The energetics of T4 lysozyme reveal a hierarchy of conformations

Manuel Llinás[1], Blake Gillespie[2], Frederick W. Dahlquist[2], and Susan Marqusee[1]

We have used native state exchange to examine the energy landscape of the well-characterized protein T4 lysozyme. Although the protein exhibits two-state behavior by traditional probes, the energy landscape determined here is much more complex. The average stability of the C-terminal subdomain is significantly higher than that for the N-terminus suggesting at least two regions of unfolding. At a more detailed level, there appears to be a broad continuum of stabilities throughout each region. The overall subdomain hierarchy of energies does not mirror data on the folding pathway for this protein, challenging the relationship between energy landscapes and folding trajectories.

Most small, single domain proteins unfold by an apparent two-state process suggesting that proteins populate one of two conformations: native or unfolded. Recently, however, it has been shown that under strongly native conditions, the unfolding energetics of each residue are not equivalent, suggesting that partially unfolded conformations are accessible[1–5]. These partially unfolded forms have been described as excited states since they are populated in the native state ensemble according to a Boltzmann distribution. For example, for a protein with a global stability ($\Delta G_{U\text{-}N}$) of 12 kcal mol^{-1} at 25 °C, 1 in 6×10^8 molecules will adopt the fully unfolded conformation. Any partially unfolded conformations within a few kilocalories of the unfolded conformation will also be rarely populated (less than ~1 in 10^6 molecules). Although the nature of these partially folded species and their relationship to kinetic folding intermediates are largely disputed[6–9], identifying regions of low stability defines an energy landscape and is important in determining whether a protein assembles via hierarchical folding pathways.

Experimentally, these rare, unfolded conformations are detected by their accessibility to hydrogen exchange, monitored at the level of individual residues by nuclear magnetic resonance (NMR) spectroscopy. The populations of these conformations can be modulated by environmental perturbation such as subdenaturing amounts of chemical denaturant. This experiment, termed native state hydrogen exchange, has proven to be extremely powerful for examining the stability of individual amino acids and identifying regions of global and subglobal stability. To date, native state hydrogen exchange has successfully identified rare partially unfolded conformations of only five proteins (RNase A, RNase H, barstar, cytochromes *c* and *b*562)[1,10–13]. In most of these cases, the results describe an hierarchical organization for unfolding. For cytochrome *c* and RNase H, this hierarchy resembles the kinetic refolding pathway determined by other methods, suggesting that this technique may be generally useful for defining folding reactions; however, caution must be exercised in any kinetic interpretation of equilibrium data.

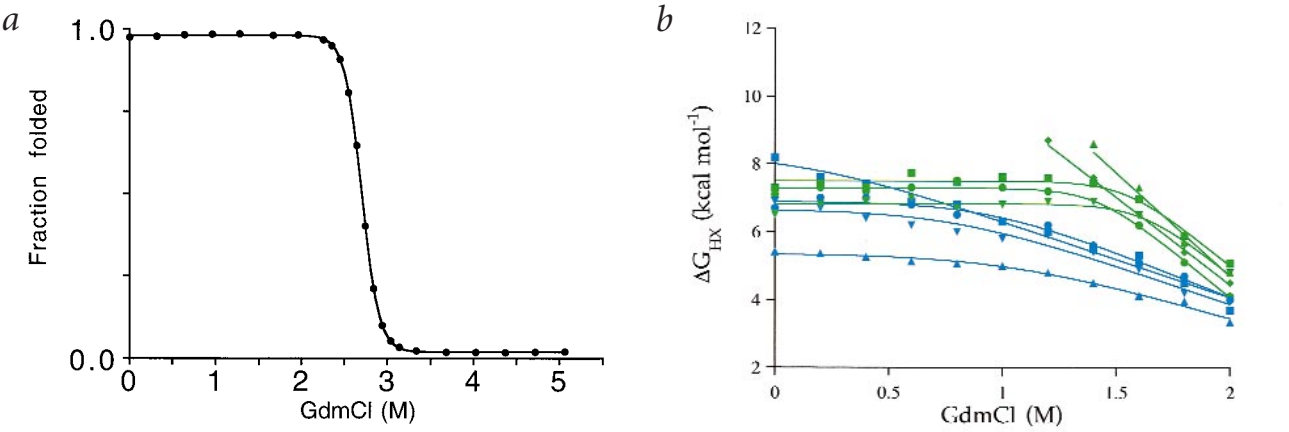

Fig. 1 Equilibrium unfolding of T4 lysozyme. ***a***, Guanidine (GdmCl) unfolding as measured by circular dichroism. ***b***, Free energy of hydrogen exchange (ΔG_{HX}) as a function of GdmCl for ten representative amides located throughout both the N-terminal subdomain (shown in blue are Lys 16 ●, Ile 27 ■, Glu 45 ▲, Ile 50 ▼) and C-terminal subdomain (shown in green are Gly 77 ■, Ala 98 ▲, Met 106 ▼, Phe 153 ◆, Glu 5 ●). The data have been fit to a model with contributions to exchange coming from both unfolding mechanisms and local fluctuations.

[1]Department of Molecular and Cell Biology, 229 Stanley Hall, University of California, Berkeley, Berkeley, California 94720, USA. [2]Institute of Molecular Biology and Departments of Chemistry and Physics, University of Oregon, Eugene, Oregon, 97403, USA.

Correspondence should be addressed to S.M. *email: marqusee@uclink4.berkeley.edu*

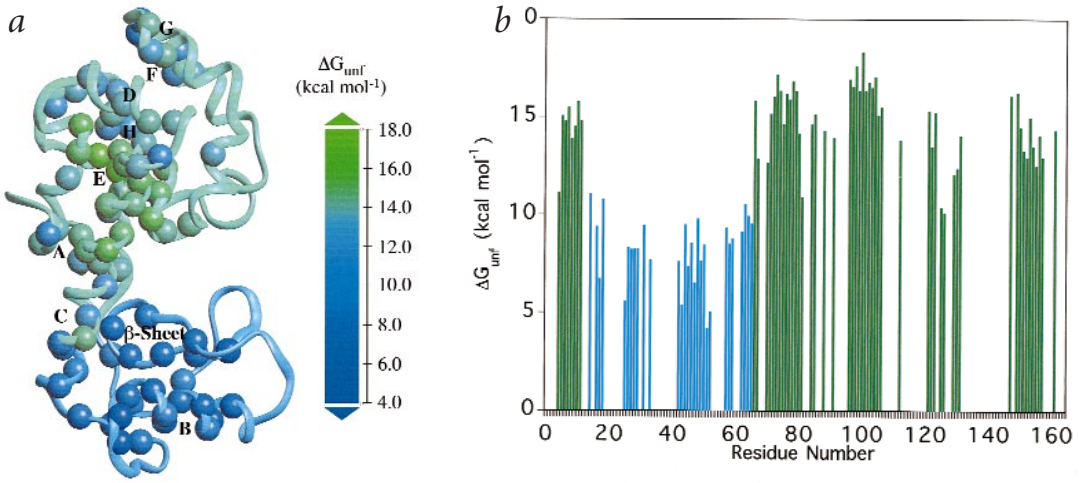

Fig. 2 Distribution of unfolding free energies (ΔG_{unf}) for T4 lysozyme. ***a***, Ribbon representation of the crystal structure of T4 lysozyme. Individual amide probes are shown as spheres and colored according to their unfolding free energy determined by native state hydrogen exchange. (Residues whose exchange was detected as occurring only through local fluctuations are not shown.) ***b***, Free energy values (ΔG_{unf}) plotted along the amino acid sequence. Two breaks are apparent (residue 12 and residue 65) separating the energetics into two regions, a stable C-terminal region (green) and a less stable N-terminal region (blue).

Here we provide a detailed and comprehensive native exchange profile for T4 lysozyme. T4 lysozyme is an extremely well-studied system — the folds and stabilities of literally hundreds of mutants have been examined[14,15]. Thermodynamic analyses of these mutants have relied on a two-state interpretation of the data[16]. Indeed, traditional equilibrium experiments fit the model well: T4 lysozyme behaves as a single domain protein unfolding in an apparently two-state manner with high cooperativity. In contrast to these thermodynamic studies, the crystal structure of T4 lysozyme clearly reveals two independent lobes[17]. These subdomains have been suggested by several different structural techniques[18–23] as well as computational studies[24–28].

Using native state exchange, we have finally been able to distinguish regions of differing stability in T4 lysozyme. The results from this thermodynamic analysis support the structural studies suggesting independence between the domains; they do not, however, directly correlate with the previously described kinetic folding intermediate of T4 lysozyme[29]. This work illustrates that the kinetic folding pathway of a protein depends on specific barriers and the relationship between the folding mechanism and the energy landscape can be obtained only by comparing a variety of experiments such as presented here.

Native state hydrogen exchange

The exchange of 114 of the 164 backbone amide protons in T4 lysozyme was monitored as a function of denaturant concentration (eleven samples between 0 M and 2 M GdmCl), making this the most complete native state hydrogen exchange data set to date. All of these denaturant concentrations are below the folding transition measured by global probes (Fig. 1*a*), and hence the protein is predominantly in the native conformation. Exchange was measured by two-dimensional ^{1}H-^{15}N HSQC spectra taken over a period of hours to months.

The 114 protons all show measurable protection in the native state. The observed exchange rates were analyzed according to the classical scheme for amide hydrogen exchange, where the exchange for solvent deuterons can occur only when a protein undergoes a transition from a 'closed' state to an 'open', exchange-competent state[30,31]:

$$\text{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{open} \xrightarrow{k_{int}} \text{exchanged}$$

k_{cl} and k_{op} are the closing and opening rates for this transition and k_{int} is the intrinsic rate for a particular amide at the defined pH, temperature, and solvent conditions[32]. The transition from the closed to open state can result from: (i) gross structural unfolding events (global or patial unfolding) or (ii) small local fluctuations within the native conformation[11,33,34].

Under EX2 conditions ($k_{int} << k_{cl}$)[30], the measured rate of amide hydrogen exchange (k_{obs}) is related to the equilibrium constant ($K_{op} = k_{obs}/k_{int}$) between the open and closed states[30], and therefore a free energy for exchange ($\Delta G_{HX} = -RT \ln K_{op}$) can be determined for each residue. For most proteins at pH 6.0, this EX2 assumption holds under native conditions. Without direct measurements for k_{cl}, however, it is difficult to determine if the EX2 assumption is valid under all of the conditions of our native state exchange study. We can, however, make some estimates for k_{cl} using the refolding rates determined as a function of denaturant for T4 lysozyme. In the absence of denaturant, we assume that the slowest k_{cl} is that of refolding, 50 s^{-1} (unpublished data). Under these conditions, the average k_{int} is approximately 1.0 s^{-1} and the EX2 assumption should be valid. This same argument holds for all conditions ≤1.0 M GdmCl; however, at 1.6 M GdmCl, the refolding kinetics suggest that $k_{cl} \approx k_{int}$. Even under these conditions, deviations from EX2 behavior will result in only a small (≤0.4 kcal mol^{-1}) overestimation of the stability of a proton and do not grossly affect the analyses.

The free energy of hydrogen exchange, ΔG_{HX}, represents a combination of opening transitions, both structural unfolding and local fluctuations[1,11,34]. Evaluating the rates of exchange as a function of [GdmCl] allows us to distinguish and separate these two processes: addition of GdmCl promotes exposure of non-polar surface area and therefore alters the equilibrium constant for unfolding (K_{unf}), but by definition does not affect local fluctuations (K_{fl}). Hence, exchange that appears denaturant-independent is attributed to local events that do not expose new non-polar surface area and have fluctuation energies (ΔG_{fl}) below those for unfolding. With the addition of denaturant, the free energy of unfolding drops and eventually the exchange process is dominated by these unfolding events. The hydrogen exchange behavior of several amide proteins in T4 lysozyme is shown in Fig. 1*b*. By fitting the data to this two-process model (see Methods), free energies were obtained for both local fluctuations and larger unfolding events. For the

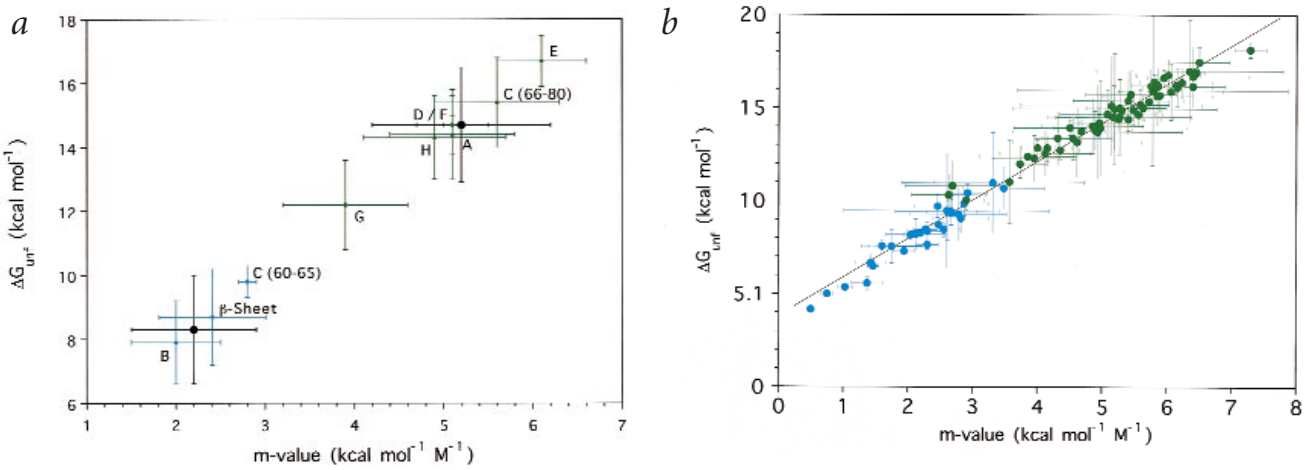

Fig. 3 The calculated unfolding free energies (ΔG_{unf}) and corresponding m-values define a broad continuum of stabilities. ***a***, The average free energy of exchange for each secondary structural element is depicted. The overall average value for each subdomain is represented by the larger filled black circles. ***b***, The individual unfolding exchange free energies for all measured amides are shown illustrating the range of the continuum.

unfolding events, we report both the extrapolated free energy of unfolding (ΔG_{unf}) in the absence of denaturant and the corresponding dependence on denaturant, or m-value (Table 1).

The importance of evaluating the denaturant dependence of exchange is also illustrated by Fig. 1*b*. For example, in the absence of denaturant, the free energy of exchange (ΔG_{HX}) for Met 106 is 6.8 kcal mol^{-1}, well below that for Ile 27 (ΔG_{HX} = 8.2 kcal mol^{-1}). Their extrapolated unfolding stabilities (ΔG_{unf}), however, are just the opposite. At low denaturant, exchange of Met 106 is dominated by local fluctuations, masking its high unfolding stability. These data highlight a major flaw in simply correlating protection factors to regional stability within a protein. If protection factors alone were being calculated, Ile 27 would appear to be the more stable residue.

Subdomain organization of T4 lysozyme

Our results clearly indicate that the energetics of T4 lysozyme can be divided into two separate regions (Fig. 2): an N-terminal (residues 13–65) and a C-terminal subdomain (residues 1–12 and 65–164). The average unfolding free energy for residues in the N-terminal subdomain (blue) is 8.3 (±1.7) kcal mol^{-1}, while that for residues in the C-terminal subdomain (green) is 14.7 (±1.8) kcal mol^{-1}. These results agree well with a protein dissection study on T4 lysozyme, which demonstrated that only the C-terminal subdomain was capable of folding in isolation[23]. The separation of these two subdomains is seen as a change in the energetics along helix C, which spans both regions. While individual helices are usually found to be cooperative, this long helix shows an unusual break near residue 65. The position of this break was unexpected (see below).

Helices E, H, A and the C-terminal end of helix C come together to form the most stable portion of T4 lysozyme, which we refer to as the C-terminal subdomain. The data clearly indicate that, in addition to the C-terminal regions, helix A from the N-terminus contributes to this domain. The average ΔG_{unf} for all of the protons in this region correlates well with the global stability of T4 lysozyme under the same conditions (16.0 kcal mol^{-1}). This implies that the unfolding events that give rise to hydrogen exchange coincide on average with the global unfolding of the molecule.

In contrast, residues in the N-terminal region of the protein (the β-sheet region) show much lower extrapolated ΔG_{unf}. Even though the N-terminal region is substantially less stable than the C-terminal region (average $\Delta\Delta G_{unf}$ = 6.3 kcal mol^{-1}), unfolding and hydrogen exchange in this N-terminal domain is still rare (1 in 5 × 10^7). This rare event results in a partially unfolded conformation comprised of a folded C-terminal subdomain and an unfolded N-terminal region. Hence, the structural independence inferred from the crystallography is mirrored in the distribution of energy within the protein.

The stability of T4 lysozyme has also been evaluated by the structure-based statistical thermodynamics algorithm COREX, developed by Freire and coworkers[28,35] . This algorithm calculates a stability parameter, which is then related to a hydrogen exchange protection factor (K_{op}^{-1}), for each residue in a protein. The results from our experiments allow a detailed evaluation of the relationship between the residue stability parameters determined from COREX and accessibility to hydrogen exchange. In general, COREX does a good job of predicting the lower stability we see for the N-terminal subdomain. While it accurately identifies the E-helix as the most stable region of the protein, the relative contributions of the other helices are not accurately identified. The detailed data set of hydrogen exchange rates we present here will allow an in depth evaluation of the COREX algorithm and the assignment of protection factors from calculated stability parameters.

Recently, we investigated the independent folding of fragments corresponding to the two regions of T4 lysozyme; only the C-terminal subdomain was capable of autonomous folding[23]. For these fragment studies the subdomain boundary was placed at residue 75. Our native-state exchange results, however, indicate that this break occurs at residue 65 and therefore extends the C-terminal subdomain further along the C-helix. These additional eleven residues, 65–75, may also contribute to the energetics of the C-terminal subdomain, suggesting a likely explanation for the low stability of the previously studied C-terminal fragment.

Discrete intermediates or a continuum

Although the average stabilities measured by hydrogen

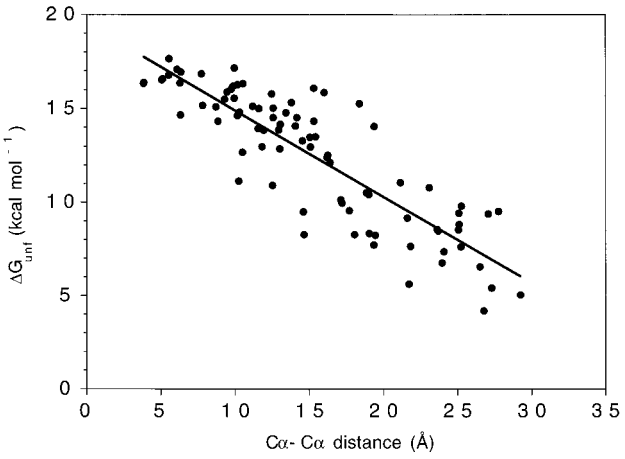

Fig. 4 The measured free energy of exchange (ΔG_{unf}) is correlated with the Cα–Cα distance from Ile 100 in the E-helix. This helix is the most stable region of the protein and may serve as the nucleation site for the folding of T4 lysozyme. Interatomic Cα–Cα distances were calculated using EDPDB[41] with the WT* T4 lysozyme crystal structure (PDB accession number 1l63)[42] as input.

exchange imply two separable subdomains (Fig. 2), the stabilities for the individual residues within each region span a large range, much larger than the errors associated with the determination of each ΔG_{unf}. The average extrapolated free energy of unfolding (ΔG_{unf}) and m-value for each element of secondary structure in the protein is shown in Fig. 3*a*. With the exception of the long C-helix, the stabilities of each secondary structure appear to cluster, implying cooperative unfolding of the individual helices. While it is clear that the C-terminal subdomain is more stable than the N-terminal subdomain, a distinct clustering of each subdomain is not obvious. For instance the E-helix appears significantly more stable (and has a higher m-value) than the G-helix. This broad distribution in unfolding energies is even more apparent when the individual unfolding free energies and corresponding m-values are plotted for each residue independently.

This observed continuum of energies suggests that the assignment of discrete partially unfolded forms may be an overly simplistic interpretation of the data. Alternatively, this continuum could be an artifact due to the long extrapolation from data taken in higher denaturant concentrations. Errors in the determination of the slope (m-value) would result in a range of calculated stabilities. This is unlikely to be the basis for the observed distribution in stability. First, attempts to fit all the data in the C-terminal subdomain to a single ΔG_{unf} and m-value, only allowing the local fluctuations to vary, were unsuccessful (data not shown). Second, such correlated errors would be unlikely to have a structural interpretation; residues in the E-helix would not be significantly higher than residues in the G-helix. Moreover, there is no correlation between residues with unusually high or low stabilities and the degree of uncertainty in their m-values (Fig. 3*b*). Hence, while it is clear that residues in the C-terminal subdomain are more stable than those in the N-terminal region, the distribution of stabilities within these subdomains appears to be significant.

A suprisingly linear correlation between ΔG_{unf} and m-value is also revealed in Fig. 3*b*. It is very difficult to make conclusions based on this linearity, since for each residue both parameters (ΔG_{unf} and m-value) are derived from a least squares fit to a single data set. A correlation between two parameters extracted from one independent variable can be a result of the so-called 'isokinetic effect'. This artifactual correlation is most significant when the variations in the parameter estimates are as small as the statistical uncertainty in each estimate[36]. This however, does not appear to be the case in our study. While the statistical uncertainties for the individual residue stabilities are quite large (as much as ±3 kcal mol^{-1}), the overall range is greater than 10 kcal mol^{-1}. Nevertheless, this statistical compensation undoubtedly plays a dominant role and it is therefore difficult to assess the true functional dependence between ΔG_{unf} and m-value.

Additional insight into a structural interpretation of the distribution in ΔG_{unf} values is seen in Fig. 4 as a plot of ΔG_{unf} for each residue shown as a function of the distance between the Cα atoms of Ile 100 and the residue in question. We chose this residue as a reference point because it, like other members of the E helix, is one of the residues most protected from hydrogen exchange and it lies near the center of the E helix. There is an apparent linear decrease in the free energy value as a function of the distance to residue 100. Using other E helix residues as the reference point gives a very similar plot while using residues that are more moderate in their exchange properties gives more complex curves. These data support a hierarchical view of the energetics of partially folded forms of T4 lysozyme that lead to progressive solvent protection and protection from exchange. In this view, the E helix and neighboring residues form a structural and energetic core about which additional structural components are added. These additional components add stability and increase the size of the effective core. The process of progressive addition of structure and the corresponding increases in stability continue until the final folded state is formed as regions most distant from the E helix are added. Since the native coordinates were used to produce Fig. 4, it appears that the various intermediate forms proposed have essentially native-like structure.

Local fluctuations

For the majority of the protons measured (77%), the dominant mechanism for exchange under native conditions (0 M GdmCl) is by local fluctuation, ΔG_{fl} (Table 1). Only sixteen of the probes in the C-terminal subdomain do not demonstrate this mode of exchange, indicating that even in the most stable regions of the protein, the dynamics are dominated by these more local events. A total of 26 probes in the protein show exchange exclusively by these local fluctuations, with no observable denaturant dependence to their exchange. For these residues, local fluctuations are so dominant (low energy) that unfolding events cannot be detected. Most of these residues are found in highly surface exposed areas. In addition, of the 50 residues that are not probes in this study, many are only weakly protected, either because exchange is dominated by local fluctuations with free energies below the detection limit (2.4 kcal mol^{-1}), or because they are completely unprotected in the native conformation. The nature of exchange by these local fluctuations is not well understood although we believe that, due to the lack of denaturant dependence, it must not involve major conformational changes.

The free energies determined for these fluctuations do not show any direct correlation with the thermal B-values determined by X-ray crystallography. Therefore, the factors determining the exchange of protons in the native state are different from thermal factors that also depend on a host of different factors.

Table 1 Calculated $\Delta G_{unf}(H_2O)$, m-value and ΔG_{fl} energies for T4 lysozyme

Residue	$\Delta G_{unf}(H_2O)$ (kcal mol^{-1})	m-value (M^{-1})	ΔG_{fl} (kcal mol^{-1})	2° Structure	Residue	$\Delta G_{unf}(H_2O)$ (kcal mol^{-1})	m-value (M^{-1})	ΔG_{fl} (kcal mol^{-1})	2° Structure
Phe 4	11.1	3.6	4.5	Helix A	Ser 90			5.0	Helix D
Glu 5	15.0	5.5	7.3	Helix A	Leu 91	14.0	4.9	6.8	
Met 6	14.8	5.1	7.7	Helix A	Val 94			2.4	Helix E
Ile 7	15.5	5.7	7.9	Helix A	Arg 96	16.9	6.0	9.1	Helix E
Arg 8	13.9	4.7		Helix A	Ala 97	16.6	5.8		Helix E
Ile 9	14.5	5.3		Helix A	Ala 98	17.7	6.5		Helix E
Asp 10	15.8	5.9	7.1	Helix A	Leu 99	16.1	6.0		Helix E
Glu 11	14.8	5.6	7.6	Helix A	Ile 100	18.3	7.3		Helix E
Gly 12			4.3		Asn 101	16.4	5.8		Helix E
Arg 14	11.1	3.3	5.1	Sheet	Met 102	16.8	6.0		Helix E
Lys 16	9.4	2.7	6.9	Sheet	Val 103	16.5	6.2		Helix E
Ile 17	6.8	1.4	5.1	Sheet	Phe 104	17.1	6.5		Helix E
Tyr 18	10.8	3.5	5.5	Sheet	Gln 105	15.1	5.2	6.3	Helix E
Asp 20			3.5	Sheet	Met 106	15.6	5.4	6.8	Helix E
Tyr 25	5.6	1.4	4.4	Sheet	Gly 107			5.0	
Thr 26	8.3	2.2		Sheet	Val 111			4.0	
Ile 27	8.2	2.0		Sheet	Ala 112	13.9	4.9	5.1	
Gly 28	8.3	2.1			Leu 121	15.3	5.6	7.0	Helix F
Ile 29	7.7	2.1	4.8		Gln 122	13.5	4.6	6.1	Helix F
His 31	9.5	2.7	5.7	Sheet	Gln 123	15.3	5.1	6.1	Helix F
Leu 33	7.7	2.3	6.4	Sheet	Lys 124			4.6	
Thr 34			2.8	Sheet	Arg 125	10.4	2.6	5.9	
Ala 42	7.6	1.8	4.3	Helix B	Trp 126	10.1	2.9	5.2	Helix G
Lys 43	5.4	1.0		Helix B	Ala 129	12.1	3.7	6.6	Helix G
Ser 44	9.5	2.6	5.0	Helix B	Ala 130	12.4	4.0	5.8	Helix G
Glu 45	7.4	1.9	5.4	Helix B	Val 131	14.1	5.0	5.5	Helix G
Leu 46	8.6	2.6	6.1	Helix B	Asn 132			3.8	Helix G
Asp 47	6.6	1.5		Helix B	Leu 133			4.0	Helix G
Lys 48	9.8	2.5	6.8	Helix B	Ala 134			4.6	Helix G
Ala 49	7.7	1.6	6.3	Helix B	Lys 135			4.0	
Ile 50	8.5	2.3	6.7	Helix B	Trp 138			3.9	
Gly 51	4.2	0.5			Tyr 139			4.3	
Val 57	9.4	2.8	4.3	Sheet	Asn 140			4.9	
Ile 58	8.5	2.3	7.0	Sheet	Gln 141			4.0	
Thr 59	8.8	2.5	7.8		Thr 142			4.4	Helix H
Glu 62	9.2	2.8	4.4	Helix C	Ala 146			6.5	Helix H
Ala 63	10.5	2.9	8.3	Helix C	Lys 147	16.1	5.8	5.8	Helix H
Glu 64	10.0	2.9	5.7	Helix C	Arg 148			5.3	Helix H
Lys 65	9.6	2.6	4.7	Helix C	Val 149	16.3	6.2	6.7	Helix H
Leu 66	15.9	5.9	7.5	Helix C	Ile 150	14.5	5.4		Helix H
Phe 67	12.9	4.4	7.8	Helix C	Thr 151	13.3	4.6	5.9	Helix H
Asn 68			5.5	Helix C	Thr 152	13.0	4.0		Helix H
Gln 69			4.8	Helix C	Phe 153	15.0	5.3		Helix H
Asp 70	12.7	4.1	7.6	Helix C	Arg 154	13.5	4.3		Helix H
Val 71	15.2	5.6	7.1	Helix C	Thr 155	12.5	3.9	5.7	Helix H
Asp 72	16.0	6.1	6.0	Helix C	Gly 156	14.1	4.5	9.1	
Ala 73	17.2	6.4	6.3	Helix C	Thr 157	13.0	4.2	6.5	
Ala 74	16.4	5.9	8.4	Helix C	Ala 160			4.0	
Val 75	14.7	5.3	6.7	Helix C	Tyr 161	14.4	5.0	6.9	
Arg 76	16.2	5.8		Helix C	Lys 162			3.2	
Gly 77	15.9	5.4	7.5	Helix C					
Ile 78	16.9	6.4	6.9	Helix C					
Leu 79	16.3	6.4		Helix C					
Arg 80	14.2	4.9	6.2	Helix C					
Asn 81	10.9	2.7	6.7						
Leu 84	14.6	5.2	5.8	Helix D					
Lys 85	15.1	5.3	8.0	Helix D					
Val 87			3.3	Helix D					
Tyr 88	14.3	5.0		Helix D					
Asp 89			6.7	Helix D					

Relationship to protein folding kinetics

The folding of many proteins involves the transient formation of partially folded intermediates[37,38]. Recently, there has been some debate as to whether the partially unfolded conformations determined by native state hydrogen exchange mimic those intermediates seen in the refolding pathway of proteins[6]. For the two best characterized examples, RNase H and cytochrome *c*, there does appear to be some correlation between the two[1,11]. For example, in RNase H, native state hydrogen exchange identified a rare, partially unfolded conformation whose structure resembled both the kinetic folding intermediate and the low pH molten globule form[33]. Similar results were seen for cytochrome *c*, implying that the partially unfolded forms of proteins determined by native state exchange may represent kinetically accessible intermediates present during the folding reaction. As shown here, however, this correlation cannot necessarily be generalized for all proteins.

There is substantial evidence that T4 lysozyme forms a kinetic intermediate during its folding reaction. In an analysis of several mutants at low temperature, Chen *et al.*[39] observed a variable effect on the rate limiting step of folding, implying the presence of an intermediate. This kinetic intermediate was further characterized by pulse-labeling hydrogen exchange[29]. Rapid protection (<10 ms) was observed most strongly in the E-helix of the C-terminal subdomain. Residues in the β-sheet of the N-terminal subdomain also showed strong protection at early time points. This pulse-labeling study, therefore, does not directly correlate with the thermodynamic results reported here. While we find the C-terminal region is most stable, the early folding intermediate shows protection from residues in both the N- and C-terminal subdomains of T4 lysozyme. In addition, the E-helix is the only region in the C-terminal subdomain to show protection in the kinetic intermediate. Therefore, the kinetic folding intermediate does not mimic the partially unfolded conformations of T4 lysozyme present under native conditions. These results underscore the fact that the native state exchange experiment can expose free energy minima on either side of the folding transition state and not necessarily reveal kinetic folding intermediates.

Conclusions

A comprehensive understanding of the native state of proteins is imperative to our understanding of protein folding and to the engineering of novel proteins. Native state hydrogen exchange allows an in-depth analysis of the structural energetics of proteins. The current study of T4 lysozyme is the most extensive examination of a protein under native conditions to date. A clear picture emerges in which the two structural subdomains behave as energetically independent units. Together, the A, E, C and H helices form the most stable element with a stability on the order of the global unfolding free energy for T4 lysozyme. Despite the energetic distinction between the N- and C-terminal subdomains of T4 lysozyme, the data presented herein demonstrate a broad continuum of stabilities supporting a hierarchy of protein stability. The differences between the kinetic and equilibrium studies underscore the difficulty in using exclusively thermodynamic data to describe a kinetic folding pathway.

Methods

Protein sample preparation. WT* (cysteine-free (C54T, C97A)) T4 lysozyme was overexpressed in BL21/DE3 pLysS cells grown in M9 minimal media with ^{15}N ammonium chloride. Protein purification was as described[23] except that the protein was dialyzed extensively against 50 mM ammonium bicarbonate and lyophilized. All samples used for the exchange experiment were from the same protein stock (1.73 mM) in 50 mM sodium phosphate, pH 6.0, 0.1 M KCl. Varying amounts of buffered GdmCl (final 0.2–2.0 M) and buffer were added to give a final protein concentration of 1.25 mM.

Hydrogen-deuterium exchange. Exchange was initiated by spinning the samples through a Quick-sep spin column (IsoLab Inc.) packed with G-25 resin (Pharmacia) pre-equilibrated in deuterated buffer (50 mM sodium phosphate, pD 6.0, 0.1 M KCl) with varying amounts of per-deuterated GdmCl. Per-deuterated GdmCl was made by exchanging 50 g of ultra pure GdmCl (ICN Biomedicals Inc.) in 99.9% deuterium oxide (Isotec Inc.) by three successive 60 ml washes followed by rotary evaporation and subsequent lyophilization. Correction for the isotope effect was made using pD = pH(read) + 0.4. The initiation of exchange (t_0) was marked after the 1.5 min spin.

Amide hydrogen exchange was monitored by collecting HSQC spectra. For each spectrum, 4 scans and 2,048 points were collected in the direct dimension with 128 points in the indirect dimension. A total of 17–25 HSQC spectra were collected for each sample at 25 °C over a period of five months on a Varian Inova-600 MHz 4 channel NMR spectrometer equipped with a 5 mm ^1H ($^{13}C/^{15}N$) PFG triple-NMR probe. The data were processed using Felix 97.0. The assignment of the measured peaks at pH 6.0 was accomplished by comparison with the previously determined assignments at pH 5.6 (ref. 40). No significant pH-induced peak shifts were detected, allowing a direct assignment of the higher pH spectrum. Peak intensities were measured using the Felix macro XPKvolheight written by Michael Akke.

Amide exchange rates (k_{obs}) were calculated by fitting the peak intensities as a function of exchange time to a single exponential using Kaleidagraph (Synergy Software, Reading, Pennsylvania). K_{op} was calculated using $K_{op} = (k_{obs}/k_{int})$ where k_{int} is the known intrinsic exchange rates for a residue in a specific tripeptide context[32]. The resulting free energy of hydrogen exchange, $\Delta G_{HX} = -RT \ln(K_{op})$, was then plotted as a function of denaturant for each probe and was fit according to the equation $\Delta G_{HX} = -RT \ln(K_{unf}+K_{fl})$ where $K_{unf} = \exp((m \times [GdmCl]-\Delta G_{unf})/RT)$ and $K_{fl} = \exp(-\Delta G_{fl}/RT)$. Here, ΔG_{unf} represents the unfolding free energy extrapolated to 0 M denaturant, while m represents the denaturant dependence of the unfolding free energy. All data were fit by this model except in cases where exchange did not show any notable denaturation-independent portion, in which case the data were simply fit to a straight line $\Delta G_{HX} = \Delta G_{unf} - m[GdmCl]$. Fits were deemed unacceptable if the error in ΔG_{unf} was greater than ±3.0 kcal mol^{-1}. The ΔG_{unf} for most residues is well below this limit. The ΔG_{fl} fit error was less than ±0.2 kcal mol^{-1} in all cases.

Acknowledgments

We thank K. Fischer, I. Griswold, J. Hollien, G. Lazar, and M.J. Parker for discussion and critical reading of the manuscript. We also thank A.J. Wand and E. Fuentes for discussion about data analysis. This work was supported by grants from the National Institutes of Health (S.M. and F.W.D.) and the W. M. Keck Foundation. M.L. was supported by an NIH Molecular Biophysics Training Grant and B.G. was supported by a Department of Health and Human Services, National Research Service Award.

Received 9 March, 1999; accepted 5 August, 1999.

1. Chamberlain, A.K., Handel, T.M. & Marqusee, S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nature Struct. Biol.* **3**, 782–787 (1996).
2. Clarke, J. & Itzhaki, L.S. Hydrogen exchange and protein folding. *Curr. Opin. Struct. Biol.* **8**, 112–118 (1998).
3. Raschke, T.M. & Marqusee, S. Hydrogen exchange studies of protein structure. *Curr. Opin. Biotechnol.* **9**, 80–86 (1998).
4. Englander, S.W., Sosnick, T.R., Englander, J.J. & Mayne, L. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6**, 18–23 (1996).
5. Parker, M.J., Dempsey, C.E., Hosszu, L.L., Waltho, J.P. & Clarke, A.R. Topology, sequence evolution and folding dynamics of an immunoglobulin domain. *Nature Struct. Biol.* **5**, 194–198 (1998).
6. Clarke, J., Itzhaki, L.S. & Fersht, A.R. Hydrogen exchange at equilibrium: a short cut for analysing protein-folding pathways? *Trends Biochem. Sci.* **22**, 284–287 (1997).
7. Englander, S.W. Native-state HX. *Trends Biochem. Sci.* **23**, 378; (1998).
8. Woodward, C. & Li, R. The slow-exchange core and protein folding. *Trends Biochem. Sci.* **23**, 379–381 (1998).
9. Clarke, J., Itzhaki, L.S. & Fersht, A.R. A reply to Englander and Woodward. *Trends Biochem. Sci.* **23**, 379–381 (1998).
10. Mayo, S.L. & Baldwin, R.L. Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science* **262**, 873–876 (1993).
11. Bai, Y., Sosnick, T.R., Mayne, L. & Englander, S.W. Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192–197 (1995).
12. Fuentes, E.J. & Wand, A.J. Local dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* **37**, 3687–3698 (1998).
13. Bhuyan, A.K. & Udgaonkar, J.B. Two structural subdomains of barstar detected by rapid mixing NMR measurement of amide hydrogen exchange. *Proteins* **30**, 295–308 (1998).
14. Matthews, B.W. Studies on protein stability with T4 lysozyme. *Adv. Protein Chem.* **46**, 249–278 (1995).
15. Matthews, B.W. Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB J.* **10**, 35–41 (1996).
16. Elwell, M. & Schellman, J. Phage T4 lysozyme. Physical properties and reversible unfolding. *Biochim. Biophys. Acta* **386**, 309–323 (1975).
17. Weaver, L.H. & Matthews, B.W. Structure of bacteriophage T4 lysozyme refined at 1.7 A resolution. *J. Mol. Biol.* **193**, 189–199 (1987).
18. Faber, H.R. & Matthews, B.W. A mutant T4 lysozyme displays five different crystal conformations. *Nature* **348**, 263–266 (1990).
19. Dixon, M.M., Nicholson, H., Shewchuk, L., Baase, W.A. & Matthews, B.W. Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3 to Pro. *J. Mol. Biol.* **227**, 917–933 (1992).
20. McHaourab, H.S., Lietzow, M.A., Hideg, K. & Hubbell, W.L. Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry* **35**, 7692–7704 (1996).
21. McHaourab, H.S., Oh, K.J., Fang, C.J. & Hubbell, W.L. Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry* **36**, 307–316 (1997).
22. Anderson, D., Becktel, W. & Dahlquist, F. The folding, stability, and dynamics of T4 lysozyme: A perspective using nuclear magnetic resonance. In *NMR of proteins.* (eds Clore, G. & Gronenborn, A.) 258–304 (CRC Press, Boca Raton, Florida; 1993).
23. Llinás, M. & Marqusee, S. Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Sci.* **7**, 96–104 (1998).
24. Bahar, I., Erman, B., Haliloglu, T. & Jernigan, R.L. Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry* **36**, 13512–13523 (1997).
25. de Groot, B.L., Hayward, S., van Aalten, D.M., Amadei, A. & Berendsen, H.J. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. *Proteins* **31**, 116–127 (1998).
26. Arnold, G.E., Manchester, J.I., Townsend, B.D. & Ornstein, R.L. Investigation of domain motions in bacteriophage T4 lysozyme. *J. Biomol. Struct. Dyn.* **12**, 457–474 (1994).
27. Arnold, G.E. & Ornstein, R.L. Protein hinge bending as seen in molecular dynamics simulations of native and M61 mutant T4 lysozymes. *Biopolymers* **41**, 533–544 (1997).
28. Hilser, V.J., Townsend, B.D. & Freire, E. Structure-based statistical thermodynamic analysis of T4 lysozyme mutants: structural mapping of cooperative interactions. *Biophys. Chem.* **64**, 69–79 (1997).
29. Lu, J. & Dahlquist, F.W. Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry* **31**, 4749–4756 (1992).
30. Hvidt, A. & Nielsen, S.O. Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**, 287–386 (1966).
31. Englander, S.W. & Kallenbach, N.R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16**, 521–655 (1983).
32. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
33. Chamberlain, A.K. & Marqusee, S. Touring the landscapes: partially folded proteins examined by hydrogen exchange. *Structure* **5**, 859–863 (1997).
34. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. Protein stability parameters measured by hydrogen exchange. *Proteins* **20**, 4–14 (1994).
35. Hilser, V.J. & Freire, E. Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* **262**, 756–772 (1996).
36. Krug, R.R., Hunter, W.G. & Greiger, R.A. Enthalpy-entropy compensation. 1. some fundamental statistical problems associated with the analysis of van't Hoff and Arrhenius data. *J. Phys. Chem.* **80**, 2335–2341 (1976).
37. Jackson, S.E. How do small single domain proteins fold? *Folding & Design* **3**, R81–91 (1998).
38. Roder, H. & Colon, W. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15–28 (1997).
39. Chen, B.L., Baase, W.A. & Schellman, J.A. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry* **28**, 691–699 (1989).
40. McIntosh, L.P., Wand, A.J., Lowry, D.F., Redfield, A.G. & Dahlquist, F.W. Assignment of the backbone 1H and 15N NMR resonances of bacteriophage T4 lysozyme. *Biochemistry* **29**, 6341–6362 (1990).
41. Zhang, X.J. & Matthews, B.W. EDPDB: a multifunctional tool for protein structure analysis. *J. Appl. Crystallogr.* **28**, 624 (1995).
42. Nicholson, H., Anderson, D.E., Dao-pin, S. & Matthews, B.W. Analysis of the interaction between charged side chains and the alpha-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* **30**, 9816–9828 (1991).