

- These dimensionless values, multiplied by our final estimate of $\alpha = 0.204 \pm 0.013 \text{ year}^{-1}$, provide the actual values ($1/\tau$). The latter $1/\tau$, in the four lowest boxes are augmented to include an ocean sink with global lifetime (global amount of CH_3CCl_3 divided by global rate of removal by ocean) of 85 years [J. H. Butler *et al.*, *J. Geophys. Res.* **96**, 22347 (1991)] distributed in proportion to the ocean area in each box (ignoring any possible minor temperature effects).
18. The deduced OH concentrations ($10^5 \text{ radical cm}^{-3}$) in the eight lower atmospheric boxes are (from south to north and bottom to top) 5.5, 15.9, 12.2, 4.3, 6.4, 15.8, 13.7, and 5.7. See W. B. DeMore *et al.*, *NASA Jet Propulsion Laboratory Publ.* 94-26 (1994) for rate constants used here.
 19. Going from south to north and bottom to top, the appropriate average temperatures T_i in the eight lower atmospheric boxes are 278, 293, 293, 278, 241, 253, 253, and 241 K, respectively. A simplified version of Eq. 5 regards the whole troposphere as a single box, in which case the appropriate average tropospheric temperature is 277 K [M. J. Prather and C. M. Spivakovsky, *J. Geophys. Res.* **95**, 18723 (1990)] and, from our results, the lower atmospheric CH_3CCl_3 lifetime to use is $\alpha^{-1} = 4.9 \pm 0.3$ years.
 20. The model of Golombek and Prinn (25) has total atmospheric lifetimes resulting from upper atmospheric destruction alone of 112 and 131 years for CH_4 and CHClF_2 , respectively. Total lifetime uncertainty includes stratospheric destruction uncertainty for CH_3CCl_3 (17), but not for CH_4 and CHClF_2 .
 21. E. J. Dlugokencky *et al.*, *Geophys. Res. Lett.* **21**, 45 (1994); P. C. Novelli, K. A. Masarie, P. P. Tans, P. M. Lang, *Science* **263**, 1587 (1994).
 22. S. Madronich and C. Granier, *Geophys. Res. Lett.* **19**, 465 (1992).
 23. D. M. Cunnold *et al.*, *J. Geophys. Res.* **99**, 1107 (1994).
 24. P. M. Midgley, personal communication (letter dated 6 July 1994).
 25. A. Golombek and R. G. Prinn, *J. Geophys. Res.* **91**, 3985 (1986).
 26. This research was supported by NASA grants

NAGW-732 and NAGW-474 (R.G.P. and J.H.), NAGW-732 (F.N.A., D.M.C., D.E.H.), and NAGW-2034 (R.F.W. and B.R.M.); National Oceanic and Atmospheric Administration (NOAA) contract NA85-RAC05103 (P.G.S.); Chemical Manufacturers Association contracts FC-85-567 (P.G.S.) and FC-87-640 (P.J.F.); United Kingdom Department of Environment contract PECD 7/10/54 (P.G.S.); Commonwealth Scientific and Industrial Research Organization (P.J.F.); and Bureau of Meteorology, Melbourne, Victoria, Australia (P.J.F.). Valuable local support at the ALE-GAGE stations was provided by D. Brown (Mace Head, Ireland), P. Sealey (Ragged Point, Barbados), C. Brunson and M. Winey (Cape Matatula, Samoa), and L. Porter (Cape Grim, Tasmania). We thank R. A. Rasmussen and A. J. Crawford for their contributions to the ALE-GAGE experiment, including operation of the Cape Meares, Oregon, and Cape Matatula, Samoa, stations before 1989. Finally, we thank M. Ko, N. D. Sze, D. Weisenstein, C. H. Jackman, and A. R. Douglass for kindly providing results from their two-dimensional models used in our analysis.

RESEARCH ARTICLE

Protein Folding Intermediates: Native-State Hydrogen Exchange

Yawen Bai, Tobin R. Sosnick, Leland Mayne,
S. Walter Englander*

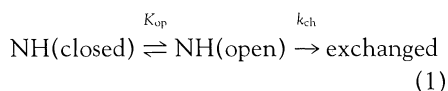
The hydrogen exchange behavior of native cytochrome c in low concentrations of denaturant reveals a sequence of metastable, partially unfolded forms that occupy free energy levels reaching up to the fully unfolded state. The step from one form to another is accomplished by the unfolding of one or more cooperative units of structure. The cooperative units are entire omega loops or mutually stabilizing pairs of whole helices and loops. The partially unfolded forms detected by hydrogen exchange appear to represent the major intermediates in the reversible, dynamic unfolding reactions that occur even at native conditions and thus may define the major pathway for cytochrome c folding.

Under native conditions, a small fraction of any population of protein molecules occupies each possible higher energy, partially unfolded state, including even the fully unfolded state, as described by the Boltzmann distribution. The study of these partially unfolded forms (intermediates) may illuminate the fundamental cooperative nature of protein structure and define the unfolding and refolding pathways of a protein even though the intermediates are normally invisible to measurement. The energy levels and therefore the occupation of these conformationally excited states can be manipulated by denaturants and temperature. Hydrogen exchange experiments can then determine the hydrogens exposed in each higher energy form, their rates of exchange

with solvent, and their sensitivity to the perturbant. From this we can infer, respectively, the structure, the free energy, and the surface exposure of each protein form.

Results for cytochrome c reveal a small sequence of distinct partially unfolded forms with progressively increasing free energy and degree of unfolding. These appear to represent the major intermediates in the unfolding and refolding pathways of cytochrome c.

Hydrogen exchange theory. Exchangeable amide hydrogens (NH) that are involved in hydrogen-bonded structure can exchange with solvent hydrogens only when they are transiently exposed to solvent in some kind of closed to open reaction (1–3), as indicated in Eq. 1.



In the almost universally observed lim-

iting case, referred to as EX2 (for bimolecular exchange) (1), the structural opening reaction enters the rate expression as a pre-equilibrium step. The exchange rate of any hydrogen, k_{ex} , is then determined by its chemical exchange rate in the open form, k_{ch} , multiplied by the equilibrium opening constant, K_{op} (Eq. 2).

$$k_{\text{ex}} = K_{\text{op}} k_{\text{ch}} \quad (2)$$

Since the free peptide rate, k_{ch} , is known from model studies (4–6), the measurement of k_{ex} leads to K_{op} (Eq. 2) and the free energy for the dominant opening reaction, ΔG_{op} (Eq. 3).

$$\Delta G_{\text{op}} = -RT \ln K_{\text{op}} = -RT \ln (k_{\text{ex}}/k_{\text{ch}}) \quad (3)$$

where R is the gas constant and T is the absolute temperature. Any structurally blocked hydrogen may be exposed to exchange by many different unfolding reactions with differing K_{op} values. The opening with the greatest K_{op} (smallest ΔG_{op}) will dominate the hydrogen exchange behavior.

The hydrogen exchange behavior measured here is in the EX2 limit. This is shown by the agreement between the ΔG_{op} values calculated for various hydrogens with disparate k_{ch} values (Figs. 1 to 5). In the alternative EX1 case (1) (monomolecular exchange) the exchange rate is independent of k_{ch} , and therefore correcting the different hydrogens for their k_{ch} differences would make ΔG_{op} values controlled by the same opening appear to be disparate.

Opening reactions that may determine protein hydrogen exchange rates have been discussed in terms of the breakage of single hydrogen bonds (7), the concerted local unfolding of small protein segments (3, 8), and whole molecule unfolding (9–11). Here we deal with a new class of large but still subglobal unfolding reactions. All these cases, although different structurally, are governed by the same mathematical relations (Eqs. 2 and 3).

The authors are at The Johnson Research Foundation, Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104–6059, USA.

*To whom correspondence should be addressed.

Global, local, and subglobal unfolding. The Gibbs free energy for whole molecule unfolding (ΔG_u) is conventionally obtained by carrying a protein through its cooperative unfolding transition, either at high temperature or high denaturant concentration. In the transition zone, the unfolding equilibrium constant, K_u , can be measured and converted into ΔG_u . The dependence of unfolding free energy on denaturant concentration [$\Delta G_u(\text{den})$] is often assumed (12–14) to be linear (Eq. 4).

$$\Delta G_u(\text{den}) = \Delta G_u(0) - m[\text{den}] \quad (4)$$

The slope (m) depends on the additional denaturant sensitive surface exposed in the unfolding reaction. The opening reactions that determine protein hydrogen exchange may expose much or little new surface. They will be promoted accordingly (Eq. 4).

The ΔG_{op} values for the most slowly exchanging hydrogens in native cytochrome c can be computed (Eq. 3) from the data of Bai

et al. (11). These hydrogens, exemplified by Leu⁹⁸ (Fig. 1A), show a strong dependence on guanidinium chloride (GdmCl) concentration, reflecting a very large unfolding reaction (Eq. 4). For these hydrogens the curve for ΔG_{op} against GdmCl concentration extrapolates to the value for global ΔG_u at zero GdmCl concentration and, at high GdmCl concentration, merges smoothly with ΔG_u values obtained from equilibrium melting data (11). Results for ribonuclease A (15, 16) show similar agreement (11). Thus the exchange of the slowest hydrogens in these proteins is dominated by and can measure the parameters of the global unfolding equilibrium (Eqs. 1 and 2), even at solution conditions far from the melting transition.

At low GdmCl concentration, the exchange of many other hydrogens (Fig. 1A) is dominated by various local structural fluctuations, such as $K_{\text{op}}(\text{local})$, that expose little new GdmCl binding surface and therefore have m values (Eq. 3) close to

zero. The same hydrogens can also exchange through the global unfolding reaction, $K_{\text{op}}(\text{g})$, as in Eq. 5.

$$k_{\text{ex}} = [K_{\text{op}}(\text{local}) + K_{\text{op}}(\text{global})]k_{\text{ch}} \quad (5)$$

The globally unfolded state is normally populated only at a very low level and makes no significant contribution to the measured exchange of most hydrogens. As GdmCl concentration is increased, the global unfolding pathway is selectively promoted due to its high m value (Eq. 4), and it comes to dominate the exchange of progressively faster hydrogens (Eq. 5) as shown in Fig. 1A. The several hydrogen exchange curves in Fig. 1A merge to define a hydrogen exchange isotherm that reveals the global unfolding reaction and defines its ΔG and m values.

One can expect that all the faster exchanging NH's should become controlled by the global unfolding reaction as denaturant is increased. This will occur if protein unfolding behavior is adequately represented by a simple, two-state transition with no intermediate states between the native and fully unfolded forms. The data obtained show a more interesting behavior, illustrated by the GdmCl and temperature isotherms in Fig. 1, B and C. Before joining the high energy, global unfolding isotherm, the various faster hydrogens in cytochrome c assemble into several lower lying isotherms, distinguished by their different ΔG_{op} and m values. To avoid a confusion of overcrowding in Fig. 1, B and C, we show the behavior of only two hydrogens for each isotherm. In Figs. 2 to 5 we show the additional hydrogens that join each isotherm.

The various hydrogen exchange isotherms mirror the behavior seen for the global unfolding isotherm but exhibit serially decreasing values for ΔG_{op} and m . Just as the highest energy isotherm reflects the globally unfolded state, the various lower lying isotherms reveal a sequence of partially unfolded states. We refer to these subglobal partially unfolded forms by the acronym PUF. That these patterns are inherent in the cytochrome c structure and are not dependent on GdmCl interactions is supported by the similar behavior seen in temperature studies at zero GdmCl (Fig. 1C).

The amino acids that converge to form each isotherm help to identify the cooperative unfolding unit that exposes those residues to exchange (Fig. 6). The common K_{op} reached in each isotherm specifies the free energy of the PUF produced by the unfolding (Eq. 3). The slope of each isotherm (m in Eq. 4) indicates the additional GdmCl sensitive surface exposed in each PUF relative to the native state.

In the native protein, the amino-terminal helix (residues 2 to 15) and the carboxyl-terminal helix (residues 87 to 104) are docked against each other (Fig. 2) and form

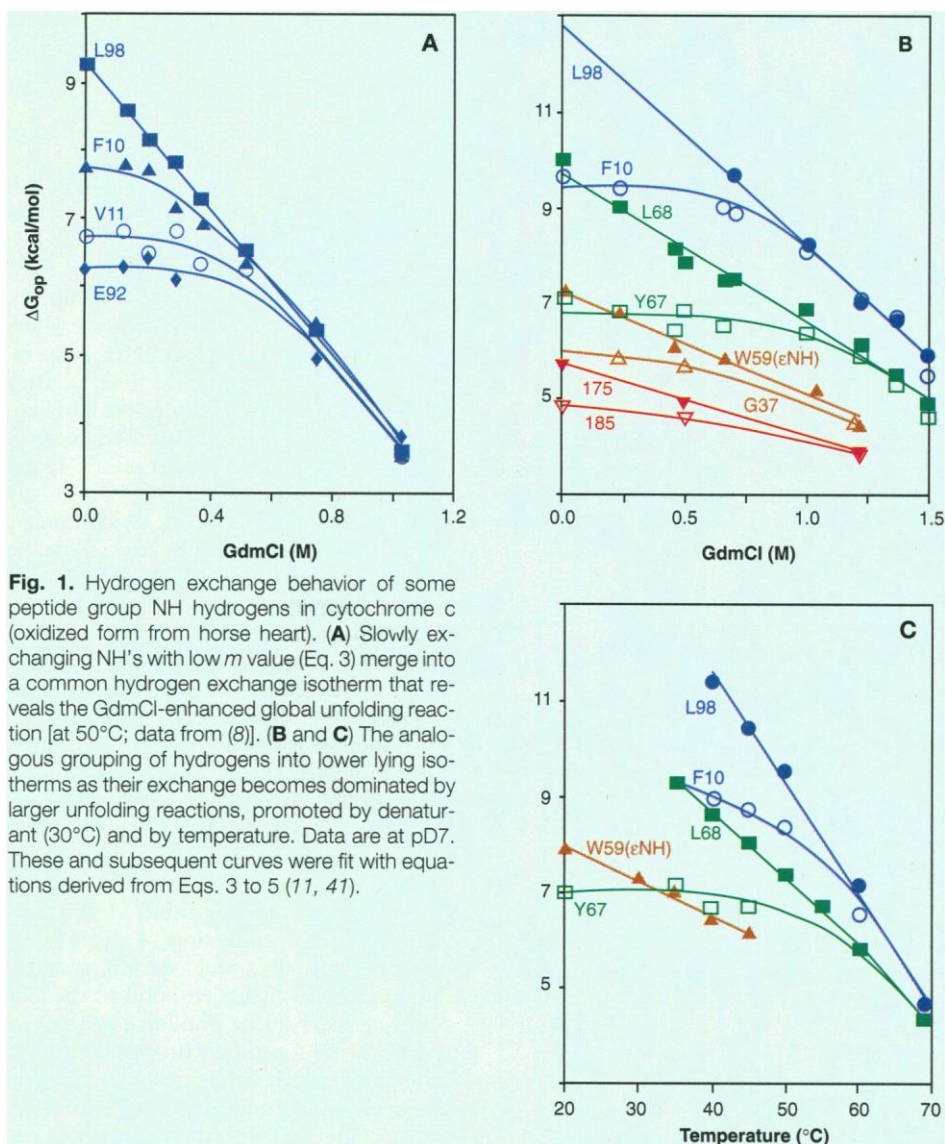


Fig. 1. Hydrogen exchange behavior of some peptide group NH hydrogens in cytochrome c (oxidized form from horse heart). (A) Slowly exchanging NH's with low m value (Eq. 3) merge into a common hydrogen exchange isotherm that reveals the GdmCl-enhanced global unfolding reaction [at 50°C; data from (8)]. (B and C) The analogous grouping of hydrogens into lower lying isotherms as their exchange becomes dominated by larger unfolding reactions, promoted by denaturant (30°C) and by temperature. Data are at pD7. These and subsequent curves were fit with equations derived from Eqs. 3 to 5 (11, 41).

a cooperative unit, the melting of which marks the final transition to the globally unfolded state (11).

All the peptide group hydrogens in the amino-terminal helix can exchange through local fluctuations with near-zero m value. As GdmCl concentration increases, these hydrogens merge with the global unfolding isotherm typified by Leu⁹⁸ (Fig. 2A). The amino-terminal segment extends to Thr¹⁹ NH, which is hydrogen bonded to a structural water (17, 18) at the far end of the amino helix.

In the carboxyl-terminal helix, the pep-

tide group hydrogens of residues 94 through 99 can exchange only through the major unfolding reaction that yields the globally unfolded form (Fig. 2B). The stable segment containing these residues must include also their hydrogen bond acceptor residues down to Glu⁹⁰ CO. Residues 100 and 101 placed near the molecular carboxyl terminus can exchange faster through local unfolding, presumably for Ala¹⁰¹, because of dynamic helix fraying that is reflected in the small but non-zero m value. Also residues 91 to 93 can exchange through some small fluctuations, presum-

ably involving motions of their hydrogen bond acceptor residues 87 to 89. These faster hydrogens all merge into the global unfolding isotherm as GdmCl concentration increases. The cooperative carboxyl-terminal segment therefore may extend beyond Glu⁹⁰ CO and include Lys⁸⁷ CO (acceptor of Arg⁹¹ NH), but it does not reach to Ile⁸⁵ NH, which exchanges in a lower energy PUF.

Residual structure in unfolded states.

Residues 14, 15, and 18 exchange through local fluctuations (Fig. 3) with near-zero m value and hydrogen exchange protection factors (2) (k_{ch}/k_{ex} in Eqs. 2 and 3) of about 10^6 in the native state. Surprisingly, when the global unfolding reaction is enhanced by GdmCl, these hydrogens cross over and exchange more slowly than the global unfolding isotherm. They experience residual protection in the globally unfolded state.

In the native protein, residues 14, 15, and 18 occupy a cyclic structure (Fig. 3) stabilized by the heme thioether bridges (Cys¹⁴ and Cys¹⁷) and the His¹⁸ to heme iron coordination. Evidently this structure persists during transient global unfolding and provides hydrogen exchange protection by a factor of at least 30 (≥ 2 kcal) (19). The continuing straight-line behavior of these hydrogens represents continuing slow exchange from the native state, which still exists 99.9998 percent of the time at the cross-over region.

The peptide group hydrogens of His³³ and Phe³⁶ also appear to show significant protection in open states (see below), which makes their assignment to a given cooperative unit somewhat insecure. Phe³⁶, His³³, and the heme-related residues exhibit additional protection compared to their free peptide exchange rates by factors of 2.5, 4, and ≥ 30 , respectively.

Partially unfolded forms. Cytochrome c contains a third sizable helix, the 60's helix, extending from residue 60 to 70. All but one of the peptide hydrogens in the 60's helix can exchange through local fluctuations with near-zero m value. The Leu⁶⁸ NH can exchange only through a more extensive unfolding reaction with large m . With increasing GdmCl, the larger unfolding is promoted (Eq. 4) and comes to dominate all the hydrogens in the 60's helix (Fig. 4A; see also Fig. 1C).

Residues 29, 32, and 33 join the hydrogen exchange isotherm defined by the 60's helix (Fig. 4B). Among these, His³³ shows some additional protection (4 times greater) in the partially unfolded form, probably due to its local hydrogen bond to the side chain of Asn³¹ (18). The large m value of the isotherm together with results for the other cooperative units point to a cooperative opening extending from Val²⁰ to Leu³⁵, nearly coincident with the omega loop, res-

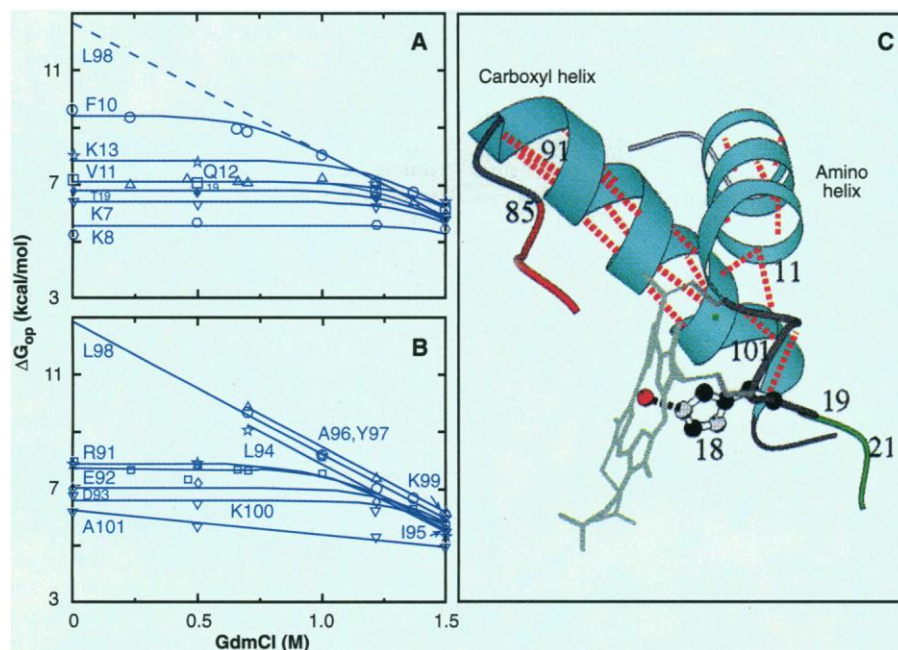


Fig. 2. Hydrogen exchange behavior defining the cooperative reversible unfolding of the amino-terminal (A) and carboxyl-terminal (B) helices. The structural positions of the hydrogens measured are illustrated (C). In these and subsequent hydrogen exchange experiments, oxidized equine cytochrome c in D₂O at pD7 and 30°C was used and then analyzed at pD5 in the reduced form by 1D and 2D NMR with the proton assignments from Wand *et al.* (42) as described (11). Molecular diagrams were obtained with Molscrip (43).

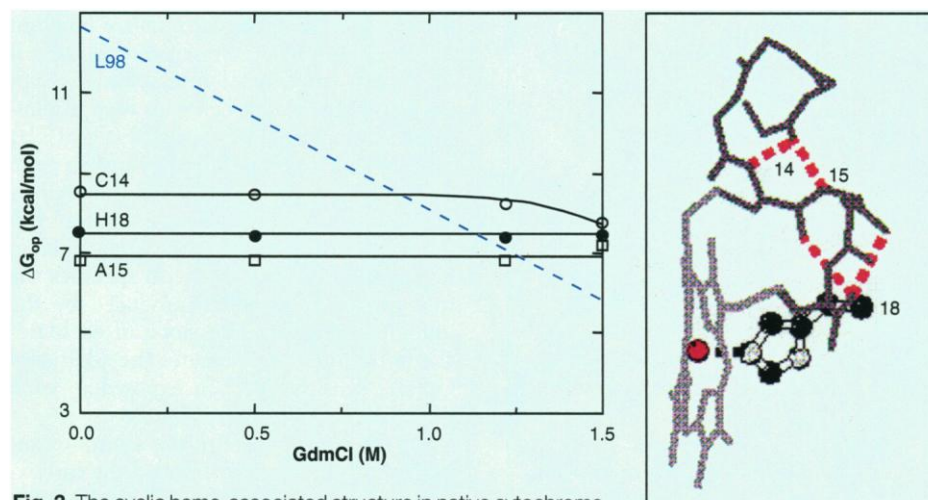


Fig. 3. The cyclic heme-associated structure in native cytochrome c. The hydrogens indicated exhibit protection against exchange in the transient globally unfolded form.

idues 18 to 33, described by Leszczynski and Rose (20).

Whether the 20 to 35 loop [green(a)] and the 60's segment [green(b)] unfold together can be questioned since they are distant in sequence. These two segments show identical m and ΔG values (Fig. 4), they are at opposite ends of a cooperative segment that may well be already open when they unfold (Fig. 6), and some of their side chains interact in the central hydrophobic core of cytochrome *c* (Leu³², Leu³⁵, Leu⁶⁴, Leu⁶⁸). The loss of one pair of core leucine residues by the unfolding of either of these segments may well destabilize the other.

The results in Fig. 5A identify a cooperative unit that exposes the peptide group hydrogens of residues 36, 37, 60, and 64 and the indole NH of Trp⁵⁹. Phe³⁶ shows some additional protection (2.5 times greater). These residues occupy the neck of a large omega loop (20) that forms the entire bottom of the protein (conventional view, Fig. 6) and connects the two segments just discussed (20 to 35 loop and 60's helix). The indole NH of Trp⁵⁹ hydrogen bonds to a heme propionate oxygen, is thoroughly buried by the bottom loop, and appears unable to exchange without disrupting it. The sizable m value observed is consistent with unfolding of the entire bottom loop.

These relationships indicate that the loop from residues 36 to 61 forms a cooperatively unfolding unit (Fig. 5). The cooperative unit must include the Lys⁶⁰ CO, the hydrogen bond acceptor of the Leu⁶⁴ NH, which joins the isotherm. It cannot include the Glu⁶¹ CO—the acceptor for Met⁶⁵ NH, which joins the more slowly exchanging 60's helix isotherm. Thus the boundary between the two cooperative units is placed at Glu⁶¹ C α .

Residues 74, 75, and 85, placed in the neck of a small omega loop that includes residues 70 to 85 (20), form a hydrogen exchange isotherm (Fig. 5B, red). Residues 74 and 75 hydrogen bond to 70 and 71, forming a short helical length orthogonal to the 60's helix. The Ile⁸⁵ NH closes the loop by hydrogen bonding either to Leu⁶⁸ CO (18) or a neighboring structural water (17). The sizable m and ΔG values indicated by the hydrogen exchange isotherm suggest that the entire 70 to 85 loop accounts for the unfolding reaction.

The cooperative units identified here are shown color coded in order of increasing free energy (red to blue) in the cytochrome *c* protein (Fig. 6).

The hydrogen exchange isotherm for the amino-terminal helix [blue(a)] extends up to Lys¹³ NH, but the pattern of protection observed requires the cooperatively unfolding unit to reach only to Ile⁹ CO, the Lys¹³ NH acceptor. The heme-

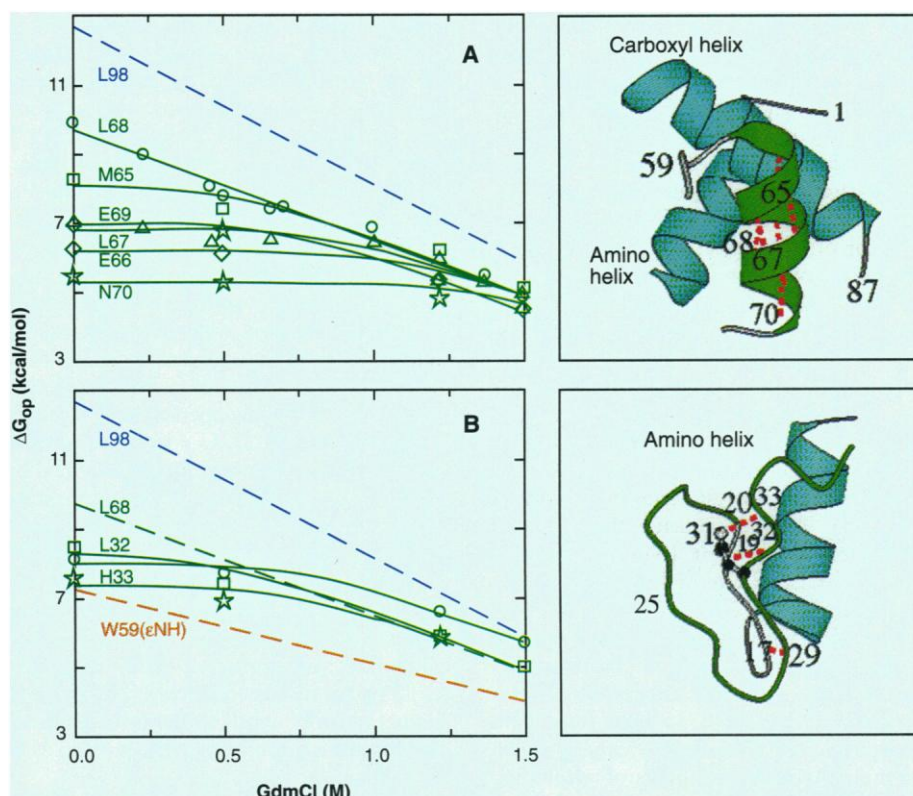


Fig. 4. Hydrogens of the 60's helix (A) and the 20 to 35 loop (B) and their common hydrogen exchange isotherm. Dashed lines show the positions of neighboring isotherms. The two cooperative segments and their neighboring structure are illustrated.

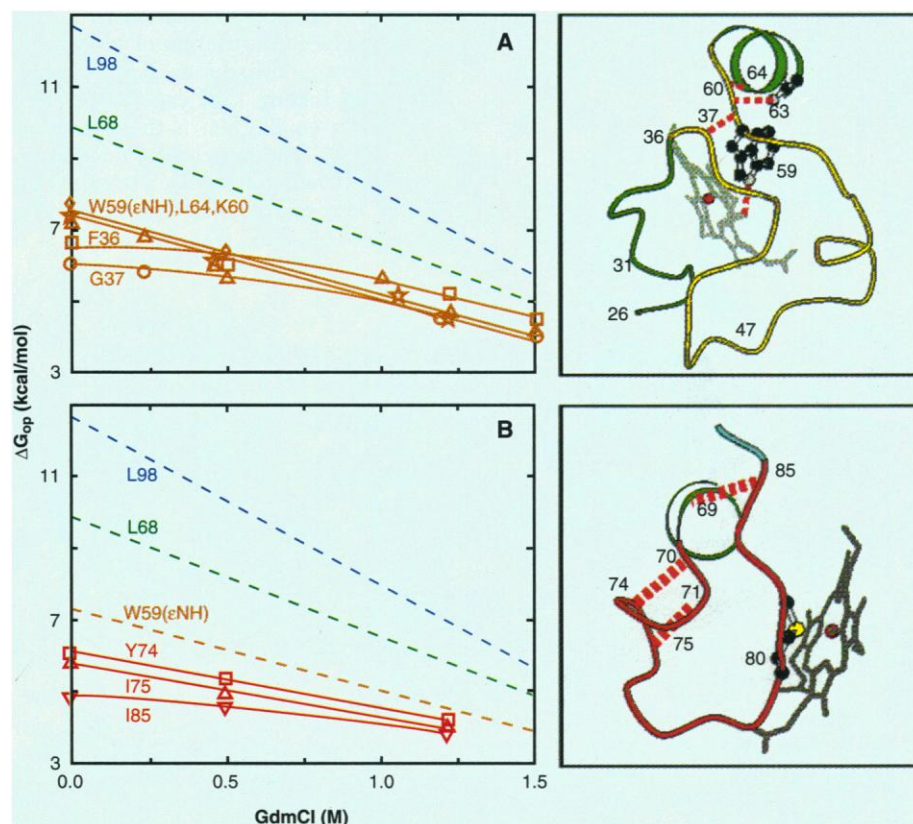


Fig. 5. The large bottom loop isotherm (A) and the small loop isotherm (B). The omega loops in the native protein are illustrated.

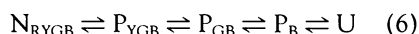
associated residual structure presumably reaches back to Phe¹⁰ CO, the hydrogen bond acceptor for residues 14 and 15 in native cytochrome c (17, 18). That the omega loop reaching from Val²⁰ to the reverse turn of residues 32 to 35 forms a cooperative segment [green(a)] is suggested by the isothermal hydrogen exchange behavior of residues 29, 32, and 33, with the slower exchange of Thr¹⁹ at one end and the faster exchange of Phe³⁶ at the other. The bottom segment, Phe³⁶ to Glu⁶¹ C α , is an omega loop (yellow). The cooperative 60's helical segment [green(b)] includes the entire helix from the C α of Glu⁶¹ to Asn⁷⁰ C α . Residues 70 to 85 comprise an omega loop (20) that represents the lowest energy unfolding unit (red). The cooperative carboxyl-terminal helical segment [blue(b)] reaches from either Lys⁸⁶ C α or Lys⁸⁷ C α to the chain terminus.

In summary, the results obtained specify all the cooperative unit boundaries within one residue, with the exception of the boundary at the 32 to 35 beta turn. The cooperative units are entire omega loops and mutually stabilizing pairs of whole helices and loops, with a segment size of about 15 residues.

Cooperative unit relations. The hydrogen exchange isotherm produced by each PUF identifies one cooperative unit that is unfolded but does not specify whether the cooperative units that have already exchanged their hydrogens in lower energy isotherms are folded or unfolded. The pos-

sible options for the structural identity of each PUF can be considered by comparing the m value measured for each isotherm with the additional surface area exposed (Δarea), computed for each possible PUF identity, as in Fig. 7. A fairly direct correlation between m and Δarea can be expected. One limiting option (left-most symbols) is that the hydrogen exchange patterns reflect the independent opening of the individual cooperative units from the native state (only red open, only yellow open, and so forth). At the other extreme (filled symbols), a sequential model is considered (red open $\Rightarrow$ red + yellow open $\Rightarrow$ red + yellow + green open $\Rightarrow$ U). Guidance in placing the expected correlation line is provided by the fact that the identity of the native state (N; all units intact) and the unfolded state (U; all units unfolded) are unambiguous.

The sequential model (filled symbols) provides a reasonable correlation and makes sense structurally and energetically, especially from the point of view of kinetic refolding (as discussed below). The suggested energetically uphill sequence of PUFs for cytochrome c unfolding can be written as in Eq. 6.



However, a role for the other possible PUFs (open symbols in Fig. 7) cannot be excluded.

A partly similar sequence has been inferred by Fisher and Taniuchi (21) from studies of the reconstitution of structure by complementing cytochrome c fragments.

Protein folding. Our experiments were done under equilibrium native conditions (pD7, 30°C). The principle of microscopic reversibility requires that each transient unfolding step must be accompanied by an equivalent refolding step. Unfolding steps

that carry a lot of traffic must be matched by major refolding steps, and minor with minor, so that the equilibrium concentration of each species is maintained. Therefore the sequence of reactions that carry cytochrome c from N to the various intermediate PUFs and to U must be matched by reverse reactions that carry it back to N and thus specify the refolding as well as the unfolding pathway (or pathways), as suggested in Eq. 6.

Some correlations with prior information suggest the reality of the pathway in Eq. 6. The lowest energy unfolding found here (opening of the 70 to 85 loop) includes the weak Met80 to heme ligand. The loss of this ligation is known to accompany the first step in cytochrome c unfolding (22, 23), and its reformation marks the final step in cytochrome c folding (24). This is the step $N_{\text{RYGB}} \rightleftharpoons P_{\text{YGB}}$ in Eq. 6, which is required by both the sequential and independent unfolding models.

The last unfolding amino-terminal plus carboxyl-terminal unit found here is the first element seen to fold in hydrogen exchange pulse labeling experiments (25). This is consistent with the step $\text{U} \rightleftharpoons P_{\text{B}}$ in Eq. 6. Further folding is transiently blocked at this step by a misfold-reorganization barrier (26) because of the misligation of the heme iron by either His²⁶ or His³³ (19, 26, 27). The sequence in Eq. 6 predicts this result, since both histidines are placed within the 20 to 35 loop [green(a)]. The misligation of either histidine may well allow the early folding bihelical unit (blue) to form, but it will block the folding of the 20 to 35 loop [green(a)], and therefore also the 60's helix [green(b)] and the subsequent folding of the other cooperative units. Isolated complexes of the amino and carboxyl helices are known to exhibit independent

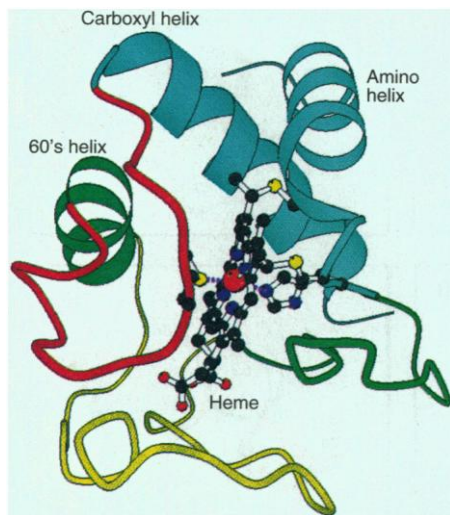
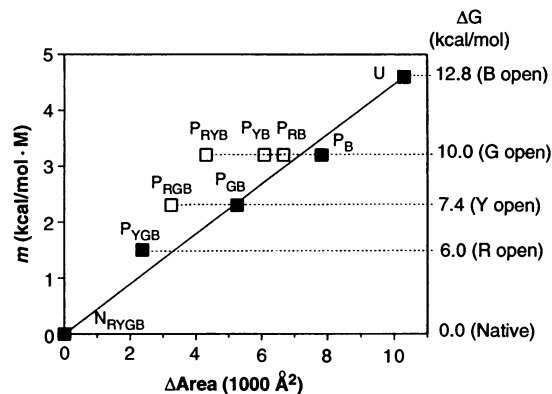


Fig. 6. The cooperative units in cytochrome c, color coded in order of increasing free energy level (red to blue), as follows: the 70 to 85 loop (red), the 36 to 61 loop (yellow), the 20 to 35 loop [green(a)] and 60's helix [green(b)], the amino-terminal helix [blue(a)] and carboxyl-terminal helix [blue(b)]. Core side chains of the green(a) and green(b) segments make contact beyond the far edge of the heme (17, 18, 44).

Fig. 7. A test for the identities of the different PUFs in terms of the cooperative units that may be folded and unfolded in each. The m value measured for each PUF is plotted against the additional solvent accessible surface (Δarea), relative to the native protein, computed for each possible PUF identity. Each possible partially unfolded form (P_x) is identified in terms of the cooperative units that remain folded, noted by the energy-related color code of Fig. 6 (R for red, Y for yellow, G for green, B for blue). Exposed surface calculated for each candidate PUF comes from the newly exposed protein and heme surface plus the cooperative segments that are unfolded, approximated as fully extended. The scale on the right shows the cooperative unit seen to be unfolded in each hydrogen exchange isotherm and the Gibbs free energy for the unfolded state, obtained from the hydrogen exchange isotherm at zero GdmCl concentration (pD7, 30°C). The correlation line drawn is determined by the fact that the unfolding of the amino helix plus carboxyl helix unit (blue) is known to produce the globally unfolded state (U) (11). The ΔG and m parameters relate to the whole molecule PUF and not merely to the individual cooperative unit newly unfolded.



stability (28, 29). The other cooperative units defined here are not independently stable and so may fold only when they are stabilized by docking against some earlier formed supporting structure. This too is consistent with the folding sequence in Eq. 6.

These observations tend to confirm the step $N \rightleftharpoons P_{YGB}$ in Eq. 6 and support the dominance of the step $P_B \rightleftharpoons U$ rather than the various alternatives shown in Fig. 7 ($U \rightleftharpoons P_{RB}$, P_{YB} , or P_{RYB}). Only the second intermediate in Eq. 6, which may in principle involve P_{RGB} as well as P_{GB} (Fig. 7), is unsupported by external evidence, but the correlation in Fig. 7 favors P_{GB} .

It seems especially promising that the hydrogen exchange analysis described here makes possible the direct study of conformationally excited, partially unfolded states of protein molecules under native conditions. Conventional equilibrium and kinetic unfolding and refolding analyses can detect intermediate states only when they are well-populated. The native state hydrogen exchange experiment may reveal all the partially unfolded intermediates that exist between the native and unfolded states even though they are only infinitesimally populated under the conditions studied. The only limitations appear to be the need for adequate probe hydrogens and the requirement for a sufficient dynamic range in ΔG_u and m . The present results point to a small number of cooperative, metastable intermediates. Forms with structures between the metastable PUFs do not appear in the hydrogen exchange analysis and therefore must exist only at higher free energy than the intermediates seen. The resolution of protein molecules into their component cooperative elements in this way promises to illuminate the fundamental nature of protein structure, including their design, construction, and biological evolution.

Other implications concern protein folding. Proteins cannot fold in one undifferentiated, whole molecule, search-dependent reaction (30), but must move to the native state through intermediate forms (31–34). The metastable, native-like inter-

mediate forms found in our study support a folding strategy that carries cytochrome c through a single linear pathway, consisting of a small number of discrete intermediates, in steps that fold one cooperative unit at a time (Eq. 6). The results seem to be inconsistent with models that suggest a large number of alternative intermediate states (35–38).

The cooperative units in cytochrome c are 15 to 27 residues in size; the individual segments involved are about 15 residues in length (Fig. 6). This reflects on the search processes visualized in the Levinthal folding paradox (30). It is far easier to fold a number of cooperative 15-residue segments (time $\propto 7 \times 3^{15} \approx 10^8$) than to concertedly fold one whole-molecule 100-residue segment (time $\propto 3^{100} \approx 10^{48}$), since the size of the cooperative unit appears in the exponent (39). Further, the native-like nature of the PUFs found here rationalizes the problem of an amino acid code for protein folding, since the formation of intermediates can then be driven by the amino acid sequence according to the same rules that determine the native state. Finally, it is interesting to consider the relations between cooperative structural units, their folding sequence, and their role in the biological evolution of contemporary protein structures. For more discussion of these issues see reference (40).

REFERENCES AND NOTES

1. A. Hvdt and S. O. Nielsen, *Adv. Protein Chem.* **21**, 287 (1966).
2. S. W. Englander and N. R. Kallenbach, *Q. Rev. Biophys.* **16**, 521 (1984).
3. S. W. Englander *et al.*, *Science* **256**, 1684 (1992).
4. R. S. Molday, S. W. Englander, R. G. Kallen, *Biochemistry* **11**, 150 (1972).
5. Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins* **17**, 75 (1993).
6. G. P. Connelly, Y. Bai, M. Jeng, S. W. Englander, *ibid.*, p. 87.
7. K. U. Linderström-Lang, in "Symposium on Peptide Chemistry," *Chem. Soc. Spec. Publ.* **2**, 1 (1955).
8. S. W. Englander, *Ann. N.Y. Acad. Sci.* **244**, 10 (1975).
9. A. Rosenberg and K. Chakravarti, *J. Biol. Chem.* **243**, 5193 (1968).
10. C. Woodward, *Trends Biochem. Sci.* **18**, 359 (1993).
11. Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins* **20**, 4 (1994).
12. C. N. Pace, *Methods Enzymol.* **131**, 266 (1986).
13. J. A. Schellman, *Biopolymers* **26**, 549 (1987).
14. G. I. Makhataдзе and P. L. Privalov, *J. Mol. Biol.* **226**, 491 (1992).
15. S. L. Mayo and R. L. Baldwin, *Science* **262**, 873 (1993).
16. H. Qian, S. L. Mayo, A. Morton, *Biochemistry* **33**, 8167 (1994).
17. P. X. Qi, D. L. Di Stefano, A. J. Wand, *Biochemistry* **33**, 6408 (1994).
18. G. W. Bushnell, G. V. Louie, G. D. Brayer, *J. Mol. Biol.* **214**, 585 (1990).
19. Elöve *et al.* measured a hydrogen exchange protection factor of 140 for His¹⁸ in unfolded cytochrome c [G. A. Elöve, A. K. Bhuyan, H. Roder, *Biochemistry* **33**, 10271 (1994)].
20. J. F. Leszczynski and G. D. Rose, *Science* **234**, 849 (1986).
21. A. Fisher and H. Taniuchi, *Arch. Biochem. Biophys.* **296**, 1 (1992).
22. B. T. Nall, *Biochemistry* **25**, 2974 (1986).
23. E. Schechter and P. Saludjian, *Biopolymers* **5**, 788 (1967).
24. J. A. Ridge, R. L. Baldwin, A. M. Labhardt, *Biochemistry* **20**, 1662 (1981).
25. H. Roder, G. A. Elöve, S. W. Englander, *Nature* **335**, 700 (1988).
26. T. R. Sosnick, L. Mayne, R. Hiller, S. W. Englander, *Nature Struct. Biol.* **1**, 149 (1994).
27. D. N. Brems and E. Stellwagen, *J. Biol. Chem.* **258**, 3655 (1983).
28. Y. Kuroda, *Biochemistry* **32**, 1219 (1993).
29. L. C. Wu *et al.*, *ibid.*, p. 10271.
30. C. Levinthal, *J. Chim. Phys.* **65**, 44 (1968).
31. P. S. Kim and R. L. Baldwin, *Annu. Rev. Biochem.* **59**, 631 (1990).
32. C. R. Matthews, *ibid.* **62**, 653 (1993).
33. A. R. Fersht and K. A. Dill, *Curr. Opin. Struct. Biol.* **4**, 67 (1994).
34. O. B. Ptitsyn, *Protein Eng.* **7**, 593 (1994).
35. P. G. Wolynes, J. N. Onuchic, D. Thirumalai, *Science* **267**, 1619 (1995).
36. A. Sali, E. Shakhnovich, M. Karplus, *Nature* **369**, 248 (1994).
37. K. Kim and C. Woodward, *Biochemistry* **32**, 9609 (1993).
38. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins* **21**, 167 (1995).
39. O. B. Ptitsyn and A. A. Rashin, *Biophys. Chem.* **3**, 1 (1975).
40. M. F. Jeng and S. W. Englander, *J. Mol. Biol.* **221**, 1045 (1992).
41. Y. Bai, *thesis*, University of Pennsylvania (1994).
42. A. J. Wand *et al.*, *Biochemistry* **28**, 186 (1989).
43. P. J. Kralius, *J. Appl. Crystallogr.* **24**, 946 (1991).
44. T. Takano and R. E. Dickerson, *J. Mol. Biol.* **153**, 79 (1981).
45. We thank T. Hancock and J. S. Milne for technical assistance, A. J. Wand and G. D. Brayer for supplying structural coordinates and hydrogen bonding patterns for cytochrome c derived from NMR and x-ray results, and R. L. Baldwin, W. F. deGrado, N. R. Kallenbach, C. R. Matthews, G. D. Rose, and the members of this laboratory for helpful discussion. Supported by NIH research grant GM31847.

26 January 1995; accepted 7 June 1995