doi:10.1016/j.jmb.2004.08.005 J. Mol. Biol. (2004) 343, 223–233

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

How Cytochrome *c* Folds, and Why: Submolecular Foldon Units and their Stepwise Sequential Stabilization

Haripada Maity*†, Mita Maity† and S. Walter Englander

Johnson Research Foundation Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

**Corresponding author*

Native state hydrogen exchange experiments have shown that the cytochrome *c* (Cyt *c*) protein consists of five cooperative folding–unfolding units, called foldons. These are named, in the order of increasing unfolding free energy, the nested-Yellow, Red, Yellow, Green, and Blue foldons. Previous results suggest that these units unfold in a stepwise sequential way so that each higher energy partially unfolded form includes all of the previously unfolded lower free energy units. If this is so, then selectively destabilizing any given foldon should equally destabilize each subsequent unfolding step above it in the unfolding ladder but leave the lower ones before it unaffected. To perform this test, we introduced the mutation Glu62Gly, which deletes a salt link in the Yellow unit and destabilizes the protein by 0.8 kcal/mol. Native state hydrogen exchange and other experiments show that the stability of the Yellow unit and the states above it in the free energy ladder are destabilized by about the same amount while the lower lying states are unaffected. These results help to confirm the sequential stepwise nature of the Cyt *c* unfolding pathway and therefore a similar refolding pathway. The steps in the pathway are dictated by the concerted folding–unfolding property of the individual unit foldons; the order of steps is determined by the sequential stabilization of progressively added foldons in the native context. Much related information for Cyt *c* strongly conforms with this mechanism. Its generality is supported by available information for other proteins.

Keywords: protein folding; hydrogen exchange; cytochrome *c*; stability labeling; alkaline transition

Introduction

How do proteins fold? The classical macroscopic view is that proteins are guided downhill to the native state through a pathway of more or less distinct intermediate forms.[1–4] A newer view has figured prominently in the search for the microscopic details of folding processes.[5–12] One interpretation considers that proteins might fold through multiple indefinite trajectories and intermediates guided simply by the downhill converging nature of the folding energy landscape.

To distinguish these alternative mechanisms and other uncertainties, one would like to detect and characterize the intermediates that in fact determine the protein folding process. This has been exceedingly difficult. The entire folding process is usually complete in less than one second. During kinetic folding, intermediates often do not accumulate. Even when they do accumulate, they cannot be isolated and studied by the usual structural methods. Methods that are fast enough to follow kinetic folding behavior, mainly spectroscopic in nature, provide little detailed structural information.

Fortunately another option exists. Thermodynamic principles require that proteins must continually unfold and refold even under native conditions, albeit at a very low level. In this process all possible higher energy intermediate states are

† H.M. and M.M. contributed equally to this work.

Abbreviations used: Cyt *c*, cytochrome *c*; pWT, pseudo wild-type; N, native state; U, globally unfolded state; foldon, concerted folding–unfolding unit, written as nested-Yellow (N), Red (R), Yellow (Y), Green (G) and Blue (B); PUF, partially unfolded form consisting of one or more foldon units; HX, hydrogen exchange; EX1 exchange, HX rate determined by structural unfolding rate; EX2 exchange, HX rate determined by a structural unfolding pre-equilibrium; GdmCl, guanidinium chloride; GdmSCN, guanidinium thiocyanate.

E-mail address of the corresponding author: hmaity@hx2.med.upenn.edu

occupied according to their equilibrium Boltzmann factor and all of the molecules continually cycle through all of these states. Thus intermediates that determine the folding process might in principle be studied over times much longer than one second. However, the infinitesimally populated high energy states are invisible to most measurements, which detect only the overwhelmingly populated native state. This is not true for hydrogen exchange (HX). HX measurements receive no contribution from the unperturbed native state. HX rates of hydrogens that are protected in the native state are wholly determined by deprotection (H-bond breaking) reactions that occur when proteins transiently visit their higher energy states.

In favorable cases, the high energy intermediate forms can be studied by a native state HX (NHX) method. Previous NHX experiments have found that amide hydrogens in a number of proteins can become exposed in concerted unit steps. Figure 1(a) illustrates the five concerted folding–unfolding units, called foldons, found in cytochrome *c* (Cyt *c*). Figure 1(b) illustrates the NHX method. The experiment identifies the foldon units in terms of the amino acids involved and can also characterize the energetics, kinetics, and other physical parameters of the high energy partially unfolded forms (PUFs) that are produced by the dynamic reversible unfolding of one or more foldon units.

The physical parameters found for Cyt *c* suggest that the high energy space between the native and fully unfolded states is dominated by PUFs formed in a stepwise pathway by the sequential reversible unfolding of the foldon units, as in reaction Scheme 1 (letters refer to the color-coded foldons in Figure 1(a)):

$$\text{NRYGB(Native)} \leftrightarrow \text{nRYGB} \leftrightarrow \text{nrYGB} \leftrightarrow \text{nryGB} \leftrightarrow \text{nrygB} \leftrightarrow \text{nrygb(Unfolded)} \quad (1)$$

If Scheme 1 represents the major pathway for reversible Cyt *c* unfolding, then the same sequence of steps in the reverse direction must specify the major folding pathway. This is so because the NHX experiments are done under equilibrium native conditions. To maintain equilibrium, each individual unfolding reaction in Scheme 1 must be matched by an equal and opposite folding reaction.

NHX results like those in Figure 1(b) identify the amides newly exposed in each higher energy PUF and the amides that are still protected (not yet exchanged) but they do not reveal the condition of the lower lying units (exchanged earlier). This information is necessary to tell whether the foldon units actually unfold and refold in a sequential pathway manner, as in Figure 1(c) where each higher energy PUF includes all of the prior unfoldings, or whether they unfold independently as in Figure 1(d), or in some mixed mode.

A stability labeling method can make this distinction.[13] The present work attempted this approach. The solvent-exposed Glu62 residue in the Yellow unit of Cyt *c* was mutated to glycine. This change deletes an ion pair bond and destabilizes the protein by 0.8 kcal/mol. In a sequential pathway situation (Figure 1(c)), one can expect that the Yellow unit itself and all of the higher lying states (Green foldon open; Blue foldon open) will be destabilized more or less equally while lower lying states (Red open; nested-Yellow open) are unaffected. Independent unfolding would produce a quite different result (Figure 1(d)).

Results

Global protein stability

We used a recombinant pWT Cyt *c*.[14] Structure in the pWT protein is essentially identical to the highly studied parent wild-type equine Cyt *c* as indicated by the chemical shifts of amide NH and $C^{\alpha}H$ protons, which are the same within our digital resolution (± 0.02 ppm) except near the altered positions (N terminus not acetylated; His26 and 33 are Asn). We studied the effect of a destabilizing mutation in which a surface glutamate at position 62 is replaced by glycine (Glu62Gly). Only chemical shifts of amide protons in the immediate vicinity of the mutant residue are affected, suggesting some minor distortion of the local structure.

Denaturant melting experiments show that the Gly62 mutation modestly destabilizes the protein (melting midpoint from 2.85 M GdmCl to 2.72 M; pH 7, 20 °C). The destabilization is due to the loss of an ion pair bond between the side-chains of Glu62 and Lys60 and the entropic cost of the newly introduced glycine. Cyt *c* stability cannot be accurately obtained by the usual denaturant melting analysis because melting is not accurately two-state; a significant population of intermediates occurs in the transition zone. The intermediate population can be predicted by extrapolation into the unfolding transition region of the foldon isotherms shown in Figure 1(b). This non-two-state behavior has been analyzed by Mayne & Englander[15] for WT Cyt *c* and by Eftink & Ionescu[16] in more general terms. The result is an artifactual broadening of the melt, which spuriously reduces the slope (*m* value) and the ΔG_u value estimated by linear extrapolation to zero denaturant. The melting midpoint is less affected, and may be shifted in either direction depending on the intermediate population.

Alternatively, protein stability can be obtained from the HX rates of amide protons that exchange by way of global unfolding[17,18] (equations (3)–(5)). A survey of the published literature[19,20] found that this approach yields results that agree with two-state melting results. In Cyt *c*, several protons in the C-terminal helix act as markers for global unfolding (Blue in Figure 2). The various globally exchanging hydrogens differ somewhat. Similar variability has been computed for the globally exchanging hydrogens in other proteins and is often attributed to

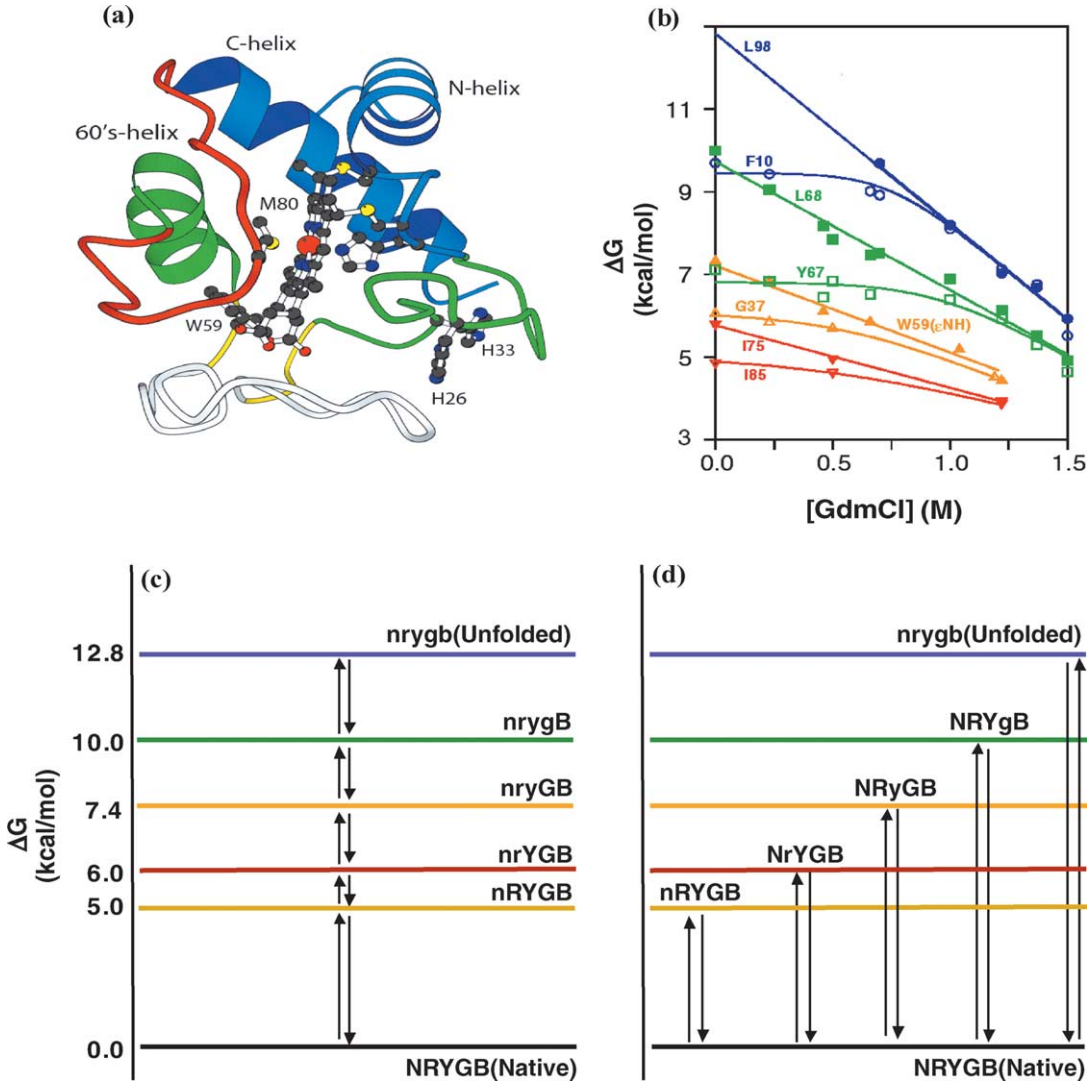

Figure 1. Illustration of the problem addressed here. A ribbon model for Cyt *c* (a) is color-coded to show the different foldon units previously found by NHX and HX pulse labeling experiments. The NHX experiment (b) uses destabilant to selectively promote large unfolding reactions so that they come to dominate the exchange of the hydrogens that they expose, in competition with small local fluctuation reactions. For each partially unfolded form (PUF), one marker proton and one merging proton is shown (data from Bai *et al.*[25]). The question is whether the foldon units unfold and refold in a sequential pathway manner as in (c) and reaction Scheme 1, or independently as in (d), or in some mixed mode.

uncertainties in the value of k_{int}, the intrinsic HX rate calibrated for fully exposed amides in small peptides (equation (3)).[21–23] In fact, the uncertainty in those calibrations is exceptionally small and seems unlikely to account for the spread in values for ΔG_{HX}. The variability is impressively consistent (Table 1) and seems to be due to real effects that are not yet understood, including especially the condition of the different residues in the unfolded state.[24]

Comparison of the global marker protons listed in Table 1 indicates a global stability for pWT Cyt *c* and the Glu62Gly mutant protein of 12.5 kcal/mol and 11.6 kcal/mol, respectively. Equivalent values measured correctly by denaturant melting will be slightly lower due to the time-dependent relaxation of proline stereoisomers.[17]

Subglobal PUF stability

Just as marker protons in the C-terminal helix exchange by way of global unfolding, Cyt *c* happens also to provide marker protons with HX rate controlled by the dynamic subglobal unfolding of its other lower lying foldons. Marker protons can be distinguished by their sharp dependence on denaturant (large unfolding reaction),[25] temperature (large ΔH_u),[26] pressure (large ΔV_u),[27] and/or pH.[28] Many other hydrogen atoms exchange by way of local fluctuational pathways, minimally distorted near-native states that have little dependence on destabilizing factors. Small local fluctuations that break one protecting hydrogen bond at a time are interesting in respect to issues of protein flexibility and dynamics[29] but they do not provide reasonable

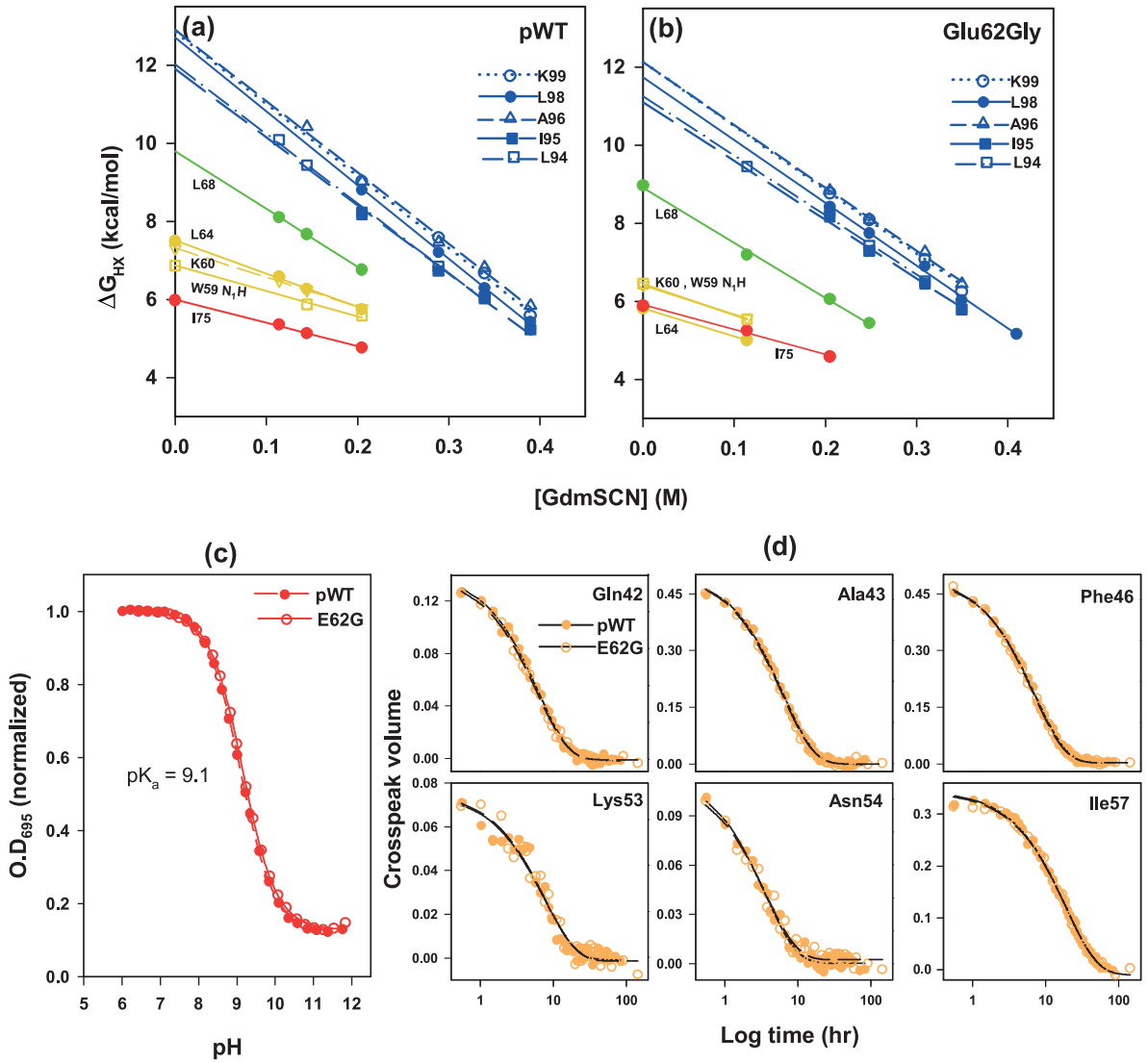

Figure 2. The data obtained. (a) and (b) Comparison of NHX data for marker protons in the various foldons (pH* 7, 20 °C) of pWT Cyt *c* and the Glu62Gly variant (see Table 1). (c) Comparison of the alkaline transition in both proteins, which is sensitive to the stability of the Red loop (pK_a=9.1). (d) HX data for residues in the nested-Yellow loop under conditions where exchange is due to the concerted loop unfolding (pH* 3.3, reduced Cyt *c*, no denaturant; single exponential fit). Results are summarized in Figure 3 and Table 1.

candidates for folding intermediates. Worse, they obscure the very interesting larger unfolding reactions.

The NHX experiment takes advantage of the special sensitivity of large unfoldings to destabilizing factors. Large unfoldings are sharply promoted by destabilants and may then come to dominate the exchange of the hydrogens that they expose. Under favorable conditions, the exchange of the measurable amide hydrogens in each foldon is then seen to merge into a common HX isotherm, as illustrated in Figure 1(b). This behavior can reveal the identity of individual foldons (amides exposed), their stability against unfolding (from HX rate in EX2 conditions), their folding and unfolding rates (from HX rate in EX1 conditions), and the sensitivity parameter of each unfolded state pertinent to the destabilant used (m, ΔH_u, ΔV_u, $\Delta \nu_u$ for denaturant, temperature, pressure, or pH, respectively).

In an effort to distinguish between the sequential and independent unfolding models (Figure 1(c) *versus* (d)) we did stability labeling experiments using the marker protons themselves as convenient probes for the effect on each foldon of the destabilizing Glu62Gly mutant.

Higher lying states: the Blue and Green foldons

Some protons in the C-terminal helix exchange by way of the transient global unfolding, marked by unfolding of the highest energy (Blue) foldon (Figure 2(a) and (b)). When taken individually, the global marker protons L94, I95, A96, L98, and I99 indicate a destabilization ($\Delta\Delta G_{HX}$) due to the Glu62Gly mutation of 0.8, 0.8, 0.8, 1.0, and 0.8 kcal/mol, respectively (Table 1), yielding an average of 0.84 kcal/mol for the Blue unfolding.

Leu68 is the sole marker proton for the 60's helix

Table 1. Parameters of pWT and Glu62Gly Cyt *c* from native state hydrogen exchange

		pWT		Glu62Gly		
Foldons	Residue	ΔG_{HX} (kcal/mol)	*m*-value (kcal mol^{-1} M^{-1})	ΔG_{HX} (kcal/mol)	*m*-value (kcal mol^{-1} M^{-1})	$\Delta\Delta G_{HX}$ (kcal/mol)
Blue	Leu94	12.0	17.8	11.2	15.2	−0.8
	Ile95	11.9	17.5	11.1	15.1	−0.8
	Ala96	12.9	18.3	12.1	16.1	−0.8
	Leu98	12.7	18.9	11.7	16.1	−1.0
	Lys99	12.9	18.4	12.1	16.3	−0.8
Green	Leu68	9.8	14.8	8.9	14.1	−0.9
Yellow	W59 N_1H	6.9	6.5	6.5	7.9	−0.4
	Lys60	7.3	7.5	6.4	7.5	−0.9
	Leu64	7.5	8.5	5.8	7.2	−1.7
Red	Tyr74	6.1	5.5	6.0	5.5	−0.1
	Ile75	6.0	6.0	5.9	6.3	−0.1
Nested Yellow	Thr40	2.23		2.20		−0.03
	Gln42	2.20		2.19		−0.01
	Ala43	2.21		2.18		−0.03
	Phe46	2.05		2.02		−0.03
	Lys53	2.36		2.36		0.00
	Asn54	2.29		2.29		0.00
	Lys55	2.32		2.34		0.02
	Ile57	2.33		2.34		0.01

Data were obtained by native state HX measurements as a function of GdmSCN concentration at pH* 7, 20 °C, in oxidized Cyt *c* (Figure 2(a) and (b)). The nested-Yellow hydrogens were measured at pH* 3.3 in reduced Cyt *c* (Figure 2(d)).

unfolding (Green foldon). It indicates a $\Delta\Delta G_{HX}$ value for destabilization of the Green unit due to the mutation of 0.9 kcal/mol (Figure 2(a) and (b); Table 1).

The Yellow unit

The Yellow unit is a short irregular structure that necks together the two ends of the nested-Yellow loop and connects them to the Green helix on one end and the Green loop on the other (Figure 1(a)). Previous work with oxidized WT Cyt *c* found that the indole NH of Trp59 and the amide hydrogen atoms of residues 60 and 64 all behaved identically as markers for the Yellow unfolding. The Leu64-NH is the first H-bonded amide NH in the 60's helix (part of the Green foldon) but it is exposed to exchange when the Yellow unit unfolds because its H-bond acceptor is the Lys60-C=O, in the Yellow unit (the Yellow neck to Green helix boundary seems to be at Glu61-C^{α} in oxidized WT Cyt *c*[27]).

The Glu62Gly mutation used here is placed between two markers for the Yellow foldon (residues 60 and 64). It directly destabilizes the Yellow unit by removing the Glu62 to Lys60 salt link. We want to use the Yellow marker protons to assess this effect. However, the Yellow neck markers do not always exchange as a clear grouping. They have been seen to separate in reduced WT Cyt *c*[13,26] and they do so also in the present case. The Yellow marker residues 59, 60 and 64 register a destabilization due to the Gly62 mutation of 0.45, 0.90, and 1.65 kcal/mol, respectively, for an average of 1 kcal/mol (Figure 2(a) and (b); Table 1), with considerable uncertainty due to the different behavior of the marker protons in the different proteins.

The cause for the break up of the Yellow markers is unclear. The Gly62 mutation produces changes in chemical shift of its immediate neighbors (amide protons of residues 61 and 63 change by +0.1 ppm and −0.1 ppm, respectively). The HX changes may receive contributions from the mutation-dependent changes in local structure indicated by chemical shift, and/or due to the promotion of local fluctuation reactions by the glycine mutation, as was seen before for the Lys8Gly mutation in the N-terminal helix.[29] In addition, the effects that cause spreading in other markers (Figure 2(a) and (b)) may play a role.

Whatever the source it is clear that the Yellow unit is destabilized by the mutation, like the foldons above it, and by about the same amount.

Lower lying states: the Red loop

The HX marker for the Red loop, Ile75, indicates no effect due to the Gly62 mutation (−0.01 kcal/mol measured; Figure 2(a) and (b); Table 1).

Stability information for the Red loop could also be obtained by comparing the Cyt *c* alkaline transition in the pWT and Glu62Gly proteins. Ferri Cyt *c* undergoes a reversible pH-dependent conformational transition at high pH in which the normal Met80-S to heme ligand is replaced by a nearby lysine, probably Lys79 but possibly Lys72 or 73.[30,31] All of these residues are in the Red loop. Previous information shows that the replacement reaction, indicated by the pK_a of the transition, depends on the stability of the Red and nested-Yellow loops and their unfolding rates.[32] Figure 2(c) shows that the alkaline transitions of pWT Cyt *c* and the Glu62Gly variant are identical, consistent with the HX result that the Red loop is not destabilized by the mutation.

Lower lying states: the nested-Yellow loop

The grouped unfolding behavior shown by the Yellow marker protons in oxidized WT Cyt *c* was initially taken to represent the entire bottom Ω-loop (Figure 1(a)).[27] Subsequent work found that the Ω-loop segment unfolds independently and at lower free energy than the short Yellow neck region.[28] We therefore consider the nested Ω-loop segment as a separate concerted foldon unit, referred to as the nested-Yellow loop.

Amide protons in this segment exchange too rapidly to measure under the conditions used for Figure 2(a) and (b) (pH* 7, 20 °C). Therefore, a different strategy is required in order to study the effect of the Gly62 mutation. Previous work showed that these hydrogen atoms can be measured at lower pH where HX is slowed.[28] This was done in reduced Cyt *c*, which has enough stability at low pH to make the measurement possible. Fortuitously, low pH also selectively decreases the stability of the concerted nested-Yellow loop unfolding causing its hydrogens to merge to a common ΔG_{HX}, controlled by the concerted loop unfolding. The pH sensitivity of this unfolding reaction is due to the destabilization of buried protonatable bonding groups that help to stabilize the nested-Yellow loop.[28]

To compare the nested-Yellow protons in the pWT and Glu62Gly proteins, HX was measured at the low pH condition (pH* 3.3, reduced Cyt *c*). The data (Figure 2(d)) confirm the cooperative foldon nature of the nested-Yellow loop; ΔG_{HX} values for the accurately measurable residues all merge to within 0.3 kcal/mol (Table 1). The change in ΔG_{HX} value due to the Glu62Gly mutation is <0.04 kcal/mol (Table 1).

Summary

Figure 3(a) shows the comparisons obtained here. In the mutant Glu62Gly protein, the Yellow and higher lying units are all destabilized by about 0.9 kcal/mol. The lower lying Red and nested-Yellow units are not affected. This result supports a sequential unfolding pathway.

Figure 3(b) summarizes results from a prior stability labeling experiment in which energy levels of the various foldons were compared in reduced and oxidized Cyt *c*.[13] Reduction of the heme iron selectively increases the stability of the Met80-S to heme iron bond by 3.2 kcal/mol (independently measured). As expected, the Red unit was found to be stabilized by 3.2 kcal/mol because it contains the Met80-S heme ligand and the ligation is broken when the Red unit unfolds. The higher lying Yellow and Green units were stabilized by the same amount, indicating that these unfoldings contain the Red unfolding, as expected for sequential unfolding. The highest lying Blue unit, known to reflect global unfolding, was also stabilized in reduced Cyt *c* but by an even larger amount (5.1 kcal/mol = 3.2 + 1.9 kcal/mol). One can rationalize this result; reduced Cyt *c* is more stable not only because of its stronger Fe–S bond but also because oxidized Cyt *c* carries a buried charge, which apparently becomes exposed and therefore effective only in the final Blue unfolding.[13,33] Still, the anomaly creates some uncertainty. Also, this study was done before the lowest lying nested-Yellow unit was known. The present study does not suffer from these uncertainties.

In summary, the stability labeling results now available show that the folding–unfolding behavior of the concerted foldon units that comprise the Cyt *c* protein is as expected for a sequential unfolding pathway (Figure 1(c)) but not for independent unfolding (Figure 1(d)).

Discussion

The discovery of the foldon substructure of Cyt *c* and other proteins reveals a previously unsuspected dimension of protein structural organization and properties. It seems likely that foldon behavior will be found to participate in various aspects of protein function.[28,32] We are here particularly interested in how foldons may participate in the protein folding process.

Foldons and folding in Cyt *c*

The present work adds to a body of information on the equilibrium foldon substructure and folding properties of Cyt *c*. The existence and identity of the Cyt *c* foldons was revealed in the first equilibrium NHX study.[25] In that study, the possibility that the various units unfold in the sequential stepwise manner shown in reaction Scheme 1 was suggested by the fact that their unfolding free energy and their new surface exposure (denaturant dependence) increase in the same rank order (Figures 1(b) and 2(a) and (b); Table 1). The sequential unfolding picture (Figure 1(c)) has now been validated in the present mutation-based stability labeling study (Figure 3(a)) and in a previous comparison between reduced and oxidized Cyt *c* (Figure 3(b)). These results, found at equilibrium native conditions, require an equivalent folding sequence in the reverse direction.

Kinetic methods consistently validate the same picture. According to Scheme 1, the Blue unit should be the first to fold. A recent HX pulse labeling study[34] confirmed that all of the amino acids in the Blue foldon are protected in an initial kinetic folding intermediate while the other foldons remain unprotected. Other studies indicate that this structure is on the folding pathway[35] and in fact is formed in the initial rate-limiting transition state.[36] Further, when the N and C segments are prepared in isolation and then mixed, they form some residual helical structure,[37] and especially so when they are covalently tethered.[38]

Another approach, called kinetic native state HX, used a high pH labeling pulse applied to the native protein to study the kinetic pathway in the

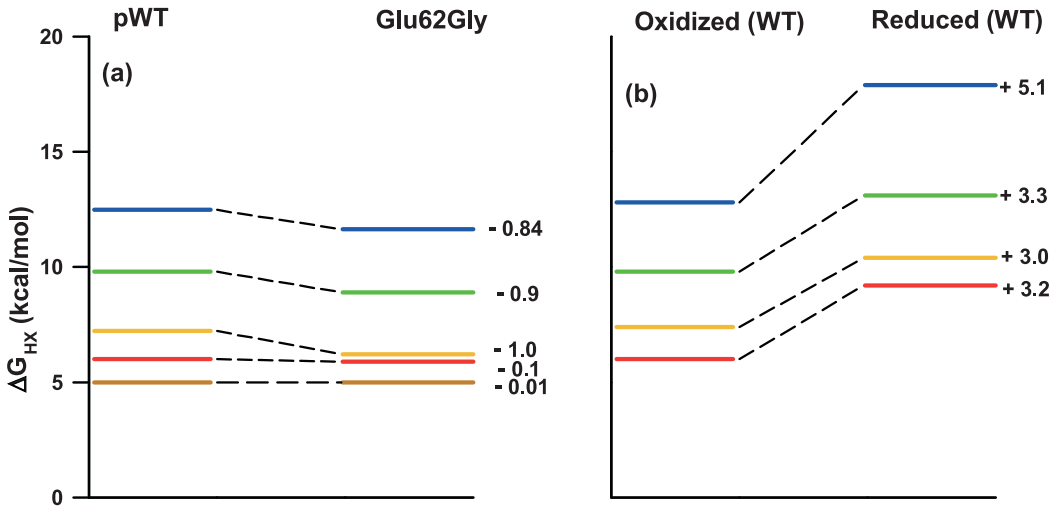

Figure 3. Summary of stability labeling results. The panels compare pWT Cyt *c* with the Glu62Gly mutant ((a), present results) and oxidized with reduced WT Cyt *c* ((b), data from Xu *et al.*[13]). For each foldon, the average value obtained from the several marker protons (Table 1) is shown. Results for the nested-Yellow unit are from the data in Figure 2(d); the ΔG_{HX} level is from minimal data for the slowest exchanging residue (Gln42) in the loop under the same conditions used for the other foldon measurements (pH* 7, 20 °C). In both cases the modification-induced change in foldon stability appears in the affected foldon and higher lying states but not in lower lying ones, as expected for sequential unfolding (Figure 1(c)) but not for independent unfolding or possible mixed modes.

unfolding direction.[39] The first unit seen to deprotect was the Red loop (the nested-Yellow loop was not yet known). The Red loop unfolding rate was identical to that indicated by stopped-flow unfolding experiments. These results confirm the unit nature of the Red loop and its role as an early unfolding step, in agreement with Scheme 1. The same work showed that the subsequent unfolding sequence occurs in the order shown in Scheme 1. Thus kinetic studies find the same units placed in the same order as the equilibrium studies.

Finally, it seems compelling that the order of folding steps indicated by all of this work is consistent with a "sequential stabilization" sequence that is dictated by the native structure.[40] In the native protein, the two interacting N and C helices form a docking surface that contacts only the Green helix and Green loop. If the pathway proceeds in the native context, then the initial formation of the Blue unit can act to guide and stabilize only the Green units. In turn, the Yellow neck region grows out of the two Green segments. Finally the Green and Yellow units are necessary to guide and stabilize the formation of the Red and nested-Yellow Ω-loops.

In summary, results obtained at multiple site-resolved positions in Cyt *c* in both equilibrium and kinetic experiments support the unit foldon nature of Cyt *c*. The same body of evidence shows that the foldon substructure of Cyt *c* determines a linear stepwise folding–unfolding pathway. The pathway sequence apparently depends on the foldon organization in the native protein together with the principle of sequential stabilization. Different topologies in other proteins may dictate a different pathway sequence, perhaps not simply linear.

Foldons and folding in other proteins

How general are the Cyt *c* results? Analogous results for other proteins are available.

Apomyoglobin (apoMb) has eight helical segments (A to H). An early HX study of the pH 4 molten globule found helices A, G, and H formed.[41] Independently, HX pulse labeling found that the A, G, and H helices form initially[42] and further stabilization induces helix B formation.[43] The same four helices and the beginnings of some others were found by direct NMR study of the pH 4 molten globule.[44] An experimental test showed that these intermediates are on the folding pathway.[45] More recent HX pulse labeling studies show that entire helices can be added to or removed from the kinetic intermediate by mutational manipulation of their relative stability.[46,47] These results connect the foldon substructure of apoMb to entire helical units and groupings thereof and indicate their role as intermediates in folding.

Marqusee and co-workers have extensively analyzed the folding behavior of ribonuclease H_1, which consists of five helices (A to E) and five β-strands (I to V). NHX results[48] found two subglobal PUFs. In the lowest energy PUF observed, strands I, II, III, V, and helix E are unfolded. The next higher energy PUF unfolds also helix B and strand IV, leaving helix A and D as the final unfolding unit. A similar ladder with slightly different groupings but the same order was found in a thermophilic homolog.[49] In agreement, kinetic HX pulse labeling, CD, and mutational results connect the earliest folding phase to formation of helix A and D and strand IV, the second phase largely to β-sheet formation, and show that

these intermediates are on the folding pathway.[50,51] An HX protection study shows that the equilibrium acid molten globule has only helices A, B and D well formed, similar to the kinetic and NHX results.[52] Also a large synthetic fragment containing the contiguous helices A to D and strand IV displays independent stability.[53] These results reveal (some of) the foldon units that compose RNase H and their role in determining a stepwise folding pathway.

Fuentes & Wand studied the four-helix bundle protein, apoCyt b562, by native state HX using the destabilants GdmCl[54] and pressure.[27] The results suggest a sequence in which the core bihelix II+III folds first followed, perhaps in random order, by helix I and IV. Bai and co-workers studied a stabilized apob562 multiple mutant[55] by NHX, kinetic folding, and phi analysis. All of the results consistently indicate that the core helices II and III fold first, perhaps in the initial folding transition state, followed by helix IV and then helix I. In further studies, the same group created a construct with helix I mutationally destabilized.[56] Direct NMR studies[57] found a well-folded structure with helix I unfolded and the rest of the structure intact, equivalent to the first PUF in the inferred unfolding sequence. The remaining three helices maintain a nearly native main-chain conformation but the partially exposed apolar side-chains energy minimize by extensive repacking. Phi analysis indicated similar repacking behavior in a kinetic intermediate of the similarly structured Im7 protein.[58] These results reveal a foldon-dependent folding pathway. They also indicate the important result that incompletely native forms (intermediates, transition states) are likely to include non-native interactions.

Koide and co-workers used native state HX to dissect the OspA protein into five distinct foldon units.[59] An extensive mutational phi analysis study found that at least two PUFs consisting of these same foldons occur as sequential intermediates in the kinetic folding pathway.[60]

The homodimer, triose phosphate isomerase (TIM), which is too large for the NMR analysis usually used to analyze HX results, was studied by Silverman & Harbury.[61] They measured the side-chain reactivity of Cys residues inserted at various buried positions. Similar methodologies have been explored by others.[62,63] The side-chain reactivity results for TIM exhibit behavior analogous to NHX results (like Figure 1(b)), with local fluctuational pathways merging into larger unfolding reactions that define two distinct subglobal PUFs. The analog of a stability labeling experiment showed that these forms fold and unfold in a sequential pathway manner.

In summary, detailed site-resolved results now available for a number of proteins consistently point to component structural folding units and their formation in folding pathways that step through distinct intermediate forms, just as found for Cyt *c*. Given the difficulties that conspire to obscure protein foldons, it seems likely that foldon substructure is far more prevalent than has been found so far.

Summary

Detailed site-resolved observations for Cyt *c* and a number of other proteins support the conclusion that protein molecules are made up of separately cooperative folding–unfolding units. This relatively unexplored dimension of protein structure seems likely to have important implications for a variety of protein properties including cooperativity, stability, design, evolution, and function.

The present work focuses on the role of foldon units in protein folding. Results now available support the following conclusions. (1) Proteins fold by stepping energetically downhill through discrete partially folded intermediate forms, progressively adding native-like foldon units to construct determinate folding pathways. (2) The sequence of steps proceeds in the native context, determined by a sequential stabilization process that is directed by the same general interactions that determine the final native structure.

The reality of foldon units and the principle of sequential stabilization provide a physical rationale for stepwise sequential folding pathways.

Materials and Methods

Materials

^2H_2O (99%), deuterated sodium formate, guanidinium hydrochloride (GdmCl) and guanidinium thiocyanate (GdmSCN) were from Isotec, ICN, and Fisher Scientific. Deuterated GdmSCN was prepared by dissolving in ^2H_2O and lyophilizing three times. All experiments were performed in phosphate buffer at pH 7 and 20 °C. Solutions contained high salt, equivalent to 0.5 M ionic strength, with denaturant as indicated. Recombinant pseudo wild-type Cyt *c* (pWT) and the Glu62Gly mutant were expressed in a high yield *Escherichia coli* system and purified as described elsewhere.[14] Protein concentration was measured using an extinction coefficient of 106 $mM^{-1}\,cm^{-1}$ at 409 nm for oxidized Cyt *c*. Refractive index measurements were used to determine GdmCl concentration.[64] GdmSCN concentration used equation (2), where C is the molar concentration and ΔN is the difference between the refractive index of the denaturant solution and aqueous buffer at the sodium D line:

$$C = 42.0(\Delta N) - 114(\Delta N)^2 + 870(\Delta N)^3 \qquad (2)$$

Stability

Circular dichroism (222 nm) and fluorescence data (280 nm excitation, 320 nm cut-off filter for emission) for GdmCl-induced unfolding were simultaneously recorded in an Aviv 202 instrument (20 °C). Protein concentration was about 7 μM. Melting curves were fit by a two-state equation.[65]

The Cyt *c* alkaline transition was measured by

absorbance at 695 nm, which detects the heme–Met80 ligation. Protein concentration was about 250 μM in 50 mM phosphate, 0.5 M KCl, at 20 °C. Samples were pH-adjusted with dilute KOH and HCl, equilibrated for five minutes before measurement, and each data point was averaged for 45 seconds. The data were corrected for dilution and fit with a one-proton titration curve.

Hydrogen exchange

Hydrogen–deuterium exchange was monitored by the time-dependent decrease in amide cross-peak volumes in two-dimensional homonuclear correlated ^{1}H NMR spectra (pulsed field gradient COSY with two transients per increment) using a Varian 500 MHz Inova spectrometer (1024 complex data points, 512 time increments, 20 ppm spectral width in both dimensions, low power presaturation to suppress the residual HO^{2}H signal; Felix package on a Silicon Graphics workstation). Cross-peak assignments for oxidized Cyt *c* were from Feng *et al*.[66] Appropriately scaled baseline footprints were subtracted from cross-peak volumes normalized to the Tyr97 cross-peak. Data analysis and display were performed automatically with home written software.

Exchange was initiated by moving protein into ^2H_2O exchange buffer by centrifugal gel filtration.[67] Dead time of the experiment was less than 25 minutes. For oxidized Cyt *c*, HX was done in 0.1 M phosphate (pH* 7) in ^2H_2O, at various concentrations of GdmSCN (GdmSCN + KCl = 0.5 M). HX of protons in the fast exchanging nested-Yellow foldon was measured in reduced Cyt *c* at pH* 3.3 (0.1 M deuterated formate, 40 mM ascorbate) to slow exchange.

HX data analysis

The chemical exchange rate (k_{ch}) of a freely exposed amide hydrogen at the neutral pH used here can be described by equation (3), where the intrinsic rate constant (k_{int}) depends on pH, temperature, isotope effects, and nearest neighbor inductive and blocking effects,[21–23] calculated using the spread sheet available at HX2.Med.UPenn.edu/download.HTML:

$$k_{ch} = k_{int}[OH^-] \quad (3)$$

For stably protected hydrogen atoms ($k_{cl} \gg k_{op}$), exchange rate in the steady state approximation is given by equation (4),[68] where k_{op} and k_{cl} are opening and reclosing rates of the protecting structure. The approximation shown holds when reclosing is faster than chemical exchange (the EX2 bimolecular exchange condition):

$$k_{ex} = (k_{op}k_{ch})/(k_{cl} + k_{ch}) \approx K_{op}k_{ch} \quad (4)$$

Measured exchange rate (k_{ex}) together with the known k_{ch} (equation (3)) then leads to K_{op} and hence ΔG°_{HX} as in equation (5):

$$\Delta G^\circ_{HX} = -RT \ln K_{op} \quad (5)$$

When HX is determined by a large unfolding reaction, like the unfolding of one of the foldon units in Cyt *c*, this can be demonstrated by the large sensitivity of exchange rate to increasing denaturant concentration and also by the fact that all of the protected hydrogens in that unit merge into the same HX isotherm (plot of ΔG_{HX} *versus* denaturant). We determined the unfolding free energy of each foldon at zero denaturant by measuring the denaturant dependence of its marker proton(s) and extrapolating the measured data to zero denaturant.

Acknowledgements

Supported by NIH research grant NIH GM31847. We thank I. Tinoco and members of this laboratory for help and comments on the manuscript.

References

1. Levinthal, C. (1969). How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems* (Monticello, I. L., ed.), pp. 22–24, University of Illinois Press, Allerton House.
2. Creighton, T. E. (1984). Pathways and mechanisms of protein folding. *Advan. Biophys.* **18**, 1–20.
3. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631–660.
4. Ptitsyn, O. B. (1995). Structures of folding intermediates. *Curr. Opin. Struct. Biol.* **5**, 74–78.
5. Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proc. Natl Acad. Sci. USA*, **89**, 20–22.
6. Baldwin, R. L. (1995). The nature of protein folding pathways: the classical *versus* the new view. *J. Biomol. NMR*, **5**, 103–109.
7. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science*, **267**, 1619–1620.
8. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet.* **21**, 167–195.
9. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.
10. Onuchic, J. N., Nymeyer, H., Garcia, A. E., Chahine, J. & Socci, N. D. (2000). The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Advan. Protein Chem.* **53**, 87–152.
11. Mirny, L. & Shakhnovich, E. (2001). Protein folding theory: from lattice to all-atom models. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 361–396.
12. Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75.
13. Xu, Y., Mayne, L. & Englander, S. W. (1998). Evidence for an unfolding and refolding pathway in cytochrome *c*. *Nature Struct. Biol.* **5**, 774–778.
14. Rumbley, J. N., Hoang, L. & Englander, S. W. (2002). Recombinant equine cytochrome c in *Escherichia coli*: high-level expression, characterization, and folding and assembly mutants. *Biochemistry*, **41**, 13894–13901.
15. Mayne, L. & Englander, S. W. (2000). Two-state *vs.* multistate protein unfolding studied by optical melting and hydrogen exchange. *Protein Sci.* **9**, 1873–1877.
16. Eftink, M. R. & Ionescu, R. (1997). Thermodynamics of protein unfolding: questions pertinent to testing the validity of the two-state model. *Biophys. Chem.* **64**, 175–197.
17. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins: Struct. Funct. Genet.* **20**, 4–14.
18. Bai, Y., Englander, J. J., Mayne, L., Milne, J. S. &

Englander, S. W. (1995). Thermodynamic parameters from hydrogen exchange measurements. *Methods Enzymol.* **259**, 344–356.
19. Huyghues-Despointes, B. M., Scholtz, J. M. & Pace, C. N. (1999). Protein conformational stabilities can be determined from hydrogen exchange rates. *Nature Struct. Biol.* **6**, 910–912.
20. Huyghues-Despointes, B. M., Pace, C. N., Englander, S. W. & Scholtz, J. M. (2001). Measuring the conformational stability of a protein by hydrogen exchange. *Methods Mol. Biol.* **168**, 69–92.
21. Molday, R. S., Englander, S. W. & Kallen, R. G. (1972). Primary structure effects on peptide group hydrogen exchange. *Biochemistry*, **11**, 150–158.
22. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75–86.
23. Connelly, G. P., Bai, Y., Jeng, M. F. & Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 87–92.
24. Rose, G. D. (2002). Unfolded proteins. In *Advances in Protein Chemistry*, vol. 62, Academic Press, San Diego.
25. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
26. Milne, J. S., Xu, Y., Mayne, L. C. & Englander, S. W. (1999). Experimental study of the protein folding landscape: unfolding reactions in cytochrome *c*. *J. Mol. Biol.* **290**, 811–822.
27. Fuentes, E. J. & Wand, A. J. (1998). Local stability and dynamics of apocytochrome *b*562 examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry*, **37**, 9877–9883.
28. Krishna, M. M. G., Lin, Y., Rumbley, J. N. & Walter Englander, S. (2003). Cooperative omega loops in cytochrome *c*: role in folding and function. *J. Mol. Biol.* **331**, 29–36.
29. Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. (2003). Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci.* **12**, 153–160.
30. Rosell, F. I., Ferrer, J. C. & Mauk, A. G. (1998). Proton-linked protein conformational switching: definition of the alkaline conformational transition of yeast iso-1-ferricytochrome *c*. *J. Am. Chem. Soc.* **120**, 11234–11245.
31. Assfalg, M., Bertini, I., Dolfi, A., Turano, P., Mauk, A. G., Rosell, F. I. & Gray, H. B. (2003). Structural model for an alkaline form of ferricytochrome *c*. *J. Am. Chem. Soc.* **125**, 2913–2922.
32. Hoang, L., Maity, H., Krishna, M. M., Lin, Y. & Englander, S. W. (2003). Folding units govern the cytochrome *c* alkaline transition. *J. Mol. Biol.* **331**, 37–43.
33. Cohen, D. S. & Pielak, G. J. (1995). Entropic stabilization of cytochrome c upon reduction. *J. Am. Chem. Soc.* **117**, 1675–1677.
34. Krishna, M. M. G., Lin, Y., Mayne, L. & Englander, S. W. (2003). Intimate view of a kinetic protein folding intermediate: residue-resolved structure, interactions, stability, folding and unfolding rates, homogeneity. *J. Mol. Biol.* **334**, 501–513.
35. Bai, Y. (1999). Kinetic evidence for an on-pathway intermediate in the folding of cytochrome *c*. *Proc. Natl Acad. Sci. USA*, **96**, 477–480.
36. Krantz, B. A., Srivastava, A. K., Nauli, S., Baker, D., Sauer, R. T. & Sosnick, T. R. (2002). Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nature Struct. Biol.* **9**, 458–463.
37. Wu, L. C., Laub, P. B., Elove, G. A., Carey, J. & Roder, H. (1993). A noncovalent peptide complex as a model for an early folding intermediate of cytochrome *c*. *Biochemistry*, **32**, 10271–10276.
38. Kuroda, Y. (1993). Residual helical structure in the C-terminal fragment of cytochrome *c*. *Biochemistry*, **32**, 1219–1224.
39. Hoang, L., Bédard, S., Krishna, M. M., Lin, Y. & Englander, S. W. (2002). Cytochrome *c* folding pathway: kinetic native-state hydrogen exchange. *Proc. Natl Acad. Sci. USA*, **99**, 12173–12178.
40. Rumbley, J., Hoang, L., Mayne, L. & Englander, S. W. (2001). An amino acid code for protein folding. *Proc. Natl Acad. Sci. USA*, **98**, 105–112.
41. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science*, **249**, 1544–1548.
42. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*, **262**, 892–896.
43. Loh, S. N., Kay, M. S. & Baldwin, R. L. (1995). Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc. Natl Acad. Sci. USA*, **92**, 5446–5450.
44. Eliezer, D., Chung, J., Dyson, H. J. & Wright, P. E. (2000). Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry*, **39**, 2894–2901.
45. Jamin, M. & Baldwin, R. L. (1998). Two forms of the pH 4 folding intermediate of apomyoglobin. *J. Mol. Biol.* **276**, 491–504.
46. Cavagnero, S., Dyson, H. J. & Wright, P. E. (1999). Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J. Mol. Biol.* **285**, 269–282.
47. Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H. J. & Wright, P. E. (2000). Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. *Biochemistry*, **39**, 11227–11237.
48. Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nature Struct. Biol.* **3**, 782–787.
49. Hollien, J. & Marqusee, S. (1999). Structural distribution of stability in a thermophilic enzyme. *Proc. Natl Acad. Sci. USA*, **96**, 13674–13678.
50. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Struct. Biol.* **4**, 298–304.
51. Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nature Struct. Biol.* **6**, 825–831.
52. Dabora, J. M., Pelton, J. G. & Marqusee, S. (1996). Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry*, **35**, 11951–11958.
53. Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Sci.* **8**, 2251–2257.
54. Fuentes, E. J. & Wand, A. J. (1998). Local dynamics and stability of apocytochrome *b*562 examined by hydrogen exchange. *Biochemistry*, **37**, 3687–3698.
55. Chu, R., Pei, W., Takei, J. & Bai, Y. (2002). Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. *Biochemistry*, **41**, 7998–8003.

56. Takei, J., Pei, W., Vu, D. & Bai, Y. (2002). Populating partially unfolded forms by hydrogen exchange-directed protein engineering. *Biochemistry*, **41**, 12308–12312.
57. Feng, H., Takei, J., Lipsitz, R., Tjandra, N. & Bai, Y. (2003). Specific non-native hydrophobic interactions in a hidden folding intermediate: implications for protein folding. *Biochemistry*, **42**, 12461–12465.
58. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216.
59. Yan, S., Kennedy, S. D. & Koide, S. (2002). Thermodynamic and kinetic exploration of the energy landscape of *Borrelia burgdorferi* OspA by native-state hydrogen exchange. *J. Mol. Biol.* **323**, 363–375.
60. Yan, S., Gawlak, G., Smith, J., Silver, L., Koide, A. & Koide, S. (2004). Conformational heterogeneity of an equilibrium folding intermediate quantified and mapped by scanning mutagenesis. *J. Mol. Biol.* **338**, 811–825.
61. Silverman, J. A. & Harbury, P. B. (2002). The equilibrium unfolding pathway of a $(\alpha/\beta)_8$ barrel. *J. Mol. Biol.* **324**, 1031–1040.
62. D'Silva, P. R. & Lala, A. K. (1999). Hydrophobic photolabeling as a new method for structural characterization of molten globule and related protein folding intermediates. *Protein Sci.* **8**, 1099–1103.
63. Feng, Z., Butler, M. C., Alam, S. L. & Loh, S. N. (2001). On the nature of conformational openings: native and unfolded-state hydrogen and thiol-disulfide exchange studies of ferric aquomyoglobin. *J. Mol. Biol.* **314**, 153–166.
64. Nozaki, Y. (1972). The preparation of guanidinium chloride. *Methods Enzymol.* **26**, 43–50.
65. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
66. Feng, Y., Roder, H., Englander, S. W., Wand, A. J. & Di Stefano, D. L. (1989). Proton resonance assignments of horse ferricytochrome *c*. *Biochemistry*, **28**, 195–203.
67. Jeng, M. F. & Englander, S. W. (1991). Stable submolecular folding units in a non-compact form of cytochrome *c*. *J. Mol. Biol.* **221**, 1045–1061.
68. Linderstrøm-Lang, K. U. (1958). Deuterium exchange and protein structure. In *Symposium on Protein Structure* (Neuberger, A., ed.), pp. 23–24, Methuen, London.

Edited by C. R. Matthews

(Received 7 May 2004; received in revised form 15 July 2004; accepted 3 August 2004)