient nature, transition states of the folding reaction cannot be characterized by NMR. However, the protein engineering method appeared fruitful for determining the specific residues having native-like structures in the transition state of several proteins [13]. Folding intermediates are the nonnative states whose conformations have been most extensively studied. Recent developments in NMR spectroscopy largely contributed to that. Thus, detailed structure information on kinetic intermediates has been obtained from H/D exchange pulse-labeling experiments [14, 15], as well as from NMR studies of their equilibrium counterparts [16]. The use of new NMR methods has also permitted the detection of long-range NOEs and the structural determination of partially unfolded states [17–20]. Finally, the structure of the native state, located at the opposite end of the folding reaction, can be determined for larger and larger proteins [21].

Monitoring changes in the environment of every single residue accompanying the folding process has been achieved by introducing different site-specific probes into a protein [22, 23]. However, this approach is highly time consuming, and, at the same time, the introduced probes can cause substantial distortion and/or destabilization of the structure. Real-time 1D and 2D NMR [24, 25] have been used to characterize folding reactions at a high structural resolution. A very detailed picture of a ribonuclease T1 folding intermediate has been obtained in this way, showing a high content of secondary and tertiary interactions [26, 27]. Two-dimensional NMR real-time studies of the folding of barstar [28] also showed the structural changes occurring during proline 48 *cis*/*trans* isomerization at the residue level. However, these kinetic studies only apply to relatively slow folding processes and are limited by the poor dispersion of NMR signals of unfolded species, only allowing for the monitoring of some "targeted" residues.

An alternative approach to studying protein folding at the residue level is to record a series of ^{15}N-^{1}H HSQCs at increasing denaturant concentrations under equilibrium conditions. The application of this approach to the denaturation of a human α-lactalbumin folding intermediate showed that the transition between the molten globule and the unfolded states is not cooperative [29]. A similar study of the unfolding of apoflavodoxin allowed for the determination of the stability of every native secondary structure element [30]. The contribution of 21 residues to the global stability of the native protein (compared to an intermediate state) could be quantified. Besides, the (un)folding transition between the intermediate and the unfolded state was shown to be noncooperative, but the lack of signal assignment of the unfolded state hindered a meaningful description of this transition at a residue level. Folding studies of intestinal fatty acid binding protein (IFABP) for which this approach was used showed that IFABP folding is initiated by residual structures present in the denatured state [31]. On a theoretical basis, there is no guarantee that these folding species are similar to the ones observed along the fold-

[3]Correspondence: mbruix@iqfr.csic.es

Key words: CheY; folding; intermediate; NMR; residue-specific study

ing reaction; nevertheless, a similarity between them has been found in some cases [32].

In this paper, we present a comprehensive study, at a residue level, of the species found in the folding process of CheY under equilibrium conditions. CheY is an α/β parallel protein containing 129 residues and a *cis*-prolyl bond between residues 109 and 110. There is evidence for a kinetic intermediate. Analysis of the folding reaction by the protein engineering method [33] shows that formation of the folding nucleus requires the packing of the first α helix onto the first two β strands. This study also showed that the transition state can be divided into two "folding subunits": a partially structured first half and a denatured second half, which later condenses to form the native structure. Even though these two subunits can be detected in the folding pathway of CheY, they are not found in the folded protein, which is made out of a single domain [34]. The folding intermediate appears to have a nonnative character, as shown by experiments in which each of the five helices of CheY were independently stabilized [35]. Essentially, while α helices 1–3 participate, to a certain extent, in the transition state, α helices 2–5 are partly folded in the intermediate state, but not in the transition state. Thus, going from the transition state to the intermediate state requires a partial folding of the protein.

Folding studies, at equilibrium, of wild-type (WT) CheY and some mutants have also been performed, showing a completely reversible process. Two subdomains were identified by hydrogen-exchange analysis of WT CheY and the double mutant F14N/P110G. The first subdomain extends from β strand 1 to 3, and the second subdomain extends from helix 3 to the C terminus [34]. Studies of the thermal unfolding transition of CheY have shown the existence of a folding intermediate having the properties of a molten globule [36]. Some aspects of this study were called into question in a subsequent study of the closely related protein CheY from *Salmonella typhimurium* [37]. However, the slight discrepancy between the results obtained with these proteins from different bacteria can be explained by the tendency toward self-association of the intermediate observed during the temperature-induced unfolding transition of the *E. coli* protein. On the other hand, the urea-induced unfolding process could be described by a simpler, two-state mechanism, although ANS (8-anilino-1-naphtalenesulfonic acid) binding experiments indicate the existence of a very unsteady intermediate [36]. Regarding the denatured state of CheY, protein engineering studies showed that one mutant (F14N/V83T) was likely to present a more compact unfolded form [33]. NMR characterization of this mutant in 5 M urea showed that it presents a less dynamic first half of the sequence and a transient collapse of the sequence region corresponding to the first α helix [38]. These dynamic properties would facilitate the setting up of longer-range interactions and funnel the reaction toward the transition state.

Important NMR data collected during the characterization of the denatured state have permitted a step forward in the description of this mutant (un)folding process. In the present work, we have used the previous NMR assignment of CheY mutant F14N/V83T in 0 and 5 M urea [38] to identify the residues and their corresponding resonances throughout the equilibrium transition. The extent to which individual residues participate in the structure of the native, intermediate, and denatured states can then be determined, and a subsequent thermodynamic analysis of these data provides information on the stability of the region surrounding each residue. Seventy-five single residues, well distributed throughout the protein sequence, have then been used as a probe. Two intermediates existing in the equilibrium folding reaction have been observed and characterized. By using all kinetic and equilibrium data, we then describe the entire folding process of CheY, from the folding nucleus to the native state, through the transition state and a molten globule-like intermediate.

Results

NMR Analysis of the Urea-Induced Denaturation Process

To follow the denaturation of the CheY mutant F14N/V83T, we recorded a series of 20 ^{15}N-^{1}H HSQC spectra at increasing urea concentrations under equilibrium conditions. Using the previous assignments for the native and denatured states [38], we could measure the intensity of all nonoverlapping peaks. This was possible because the exchange between these two forms is slow in the NMR timescale, and, thus, the spectra consist of a superposition of the native and denatured HSQCs, with the peaks of the native state progressively fading out and those from the denatured state gradually emerging. A sample of this feature is shown in Figure 1. An unambiguous and accurate analysis of peak intensities was possible for 75 out of the 125 peaks expected in the HSQC spectra for native or denatured CheY (after discounting the N terminus and the three prolines). This represents 60% of the residues in the protein sequence.

Potential crosspeak broadening was checked throughout the denaturation experiment, and no significant increase in peak line width was detected. The absence of changes in chemical shift and line width indicates that the different species in the sample are in slow exchange in the NMR timescale. Interestingly, this fact is consistent with the folding and unfolding rates determined for WT CheY and various mutants in a previous study [33, 39]. Changes in peak heights were compared to changes in peak volumes for a few isolated peaks, and no significant differences were found. However, estimating the population of native or denatured form from peak volumes is not easy, as many peaks in the HSQC spectra corresponding to the denatured protein are overlapped, and, also, in a few cases, appearing and disappearing peaks overlap. A normalized population of native protein (decreasing peaks) or denatured protein (increasing peaks) was then estimated from peak heights. Based on the intensities of these peaks in the native and denatured states, we could estimate how the folded and denatured populations changed with urea concentration in a residue-specific manner.

Evidence for an Intermediate State

A plot of the variation of the native and denatured populations, with increasing urea concentration for the differ-

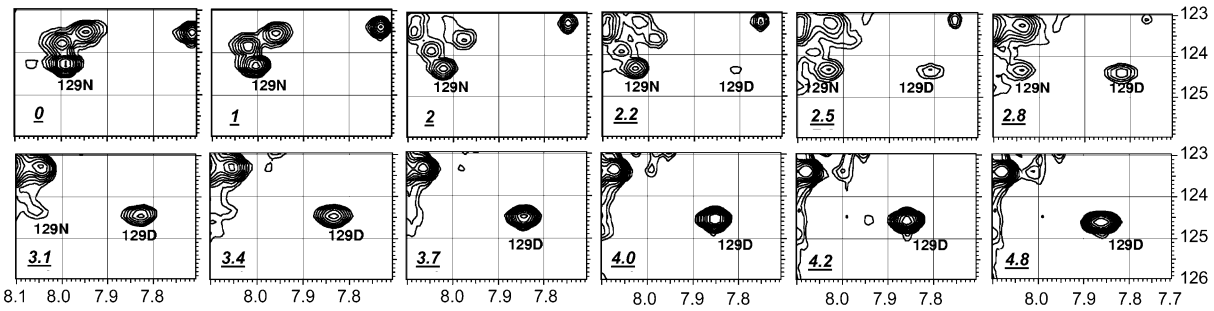

Figure 1. Evolution of Various Peaks Corresponding to Amide ^{15}N-^{1}H Nuclei of CheY Mutant F14N/V83T in a Series of Gradient-Enhanced ^{15}N-^{1}H HSQC Spectra as a Function of Urea Concentration

Peaks corresponding to methionine 129 in a native environment (129N) and in a denatured environment (129D) are labeled. For the sake of clarity, other native and denatured peaks are also shown, but not assigned, in this figure. Urea concentrations are shown, underlined, in the lower left corner of every cartoon. Axis units are in parts per million.

ent residues, shows the existence of an important lag between the disappearance of native peaks and the appearance of denatured peaks. This phenomenon is best illustrated for the NH proton of the tryptophan 58 indole ring (Figure 2A). This lag suggests that there is an intermediate state in the folding transition. At a urea concentration of around 2.8 M, the populations of both native and denatured forms represent about 20% of the total amount of protein in the sample.

The first transition would then correspond to a conformational change in which the native protein evolves to the intermediate state, and the second transition would correspond to a further denaturation of the intermediate state into the unfolded form. The simplest scheme accounting for these data would be as follows:

$$N \Leftrightarrow I \Leftrightarrow D$$

As discussed in similar cases [29–31], the missing

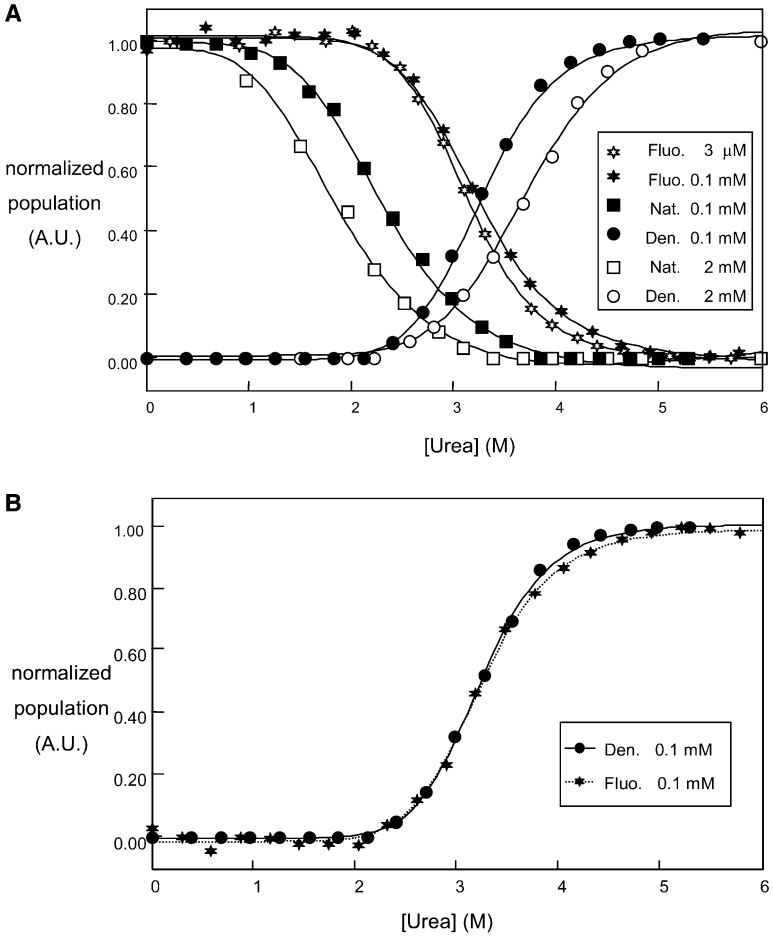

Figure 2. Normalized Urea-Induced Denaturation Curves of CheY Mutant F14N/V83T at Various Protein Concentrations, as Assessed by Various Techniques

(A) Urea-induced denaturation curves of CheY mutant F14N/V83T at various protein concentrations. Transitions monitored by fluorescence intensity changes and normalized from 100% to 0% native forms are shown as stars. Transitions followed by NMR are illustrated by plotting the variation of intensity of the peak corresponding to tryptophan 58 indole NH. Transition curves obtained by following the disappearance of NMR peaks corresponding to the native form are shown as squares, and those obtained by following the appearance of NMR peaks corresponding to the denatured forms are shown as circles. Open and filled symbols represent different protein concentrations (see inset).

(B) Superposition of the urea-induced denaturation curves of CheY mutant F14N/V83T as assessed by fluorescence (stars) and NMR (circles) when normalized from 0 to 100% in the population of the denatured form. Both denaturation experiments were carried out at 0.1 mM protein concentration.

Table 1. *Cm* as a Function of Protein Concentration

[Protein] (μM)	Technique	$N \Leftrightarrow I$	$I \Leftrightarrow D$
3	Fluorescence	—	3.14 ± 0.07
100	Fluorescence	—	3.32 ± 0.08
125	NMR	2.42 ± 0.17	3.38 ± 0.13
2000	NMR	1.78 ± 0.12	3.73 ± 0.08

Cm is given in molar concentration of urea.

crosspeaks of the intermediate can be explained by assuming that their amide resonances are broadened by conformational fluctuation on a millisecond time scale, as is characteristic of compact regions in nonnative states [40]. A further characterization of this intermediate state is presented below.

Fluorescence Analysis

It can be seen in Figure 2A that neither of the two transitions followed by NMR (2 mM protein) coincides with the denaturation transition followed by fluorescence in a previously published experiment [33]. However, as the experimental conditions vary from one experiment to the other (notably, there is a significant difference in protein concentration), it appears imperative to compare the two denaturation curves at a common protein concentration, around 0.1 mM.

Fluorescence can be easily measured at 0.1 mM protein; nevertheless, we could not run NMR HSQC spectra in a reasonable time at such a low protein concentration. However, this problem could be overcome by measuring the area of tryptophan indole HN peaks on a series of monodimensional NMR spectra. Indeed, the isolated position of this peak and the fact that the native peak (10.2 ppm) and the denatured peak (9.8 ppm) are not overlapped make it possible to measure the population of native and denatured protein at every urea concentration. The denaturation curves and the corresponding midtransition points and ΔG values are presented in Figure 2A and Tables 1 and 2, respectively, for various protein concentrations. The transition curves, as followed by fluorescence, superimpose (within experimental error) over the NMR curve corresponding to the appearance of denatured peaks for a protein concentration around 0.1 mM (Figure 2B; Tables 1 and 2).

These observations suggest that the transition, as followed by fluorescence, corresponds to the second equilibrium ($I \Leftrightarrow D$) only and that tryptophan fluorescence properties are not modified by a slight conformational change accompanying the native to intermediate transition. The fluorescence technique is well known for being extremely sensitive; however, it is also true that NMR spectroscopy can detect variations that cannot be observed by optical techniques. Indeed, the probe studied by the two techniques is the same (the indole ring), but the observed phenomenon is different. The decrease of fluorescence intensity observed during the unfolding transition can reasonably be attributed to a quenching caused by the collision of water molecules onto the indole ring [41], which becomes more accessible to solvent with denaturation. However, the increase in the indole ring's accessibility between the native and intermediate states may not be sufficient to cause a more efficient fluorescence quenching. Indeed, tryptophan 58 is already partially accessible to water molecules in the native state (accessibility estimated at 15%), and it seems likely that conformational changes accompanying this transition may not be large enough to significantly modify this accessibility. On the other hand, variations in the area of NMR peaks reflect changes in the molecular population having the indole HN nuclei in a certain magnetic environment. Changes in this environment along the folding process can be caused by a large variety of phenomena other than those related to conformational changes, such as changes in electrostatic or dynamic properties.

Table 2. ΔG as a Function of Protein Concentration

[Protein] (μM)	Technique	$N \Leftrightarrow I$	$I \Leftrightarrow D$
3	Fluorescence	—	4.84 ± 0.10
100	Fluorescence	—	4.98 ± 0.19
125	NMR	2.75 ± 0.15	5.29 ± 0.21
2000	NMR	2.25 ± 0.15	5.66 ± 0.13

ΔG is given in kilocalories per mole.

Variation of the Intermediate Population with Protein Concentration: Evidence for an Association Process

As stated above, the population of both native and denatured states can be estimated at 20% of the total amount of protein (at 2 mM and 2.8 M protein and urea concentrations, respectively). However, experiments at 0.1 mM protein concentration show that the total population of both native and denatured forms is around 40% at 2.8 M urea. The two transitions observed with NMR spectroscopy move obviously closer when lowering the protein concentration. From these observations, one can reasonably conclude that these transitions would coincide with each other at a very low protein concentration, thus reducing the population of intermediate species in the denaturation process to a negligible amount. This would explain why the formation of a very transient intermediate was merely detected in the urea-induced denaturation of WT CheY, as assessed by fluorescence or CD [36].

The fact that the population of the intermediate is increased when raising the protein concentration suggests that a self-association process might take place in the folding reaction. The data presented above are compatible with an off-pathway equilibrium between monomeric and associated states. Indeed, for a denaturant concentration favoring the intermediate formation, such as 2.8 M urea, increasing the protein concentration would shift an equilibrium between the monomeric and associated states toward the associated state. This would explain the loss of stability of either the native or the denatured forms relative to the intermediate state when increasing the protein concentration (Tables 1 and 2). This process can be accounted for by a simple scheme, such as

$$N \rightleftarrows \underset{\downarrow\uparrow}{I} \rightleftarrows D \qquad \text{(A)}$$
$$I_n$$

in which the intermediate would be in exchange between

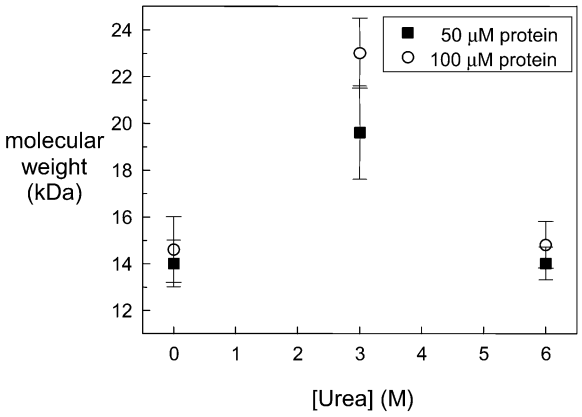

Figure 3. Molecular Weights of CheY Mutant F14N/V83T as a Function of Urea Concentration

Molecular weights were determined by analytical ultracentrifugation measurements. Black squares and open circles correspond to a protein concentration of 50 μM and 100 μM, respectively.

the monomeric and multimeric forms. As the NMR conditions and experiments used here only permit the detection of the native and denatured states, the denaturation curves observed in the present work will only give information on the $N \Leftrightarrow I$ or the $I \Leftrightarrow D$ equilibria. The study of the $I \Leftrightarrow D$ equilibrium at various protein concentrations also supports the hypothesis that the conformational changes accompanying the $N \Leftrightarrow I$ transition do not alter the fluorescence intensity of tryptophan 58. Indeed, the transition midpoints and free energies determined from denaturation curves, as assessed by fluorescence, also change with protein concentration and are equivalent to the thermodynamical parameters of the $I \Leftrightarrow D$ transition, derived from NMR, for an identical protein concentration (Tables 1 and 2).

To validate the intermediate oligomerization indicated by NMR and fluorescence, we performed small angle X-ray scattering and analytical ultracentrifugation analysis. Previous small angle X-ray scattering experiments showed that the protein is clearly monomeric in the presence of 0, 5, and 7 M urea [38]. Additional small angle X-ray scattering measurements were then performed in the presence of 3 M urea. In these conditions (corresponding to the maximum in intermediate population), an anomalous increase of the scattered signal at small scattering angles was observed (data not shown), indicating an association process. It may be noted that small angle X-ray scattering experiments were performed at the high protein concentration required by the 2D NMR experiments (~2 mM). The fact that these two techniques provide evidence for an associated form at high protein concentration strengthens the hypothesis that an equilibrium exists between monomeric and multimeric intermediate forms.

Figure 3 shows the changes in molecular weight of CheY mutant F14N/V83T corresponding to the native, intermediate, and denatured states at two different protein concentrations, as deduced from analytical ultracentrifugation measurements. The molecular weights at 0 and 6 M urea correspond to the value expected for the monomeric form, i.e., 14,000 kDa, independently of the protein concentration used. On the other hand, under the urea concentration where the intermediate state is highly populated, the molecular weight is clearly higher and increases with an increase in protein concentration. Interestingly, this value tends to approximate the molecular weight expected for dimer formation (28 kDa). At 0.1 mM protein concentration, the population of protein in both the native and unfolded states is around 40%. The intermediate population being 60%, we can estimate that, if the oligomerized intermediate is a dimer, the expected observed molecular weight would be ~22 kDa. This value is remarkably close to the value observed experimentally. In conclusion, all data supports the existence of an equilibrium between two forms of the intermediate, monomeric and dimeric, which can be shifted toward one species or the other by changing the protein concentration.

The determination of an appropriate scheme describing the unfolding transition of CheY is essential for a correct thermodynamic analysis of the extent to which an individual residue participates in each species conformation throughout this process. Once a correct scheme corresponding to this protein concentration dependency study is drawn, a further characterization of the intermediate state is necessary. The native environment of tryptophan 58 is modified during the $N \Leftrightarrow I$ transition, as deduced from NMR; nevertheless, it seems likely that the overall protein conformation is not greatly altered, as fluorescence intensity and far-UV CD signals are not modified. However, it has been shown previously that ANS can bind to WT CheY at urea concentrations much lower than the transition midpoint when denaturation is followed by fluorescence [36]. More precisely, a first transition in the ANS binding process is observed for urea concentrations between 1.5 M and 3 M, whereas tryptophan 58 fluorescence intensity is not modified until the urea concentration is around 3 M. It is worth noting that a maximum in ANS fluorescence had been observed at a concentration around 3 M urea, indicating that the intermediate state preferentially binds ANS. The intermediate then presents all properties of a molten globule state [42]. Indeed, the native secondary structure is conserved, tryptophan 58 is as accessible to water as it is in the native state, and some hydrophobic regions of the protein are exposed to the solvent, as the intermediate binds ANS more efficiently than either the native or the denatured forms.

In conclusion, the urea-induced unfolding of CheY can be described as a two-step process: (1) a first transition that leads to an on-pathway intermediate, likely to be a molten globule, which is in equilibrium with an off-pathway dimeric form and (2) a second transition between the monomeric intermediate form and the unfolded state.

Study of the Unfolding Process at the Residue Level

The aim of this study is to employ each residue as an individual probe to get information on the stability of very limited regions surrounding these residues in different conformations throughout the equilibrium denaturation of CheY. By doing so, we would be able to evaluate, in

a later stage, the sequence of the process and compare it with the conclusions derived from kinetic data.

The scheme (A) presented above will serve as a basis for the thermodynamic analysis. Data provided by NMR experiments represent the populations of native and denatured protein. If we consider the total molar fraction of protein (native, denatured, monomeric, and associated intermediates) to equal 1, the total population of intermediate, $I + I_n$, can be deduced readily. For each urea concentration this population can therefore be approximated to a unique population of intermediate, X, in equilibrium with the native and denatured forms. Thus, the scheme presented above can be approximated to the following mechanism:

$$N \underset{\leftarrow}{\overset{K_1}{\rightarrow}} X \underset{\leftarrow}{\overset{K_2}{\rightarrow}} D \qquad \text{(B)}$$

where X represents the total population of intermediate, $I + I_n$, and K_1 and K_2 are the equilibrium constants for the $N \Leftrightarrow X$ and $X \Leftrightarrow D$ equilibria, respectively. It is noteworthy that the population of intermediate X represents an ensemble of forms in association equilibrium and is not considered here as a monomer. In Experimental Procedures, we present a formal description of the equations used to obtain Equations 11 and 18, including the meaning of the various parameters involved (see Experimental Procedures). The fitting of Equation 11 to the experimental data (the sum of the normalized populations of native and denatured forms, $N + D$) provides the values of the transition midpoints for each equilibrium, C_{m1} and C_{m2}, respectively.

$$N + D = \frac{C_{m1}^{n1} \cdot C_{m2}^{n2} + x^{n1} \cdot x^{n2}}{C_{m1}^{n1} \cdot C_{m2}^{n2} + x^{n1} \cdot (C_{m2}^{n2} + x^{n2})} \qquad (11)$$

Similarly, a fitting of the same experimental data to Equation 18 will provide the values of ΔG corresponding to the two equilibria for different denaturant concentrations, thus providing information on the stability of the region surrounding each residue in each form during the folding process.

$$N + D = \frac{1 + e^{\frac{-\Delta G_1^0 + m_1 \cdot x}{RT}} \cdot e^{\frac{-\Delta G_2^6 - m_2 \cdot (x - 6)}{RT}}}{1 + e^{\frac{-\Delta G_1^0 + m_1 \cdot x}{RT}} \cdot \left(1 + e^{\frac{-\Delta G_2^6 - m_2 \cdot (x - 6)}{RT}}\right)} \qquad (18)$$

Figure 4 shows four examples of the fitting of experimental data to Equations 11 and 18. The transition midpoint provides information on the environment of a particular residue at a given denaturant concentration, helping us, in that way, to determine the sequence of events in the folding process. The ΔG of each equilibrium $A \Leftrightarrow B$ for a given residue represents the stability of the region surrounding this residue in state B as compared to state A.

Study of the Transition Midpoints: The Sequence of Events in the Unfolding Transition

In Figure 5A, we show the midpoint values for the different residues corresponding to the transition between the native and intermediate forms. The conformational change between these two states affects the environment of all residues concomitantly. Within experimental error, all residues (apart from a single residue located at the N-terminal extremity of the protein) present $N \Leftrightarrow X$ transition midpoints that are equivalent to the average midpoint value ($<C_{m1}> = 1.80$ M urea; standard deviation, SD, is 0.20 M urea). This indicates that the native conformation evolves into the intermediate state in a highly cooperative way.

On the other hand, the transition midpoints corresponding to the denaturation of the intermediate state show two sets of values (Figure 5B). Approximately the first half (from residues 1 to 71) of the protein presents an average C_{m2} value of 4.13 M urea (SD = 0.33), whereas the second half presents an average C_{m2} value of 3.36 M urea (SD = 0.37). In the C-terminal half of the sequence, a few residues deviate significantly from the average values (105, 124, and 129, with C_{m2} values of 4.14, 4.33, and 4.39, respectively), but they appear to be isolated and, therefore, are not likely to correspond to a folded conformation.

These two clusters of transition midpoints define unambiguously two folding subunits in the sequence. At 3.75 M urea, almost all residues located after lysine 70 correspond to an unfolded chain, whereas almost all residues from the first half are still in the intermediate, collapsed conformation. This can be visualized as a collapsed, globular chain with a 59-residue unfolded tail (Figure 6).

Study of the Stability of Regions Surrounding Each Residue in the Native, Intermediate, and Denatured States

The ΔG values obtained for each residue provide the stability of the region surrounding this amino acid in the various conformations existing in the equilibrium folding process. The analysis of the thermodynamic parameters of the $N \Leftrightarrow X$ transition will then give us information on the stability of the native protein compared to the intermediate state. The stability of limited regions surrounding given residues in the native state is provided by determining the free energy of the first equilibrium ($N \Leftrightarrow X$) in absence of denaturant, thereafter called ΔG_1^0. Variations of ΔG_1^0 along the protein sequence are shown in Figure 7A. The distribution of ΔG_1^0 values shows a region comprising residues 90–100 (inclusive) plus some residues at the N and C termini with lower ΔG_1^0 values. This may indicate that these regions are less stable in the native protein. Indeed, H/D exchange and structural studies on native WT CheY [34, 43] have shown that the less stable part of the native protein is α helix 4, which corresponds to residues 90–100.

Nevertheless, the ΔG_1^0 values and the stability of the native protein are found to be distributed rather uniformly throughout the structure. An additional analysis of the ΔG_1^0 values was then performed by fitting Equation 18 to all residues from each folding subdomain simultaneously. The ΔG_1^0 corresponding to residues 1–71 is 2.5 $\pm$ 0.1 kcal/mol, whereas the ΔG_1^0 corresponding to residues 72–129 is 2.6 $\pm$ 0.2 kcal/mol. The fact that these two values are identical (within experimental error) supports the idea that the disruption of the native molecule into a molten globule state is a cooperative process. This fact, together with the detection of the least stable element of the structure, in agreement with previous studies on CheY stability, validates the kind of analysis

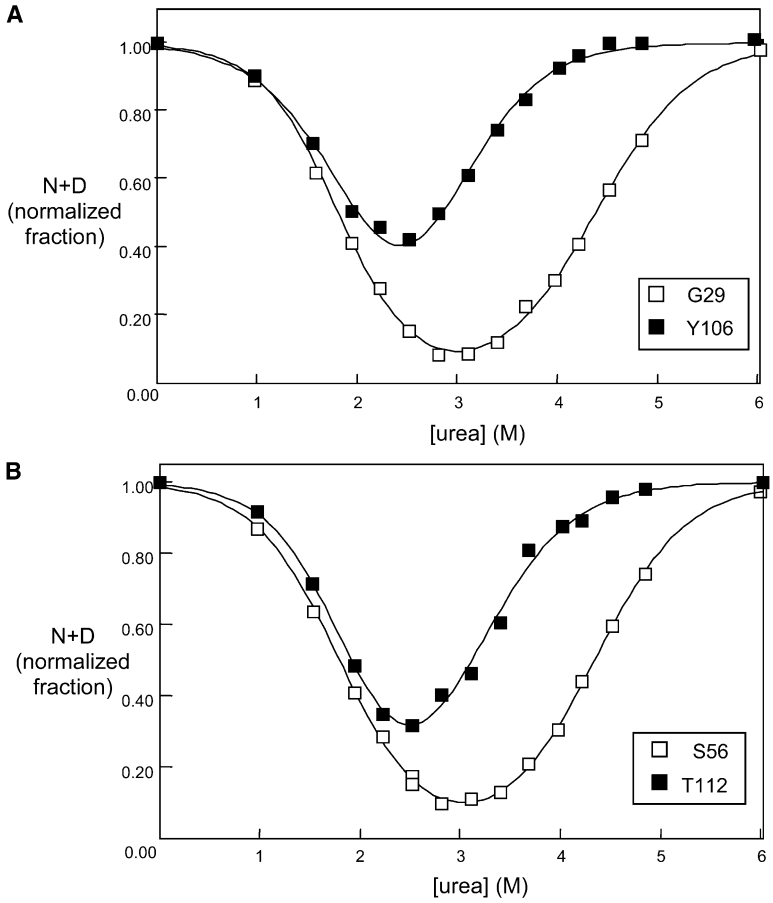

Figure 4. Examples of Fitting of 2D NMR Data to Equations 11 (A) and 18 (B)

NMR data correspond to the sum of normalized populations of native and denatured states, as determined by the evolution of peak intensities on HSQC spectra. Open and filled symbols correspond to residues located in the first or second half of the sequence, respectively.

(A) The fitting of Equation 11 to the data, which provides transition midpoints.

(B) The fitting of Equation 18 to the data, which provides ΔG values for each equilibrium.

performed here for the evaluation of a protein's stability at a residue level.

Determination of the stability of the denatured state (with respect to the intermediate) under fully denaturing conditions is preferable to the determination of the intermediate stability in the absence of denaturant. Indeed, at 0 M urea the stability of the intermediate relative to the denatured state cannot be estimated, as the intermediate itself is fully destabilized toward the native conformation. The stability of the regions surrounding every residue in the intermediate referred to the unfolded state can easily be deduced by assuming that the more stable the unfolded form, the less stable the intermediate and vice versa. In parallel to what is accepted for native proteins, stability can then be estimated by determining the ΔG value of the second equilibrium at 6 M urea, thereafter called ΔG_2^6. High ΔG_2^6 values will then refer to residues located in more stable regions in the denatured state (Figure 7B) and, consequently, located in regions that present a lower stability in the intermediate state. Figure 7B shows a relatively uniform dispersion of ΔG_2^6 values. However, residues in the first subdomain tend to show lower values. A group of residues distributed between amino acids 70 and 105 also deviate significantly and show higher values. The thermodynamic properties of this second equilibrium have then been analyzed in the same way as those for the first equilibrium, i.e., by fitting Equation 18 to residues 1–71 or 72–129 simultaneously. The ΔG_2^6 values corresponding to the first half and the second half are 2.4 $\pm$ 0.1 kcal/mol and 3.3 $\pm$ 0.2 kcal/mol, respectively. This distribution of ΔG_2^6 values indicates clearly that the first half of the polypeptide chain contributes more to the intermediate stability.

Discussion

The entire folding process of a CheY mutant has been analyzed under equilibrium conditions using NMR spectroscopy. We have shown the existence of an equilibrium intermediate and managed to determine thermodynamically the stabilities of limited regions surrounding 75 residues in all the species found in equilibrium in the urea unfolding of a protein.

Nature of the CheY Oligomeric Equilibrium Intermediate

The lag between the disappearance and appearance of peaks corresponding to the native and denatured forms, respectively, indicates the existence of an equilibrium intermediate in the folding of this CheY mutant. This is confirmed by comparing the fluorescence and NMR data, which shows that the loss of fluorescence of a tryptophan residue is not in keeping with the loss of the native signals as observed by NMR. Concentration dependence experiments and analytical ultracentrifugation suggest that the intermediate is oligomeric and, most likely, a dimer. NMR analysis at an individual level

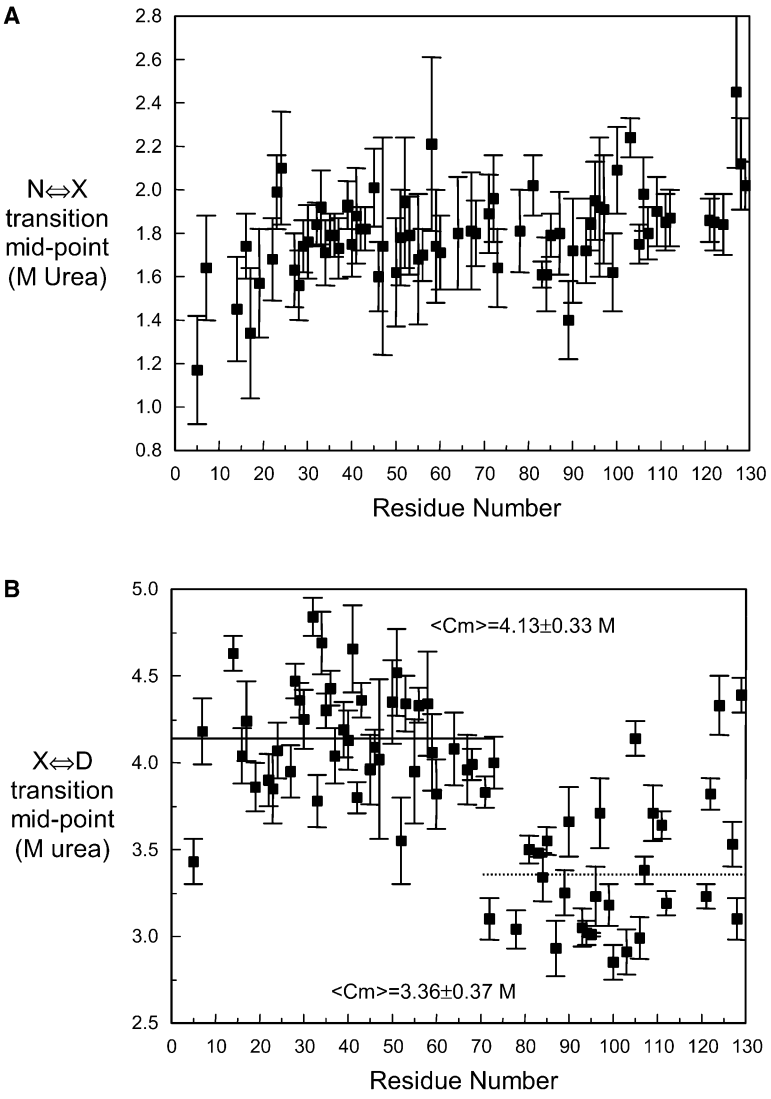

Figure 5. Variation of Transition Midpoints Corresponding to Equilibria $N \Leftrightarrow X$ (A) and $X \Leftrightarrow D$ (B) as a Function of the Sequence

In (B) the mean transition midpoints corresponding to the first half of the sequence (continuous line) and the second half of the sequence (dotted line) are also shown, together with their values.

shows a cooperative transition between the native and intermediate states.

The postulated dimeric intermediate shows all properties described for the molten globule-like form. Indeed, the native environment of each residue is altered in this state (signals with native chemical shifts are missing), and the hydrophobic surfaces are exposed (increased ANS binding [36]), but secondary structure is maintained, as no change in ellipticity is observed [36]. The absence of variation in the signal of fluorescence signal may be due to the very small change in tryptophan exposure accompanying the first transition, as discussed above. Interestingly, the possibility of a three-state mechanism had been envisaged previously, including the existence of a very unstable intermediate, only detectable by ANS binding experiments [36].

Existence of a Second Intermediate

Unfolding of the native form to the molten globule like intermediate is highly cooperative. On the other hand, the conversion of this intermediate into the unfolded state shows a sequential process. Indeed, a neat picture of CheY structure at a urea concentration of 3.75 M would show a nonnative, but collapsed, first half of the chain, together with a fully unfolded second half. This clearly defines two folding subunits that could fold independently. However, it has to be emphasized that the native structure of CheY corresponds to a single globular domain, and it is by no means constituted by two domains or subdomains. According to this ensemble of results, a new mechanism can be proposed for the unfolding of CheY at equilibrium. In the first step, CheY would unfold into a molten globule-like intermediate, called I_N, which can associate off-pathway. In the second step, I_N would unfold into another intermediate form, called I_U, in which the first half of the polypeptide chain remains collapsed, and the second half is already unfolded. In the last step, I_U would unfold entirely, to give rise to the denatured state. This model can be schematized as follows:

$$N \rightleftarrows I_N \rightleftarrows I_U \rightleftarrows D \qquad \text{(C)}$$
$$\downarrow\uparrow$$
$$(I_N)_n$$

It is interesting to note that the folding scheme deduced

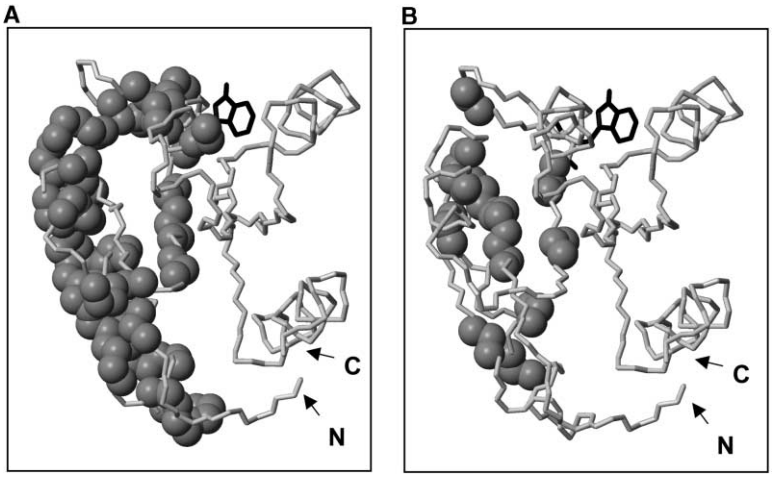

Figure 6. A Comparison of the Conformation Deduced Here for the Intermediate Observed under Equilibrium Conditions at 3.75 M Urea Characterized at Equilibrium (A) with the Conformation of the Transition State Characterized Using the Protein Engineering Method (B)

Residues corresponding to a collapsed conformation in the intermediate at 3.75 M urea are indicated with a space-filling representation in (A). Residues corresponding to a collapsed conformation in the transition state by ϕ value analysis [33] are indicated with a space-filling representation in (B). The indole ring of tryptophan 58, as well as the N and C termini, is also shown. The diagram was generated using the MolMol program, with NMR solution structure coordinates of WT CheY [56].

from this analysis is different from the scheme used for the mathematical analysis of the data. However, this scheme (C) represents a model of folding for the overall protein, whereas folding scheme A, which does not include an intermediate I_U, corresponds to the changes in every single residue environment. Each residue environment changes following two transitions, and only the observation that the $I_N \Leftrightarrow D$ transition is not concomitant for all residues permits the setting up of scheme C.

Equilibrium Versus Kinetic Data

As stated recently, "the interpretation of equilibrium denaturation studies in defining a folding pathway is always controversial" [31]. However, the thermodynamic

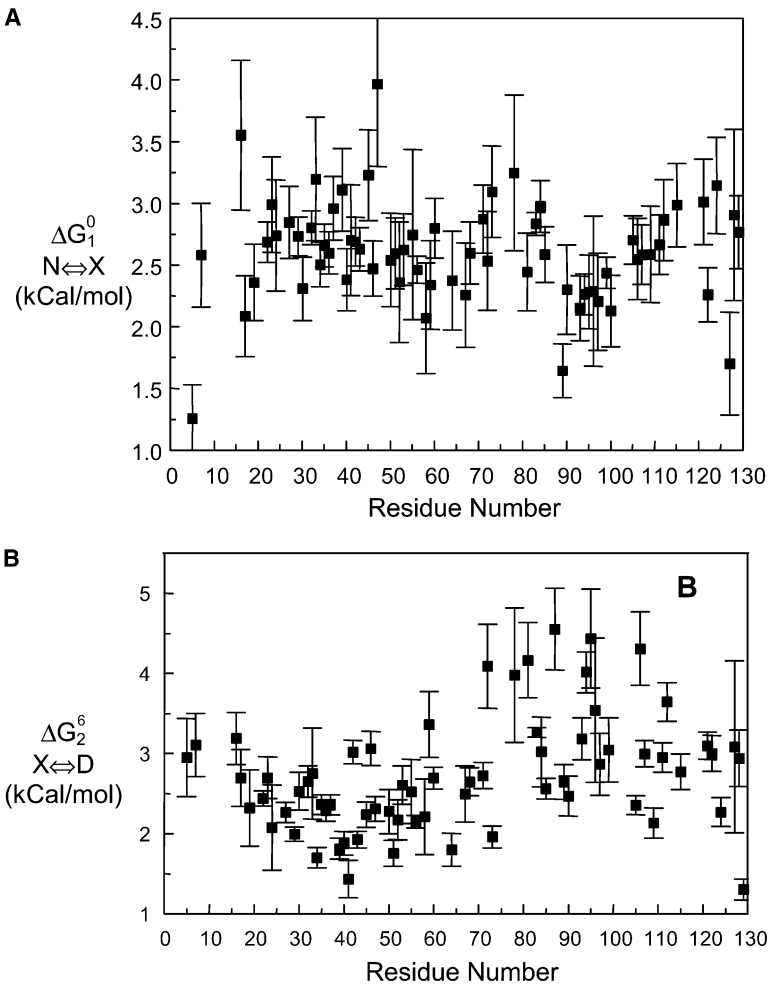

Figure 7. Variation of the Stabilities of Various Regions Surrounding Given Residues in the Native and Denatured States as a Function of the Sequence

ΔG_1^0 represents the stability of regions surrounding given residues in the native state versus the intermediate state (A). ΔG_2^6 represents the stability of regions surrounding given residues in the denatured state versus the intermediate form (B).

parameters determined here can help us estimate the energy landscape that lies between the native and denatured states and compare it with the extensive kinetic and theoretical data collected on the folding pathway of CheY.

The transition midpoints obtained for 75 residues could be represented on the molecule as a cartoon showing the progressive unfolding of CheY, but we will focus first on the most significant state revealed by this study, the intermediate I_U. Its comparison with the transition state characterized previously [33] reveals an outstanding similarity (Figure 6) on the basis of which we can speculate about the significance of this conformation in the folding pathway of CheY. Indeed, this striking similarity suggests that the intermediate I_U might certainly be formed in the folding pathway.

Turning to the dimeric intermediate, I_N, we have shown that it retains all its secondary structure elements, although its tertiary structure appears altered. Interestingly, kinetic analysis of the folding intermediate of CheY with the five α helices stabilized independently showed that all of them were present, to a certain extent, in the kinetic intermediate [35], with some of them unfolded in the transition state. This also establishes some points of similarity between the equilibrium intermediate and the kinetic intermediate.

Mutant Versus Wild-Type Proteins

One important point that merits discussion is to what extent the results obtained here can be extrapolated to the wild-type protein. In this regard, the double mutant analyzed in this work incorporates a very stabilizing mutation in the first subdomain (F14N; [39]) and a very destabilizing mutation in the second half (V83T; [33]). In this respect it is similar to mutant F14N/P110G, analyzed by hydrogen exchange [34], for which it was shown that a stabilization of the first half and a simultaneous destabilization of the second half results in a decoupling in the stability of these two CheY regions. Moreover, studies by NMR, calorimetry, CD, and fluorescence of WT CheY under acidic conditions also showed the presence of a dimeric equilibrium intermediate [36]. Thus, it appears reasonable to accept that the mutations present in the F14N/V83T CheY mutant do not lead to the presence of new intermediate but, rather, enhance their likelihood.

An important consequence of this result on the mutant versus wild-type comparison is the interpretation of the *m* value in terms of a more compact denatured state. Mutant F14N/V83T presents a lower *m* value than does WT CheY or the F14N mutant. Hovewer, the NMR characterization of this state [38] does not reveal an important degree of compactness when compared to various denatured states characterized previously [17]. This fact can now be explained by a possible stabilization of the intermediate I_U, driven by the V83T mutation. This would have introduced a bias in the interpretation of *m* values. This is a noteworthy fact that should be considered in further studies of protein folding.

The results obtained here using this mutant also give a new insight into the interpretation of WT and mutant F14N folding studies. Indeed, the detection of a populated intermediate that can self-associate in the urea-induced unfolding transition may appear to be in disagreement with a previous study where the rate constants of the unfolding/refolding of CheY were found to be concentration independent [39]. However, we show that the transition monitored by fluorescence or CD measurements corresponds to the equilibrium between the unfolded state and the intermediate I_N, which does not include the association equilibrium, as shown in scheme C. This apparent discrepancy may then come from the different technique used here (NMR instead of fluorescence), and not from the V83T mutation.

Biological Implications

The findings described here have some implications for the general problem of protein folding. Simulations have suggested that populated intermediates could act as traps that productive folding processes have to bypass for a faster track [44]. In the case of CheY, the similarity between the equilibrium dimeric intermediate I_N and the kinetic intermediate represents experimental evidence of this fact. On the other hand, the resemblance between the intermediate I_U and the transition state suggests that this populated intermediate would rather act as a funneling agent, as reported by different theoretical studies [45].

This study also provides an important insight into the energy landscape that governs the folding of CheY, such as the stabilities of small regions surrounding individual residues in the different conformations present in the pathway. Indeed, we have shown that a conformation of CheY with a collapsed first half of the sequence is energetically favorable, as indicated by ΔG_2^6 values. If this conformation actually resembles the transition state, then we may conclude that the search of the native structure has been optimized. In other words, the transition state may be the highest energy species in the CheY folding pathway, but not in the folding energy landscape. This state would then correspond to the lowest energy conformation existing for such reaction coordinates. CheY may have to search its native structure across an important energy barrier but has apparently found a mountain pass to cross it, therefore illustrating well what had been proposed previously [46]. Currently, a great number of theoretical and simulation studies tend to follow this direction. In this framework, the present study constitutes important and encouraging experimental evidence to support them.

Experimental Procedures

Protein Expression and Purification

The mutant F14N/V83T was designed, cloned, and sequenced as described previously [33]. The plasmid containing the mutated gene was transformed into a BL21 *E. Coli* strain. Cells were grown on minimal medium with $^{15}NH_4Cl$ as the only nitrogen source, and ^{15}N-labeled proteins were purified as described previously [36].

Denaturation Experiments

All experiments were performed at 25°C in a buffer containing 25 mM sodium phosphate at pH 7.

Denaturation as followed by NMR was performed by adding small amounts of a 10 M urea stock solution into an NMR tube containing 2 mM protein (in the case of HSQCs) or 0.1 mM protein (in the

case of monodimensional spectra). The protein concentration was maintained at its initial value by adding small amounts of a protein stock solution (5 mM in the case of HSQCs or 1 mM in the case of 1D spectra) at the same time that urea was added. Samples were allowed to equilibrate 30 min before spectra recording. At each urea concentration the NMR probe was tuned and matched, the magnetic field was shimmed, and the 90° pulses width were determined.

For denaturation as followed by fluorescence, a series of 1 ml samples containing 3 μM or 100 μM protein was prepared. These samples were prepared by final addition of a 5 mM protein stock solution to premixed solutions of urea and buffer.

For denaturation as followed by sedimentation equilibrium, a series of 1 ml samples containing 0, 50, or 100 μM protein in various denaturant concentrations (0, 3, or 6 M urea) was prepared.

For the denaturation series followed by fluorescence and sedimentation equilibrium, the final urea concentration of each sample was checked after acquisition by refractometry, using the relation provided by Warren and Gordon [47]. The final urea concentration of the sample used for NMR recording was also checked by refractometry.

Data Acquisition and Analysis: Fluorescence Spectroscopy

A tryptophan fluorescence spectrum was recorded between 300 and 400 nm after excitation at 290 nm for every sample on a Perkin-Elmer LS50-B fluorimeter. Excitation and emission slits were set to 2 nm. After baseline suppression, the fluorescence intensity was measured at the wavelength corresponding to the maximal change in fluorescence between native and denatured forms (315 nm). The transition curves were constructed by plotting this variation in fluorescence emission as a function of denaturant concentration.

Thermodynamic analysis was performed by using an equation developed previously [48], derived from the model of linear dependency of ΔGx upon denaturant concentration, x, as described by Tanford [49]:

$$y_x = y_n + s_n x + \left\{ \frac{e^{\left[\frac{(\Delta G_0 - mx)}{RT}\right]}}{1 + e^{\left[\frac{(\Delta G_0 - mx)}{RT}\right]}} \right\} [A + (s_d - s_n)x]$$

where y_x is the experimental signal in the presence of x molar urea, y_n is the signal of the native form, s_n and s_d are the solvent effects on the native and denatured protein signals, respectively, and A is the amplitude of the transition.

Similarly, the concentration of denaturant at the midpoint of the transition, C_m, and the cooperativity of the transition, n, were obtained by fitting the transition curves to the equation described previously [48], based on the denaturant binding model [50]:

$$y_x = y_n + s_n x + \left\{ \frac{x^n}{(C_m)^n + x^n} \right\} [A + (s_d - s_n)x]$$

where y_x, x, y_n, s_n, s_d, and A correspond to the parameters described above. Experimental data were fitted according to these equations by using a simplex procedure based on the Nelder and Mead algorithm [51].

Data Acquisition and Analysis: Sedimentation Equilibrium

Sedimentation equilibrium experiments were carried out in a Beckman Optima XI-A ultracentrifuge using a Ti60 rotor and double sector centerpieces of Epon-charcoal (12 or 4 mm optical path length, depending on protein concentration). Samples of CheY mutant F14N/V83T in the concentration range from 3 to 100 μM equilibrated in buffer were centrifuged at 28,000 rpm. Radial scans at different wavelengths (280–300 nm) were taken every 2 hr until equilibrium was reached. The weight average molecular masses of mutant F14N/V83T were determined using the program EQASSOC (Beckman) [52], with the partial specific volume calculated from its amino acid composition [53].

Data Acquisition, Processing, and Analysis: NMR Spectroscopy

All NMR experiments were performed using a Bruker AMX-600 spectrometer with a ^{1}H operating frequency of 600.13 MHz. All spectra were collected as described previously for the mutant characterization in the native and unfolded forms [38].

All NMR data were processed and analyzed using standard Bruker software and NMRPipe/NMRDraw software [54]. Various square sine bell functions with different shifts were applied to the data processing to obtain a good compromise between resolution and sensitivity.

Native and denatured state populations were obtained using the height of disappearing or emerging peaks and normalizing it from 1 to 0 in the case of native peaks and from 0 to 1 in the case of denatured peaks.

Data Analysis

In our case, a simple calculation of thermodynamic parameters cannot be performed by taking each equilibrium independently. Applying the linear dependency model to a two-state mechanism implies a process in which only the native and denatured states are populated. Transition midpoints of apoflavodoxin unfolding transition followed by NMR could be calculated in this way by van Mierlo and coworkers [30] only because these authors checked that a 100% population of intermediate species was present at the end of the transition. It is obvious that, in the case of CheY, a 100% population of intermediate species never exists. We then developed a new model for the analysis, in which every fraction of the species present in the sample is unknown. The sum of these fractions is 1. Since the X value is obtained from the total molar fraction, the potential effects of self-association in this ensemble of forms (monomer and multimers) on equilibrium constants K_1 and K_2 (and therefore on populations of N and D states) are taken into account.

According to this mechanism,

$$K_1 = X/N \text{ and } K_2 = D/X \text{ and } N + X + D = 1 \quad (1)$$

so that

$$\frac{X}{K_1} + K_2 \cdot X + X = 1 \quad (2)$$

which gives

$$X = \frac{K_1}{1 + K_1 + K_1 \cdot K_2}. \quad (3)$$

Substituting Equation 3 into Equation 1 gives

$$N + D = 1 - \frac{K_1}{1 + K_1 + K_1 \cdot K_2} \text{ or } N + D = \frac{1 + K_1 \cdot K_2}{1 + K_1 \cdot (1 + K_2)}. \quad (4)$$

According to the denaturant fixation model [50],

$$K_1 = K_1^0 \cdot x^{n_1} \quad (5)$$

and

$$K_2 = K_2^0 \cdot x^{n_2} \quad (6)$$

where x is the denaturant concentration, n_1 and n_2 are the cooperativity indices of each equilibrium, and K_1^0 and K_2^0 are the equilibrium constants for the unfolding of each equilibrium in the absence of denaturant. Substituting K_1 and K_2 into Equation 4 gives

$$N + D = \frac{1 + K_1^0 \cdot x^{n_1} \cdot K_2^0 \cdot x^{n_2}}{1 + K_1^0 \cdot x^{n_1} \cdot (1 + K_2^0 \cdot x^{n_2})} \quad (7)$$

and, subsequently,

$$N + D = \frac{\frac{1}{K_1^0} \cdot \frac{1}{K_2^0} + x^{n_1} \cdot x^{n_2}}{\frac{1}{K_1^0} \cdot \frac{1}{K_2^0} + x^{n_1} \cdot (1/K_2^0 + x^{n_2})} \quad (8)$$

for a denaturant concentration x equal to C_{m1}. Then, $K_1 = 1$ (C_{m1} being the transition midpoint for the first equilibrium) and

$$1 = K_1^0 \cdot C_{m_1}^{n_1} \text{ or } 1/K_1^0 = C_{m1}^{n_1} \quad (9)$$

and

$$1 = K_2^0 \cdot C_{m2}^{n2}. \quad (10)$$

Substituting Equations 9 and 10 into Equation 8 gives

$$N + D = \frac{C_{m1}^{n1} \cdot C_{m2}^{n2} + x^{n1} \cdot x^{n2}}{C_{m1}^{n1} \cdot C_{m2}^{n2} + x^{n1} \cdot (C_{m2}^{n2} + x^{n2})}. \quad (11)$$

A curve corresponding to Equation 11 is then fitted to the experimental points corresponding to the normalized population of native and denatured forms.

Similarly, the stability of the regions surrounding each residue in each form during the folding process can be determined, with the ΔG of every equilibrium for various denaturant concentrations. Indeed, equilibrium constants are equal to

$$K_1 = e^{-\frac{\Delta G_1}{RT}} \quad (12)$$

and

$$K_2 = e^{-\frac{\Delta G_2}{RT}} \quad (13)$$

for $N \Leftrightarrow X$ and $X \Leftrightarrow D$ equilibria, respectively. The model chosen to estimate the conformational stability of every species is the "linear dependency model" as described by Tanford [49] and later by Schellman [55].

This model provides the stability of a protein in absence of denaturant, ΔG^0, for a two-state equilibrium. However, in the case of the second equilibrium ($X \Leftrightarrow D$), determining the stability of an intermediate state in water makes no sense, as this form folds into the native state in such conditions. As for the ΔG_1^0 determined with the "linear dependency model" for the first equilibrium, it then seems more logical to determine the free energy of the second equilibrium for a 100% denatured population. In a parallel way to the ΔG_1^0, which gives the stability of the native form as compared to the intermediate form, determining a value of ΔG_2 for 6 M urea (ΔG_2^6) will then give the stability of the denatured state as compared to the intermediate form. This value does not represent the absolute stability of each region in the intermediate state but permits the comparison of this stability between the various regions surrounding all residues.

According to the "linear dependency model,"

$$\Delta G_1^x = \Delta G_1^0 - m_1 \cdot x \quad (14)$$

where ΔG_1^0 is the extrapolation of the free energy of the first equilibrium to 0 M urea. Similarly, the free energy of the second equilibrium can be extrapolated to 6 M urea, where a 100% denatured form is expected, using the following equation:

$$\Delta G_2^x = \Delta G_2^6 + m_2 \cdot (x - 6). \quad (15)$$

Substituting Equations 14 and 15 into Equations 12 and 13, respectively, one obtains

$$K_1 = e^{\frac{-\Delta G_1^0 + m_1 \cdot x}{RT}} \quad (16)$$

and

$$K_2 = e^{\frac{-\Delta G_1^6 - m_2 \cdot (x - 6)}{RT}}. \quad (17)$$

Substituting Equations 16 and 17 into Equation 4 gives

$$N + D = \frac{1 + e^{\frac{-\Delta G_1^0 + m_1 \cdot x}{RT}} \cdot e^{\frac{\Delta G_2^6 - m_2 \cdot (x - 6)}{RT}}}{1 + e^{\frac{-\Delta G_1^0 + m_1 \cdot x}{RT}} \cdot \left(1 + e^{\frac{-\Delta G_2^6 - m_2 \cdot (x - 6)}{RT}}\right)}. \quad (18)$$

Similar to the method used to obtain the transition midpoints, the fitting of experimental data corresponding to the sum of the normalized populations of native and denatured forms, $N + D$, to Equation 18 provides the values of ΔG_1^0 and ΔG_2^6 and the m values of each equilibrium.

Acknowledgments

We thank Dr. Dominique Durand (Laboratoire pour l'Utilization du Rayonnement Electromagnétique, Orsay, France) for her assistance during the SAXS experiments, Dr. José-Manuel Pérez Cañadillas and Dr. Jan Jouvert for helpful discussion, and Dr. Douglas V. Laurents for carefully reading the manuscript. This work was partly supported by an FEBS short-term fellowship, the Consejo Superior de Investigaciones Científicas, and the European Molecular Biology Laboratory.

Received: February 15, 2002
Revised: May 2, 2002
Accepted: May 28, 2002

References

1. Kim, P.S., and Baldwin, R.L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu. Rev. Biochem. *51*, 459–489.
2. Kim, P.S., and Baldwin, R.L. (1990). Intermediates in the folding reactions of small proteins. Annu. Rev. Biochem. *59*, 631–660.
3. Karplus, M., and Weaver, D.L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. Protein Sci. *3*, 650–668.
4. Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. Nat. Struct. Biol. *4*, 10–19.
5. Harrison, S.C., and Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? Proc. Natl. Acad. Sci. USA *82*, 4028–4030.
6. Fersht, A.R. (1997). Nucleation mechanisms in protein folding. Curr. Opin. Struct. Biol. *7*, 3–9.
7. Kotik, M., Radford, S.E., and Dobson, C.M. (1995). Comparison of the refolding of hen lysozyme from dimethyl sulfoxide and guanidinium chloride. Biochemistry *34*, 1714–1724.
8. Wong, K.B., Freund, S.M., and Fersht, A.R. (1996). Cold denaturation of barstar: 1H, 15N and 13C NMR assignment and characterisation of residual structure. J. Mol. Biol. *259*, 805–818.
9. Yao, J., Chung, J., Eliezer, D., Wright, P.E., and Dyson, H.J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. Biochemistry *40*, 3561–3571.
10. Alonso, D.O., and Daggett, V. (2000). Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains. Proc. Natl. Acad. Sci. USA *97*, 133–138.
11. Shortle, D., and Ackerman, M.S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. Science *293*, 487–489.
12. Baldwin, R.L., and Zimm, B.H. (2000). Are denatured proteins ever random coils? Proc. Natl. Acad. Sci. USA *97*, 12391–12392.
13. Matouschek, A., Kellis, J.T., Jr., Serrano, L., and Fersht, A.R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. Nature *340*, 122–126.
14. Roder, H., Elove, G.A., and Englander, S.W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature *335*, 700–704.
15. Miranker, A.D., and Dobson, C.M. (1996). Collapse and cooperativity in protein folding. Curr. Opin. Struct. Biol. *6*, 31–42.
16. Jennings, P.A., and Wright, P.E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science *262*, 892–896.
17. Mok, Y.K., Kay, C.M., Kay, L.E., and Forman-Kay, J.D. (1999). NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. J. Mol. Biol. *289*, 619–638.
18. Zhang, O., Kay, L.E., Shortle, D., and Forman-Kay, J.D. (1997). Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease Delta131Delta, using NMR methods with improved resolution. J. Mol. Biol. *272*, 9–20.
19. Blanco, F.J., Serrano, L., and Forman-Kay, J.D. (1998). High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. J. Mol. Biol. *284*, 1153–1164.
20. Kortemme, T., Kelly, M.J., Kay, L.E., Forman-Kay, J.D., and Serrano, L. (2000). Similarities between the spectrin SH3 domain denatured state and its folding transition state. J. Mol. Biol. *297*, 1217–1229.

21. Clore, G.M., and Gronenborn, A.M. (1998). NMR structure determination of proteins and protein complexes larger than 20 kDa. Curr. Opin. Chem. Biol. *2*, 564–570.
22. Szpikowska, B.K., and Mas, M.T. (1996). Urea-induced equilibrium unfolding of single tryptophan mutants of yeast phosphoglycerate kinase: evidence for a stable intermediate. Arch. Biochem. Biophys. *335*, 173–182.
23. Ballery, N., Desmadril, M., Minard, P., and Yon, J.M. (1993). Characterization of an intermediate in the folding pathway of phosphoglycerate kinase: chemical reactivity of genetically introduced cysteinyl residues during the folding process. Biochemistry *32*, 708–714.
24. Balbach, J., Forge, V., van Nuland, N.A., Winder, S.L., Hore, P.J., and Dobson, C.M. (1995). Following protein folding in real time using NMR spectroscopy. Nat. Struct. Biol. *2*, 865–870.
25. Dobson, C.M., and Hore, P.J. (1998). Kinetic studies of protein folding using NMR spectroscopy. Nat. Struct. Biol. *5*, 504–507.
26. Balbach, J., Steegborn, C., Schindler, T., and Schmid, F.X. (1999). A protein folding intermediate of ribonuclease T1 characterized at high resolution by 1D and 2D real-time NMR spectroscopy. J. Mol. Biol. *285*, 829–842.
27. Steegborn, C., Schneider-Hassloff, H., Zeeb, M., and Balbach, J. (2000). Cooperativity of a protein folding reaction probed at multiple chain positions by real-time 2D NMR spectroscopy. Biochemistry *39*, 7910–7919.
28. Killick, T.R., Freund, S.M., and Fersht, A.R. (1999). Real-time NMR studies on a transient folding intermediate of barstar. Protein Sci. *8*, 1286–1291.
29. Schulman, B.A., Kim, P.S., Dobson, C.M., and Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. Nat. Struct. Biol. *4*, 630–634.
30. van Mierlo, C.P., van den Oever, J.M., and Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. Protein Sci. *9*, 145–157.
31. Hodsdon, M.E., and Frieden, C. (2001). Intestinal fatty acid binding protein: the folding mechanism as determined by NMR studies. Biochemistry *40*, 732–742.
32. Paci, E., Smith, L.J., Dobson, C.M., and Karplus, M. (2001). Exploration of partially unfolded states of human alpha-lactalbumin by molecular dynamics simulation. J. Mol. Biol. *306*, 329–347.
33. López-Hernández, E., and Serrano, L. (1996). Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. Fold. Des. *1*, 43–55.
34. Lacroix, E., Bruix, M., López-Hernández, E., Serrano, L., and Rico, M. (1997). Amide hydrogen exchange and internal dynamics in the chemotactic protein CheY from Escherichia coli. J. Mol. Biol. *271*, 472–487.
35. López-Hernández, E., Cronet, P., Serrano, L., and Muñoz, V. (1997). Folding kinetics of Che Y mutants with enhanced native alpha-helix propensities. J. Mol. Biol. *266*, 610–620.
36. Filimonov, V.V., Prieto, J., Martinez, J.C., Bruix, M., Mateo, P.L., and Serrano, L. (1993). Thermodynamic analysis of the chemotactic protein from Escherichia coli, CheY. Biochemistry *32*, 12906–12921.
37. DeKoster, G.T., and Robertson, A.D. (1995). Cold denaturation of CheY. J. Mol. Biol. *249*, 529–534.
38. Garcia, P., Serrano, L., Durand, D., Rico, M., and Bruix, M. (2001). NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. Protein Sci. *10*, 1100–1112.
39. Muñoz, V., López, E.M., Jager, M., and Serrano, L. (1994). Kinetic characterization of the chemotactic protein from Escherichia coli, CheY. Kinetic analysis of the inverse hydrophobic effect. Biochemistry *33*, 5858–5866.
40. Baum, J., Dobson, C.M., Evans, P.A., and Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. Biochemistry *28*, 7–13.
41. Lakowicz, J.R. (1983). Principles of Fluorescence Spectroscopy (New York: Plenum), pp. 258–297.
42. Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins *6*, 87–103.
43. Bruix, M., Munoz, V., Campos-Olivas, R., DelBosque, J.R., Serrano, L., and Rico, M. (1997). Characterisation of the isolated Che Y C-terminal fragment (79–129)—exploring the structure/stability/folding relationship of the alpha/beta parallel protein Che Y. Eur. J. Biochem. *243*, 384–392.
44. Wolynes, P.G. (1997). Folding funnels and energy landscapes of larger proteins within the capillarity approximation. Proc. Natl. Acad. Sci. USA *94*, 6170–6175.
45. Pande, V.S., Grosberg, A., Tanaka, T., and Rokhsar, D.S. (1998). Pathways for protein folding: is a new view needed? Curr. Opin. Struct. Biol. *8*, 68–79.
46. Fersht, A.R. (1994). Characterizing transition states in protein folding: an essential step in the puzzle. Curr. Opin. Struct. Biol. *5*, 79–84.
47. Warren, J.R., and Gordon, J.A. (1970). Denaturation of globular proteins. II. The interaction of urea with lysozyme. J. Biol. Chem. *245*, 4097–4104.
48. Garcia, P., Desmadril, M., Minard, P., and Yon, J.M. (1995). Evidence for residual structures in an unfolded form of yeast phosphoglycerate kinase. Biochemistry *34*, 397–404.
49. Tanford, C. (1968). Protein denaturation. Adv. Protein Chem. *23*, 121–282.
50. Aune, K.C., and Tanford, C. (1969). Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II. Dependence on denaturant concentration at 25 degrees. Biochemistry *8*, 4586–4590.
51. Press, W.H., Flannery, B.P., Teulosky, S.K., and Veterling, W.T. (1986). Numerical Recipes in C: The Art of Scientific Computing (Cambridge: Cambridge University Press).
52. Minton, A.P. (1994). Conservation of signal: a new algorithm for the elimination of the reference concentration as an independently variable parameter in the analysis of sedimentation equilibrium. In Modern Analytical Ultracentrifugation, T. M. Schuster and T.M. Laue, eds. (Boston: Birkhauser), pp. 81–93.
53. Laue, T.M., Shak, B.D., Ridgeway, T.M., and Pelletier, S.M. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science, S.E. Harding, A.J. Rowe, and J.C. Horton, eds. (Cambridge: Royal Society of Chemistry), pp. 90–125.
54. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.
55. Schellman, J.A. (1978). Solvent denaturation. Biopolymers *17*, 1305–1322.
56. Santoro, J., Bruix, M., Pascual, J., Lopez, E., Serrano, L., and Rico, M. (1995). Three-dimensional structure of chemotactic Che Y protein in aqueous solution by nuclear magnetic resonance methods. J. Mol. Biol. *247*, 717–725.