

FUTURE DIRECTIONS

Future Directions in Folding: The Multi-State Nature of Protein Structure

Yawen Bai¹ and S. Walter Englander²

¹*Department of Molecular Biology, Scripps Research Institute, LaJolla, California 92037-9701; and* ²*The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6059*

ABSTRACT All possible protein folding intermediates exist in equilibrium with the native protein at native as well as non-native conditions, with occupation determined by their free energy level. The study of these forms can illuminate the fundamental principles of protein structure and folding. Hydrogen exchange methods can be used to detect and characterize these partially unfolded forms at native conditions and as a function of mild denaturant and temperature. This information illuminates the requirements that govern the ability of kinetic and equilibrium methods to study folding intermediates. © 1996 Wiley-Liss, Inc.

Key words: protein stability, protein dynamics, hydrogen exchange

INTRODUCTION

To understand the fundamental nature of protein molecules—their stability, folding, cooperativity, and biological evolution—one would like to be able to adopt the divide and conquer approach of the biochemist, take proteins apart into their component pieces, examine the parts, and see how they fit back together again. The cooperative nature of protein structure makes this difficult but by no means impossible. A number of promising approaches have been devised, and the characterization of partially folded intermediates by kinetic, equilibrium, synthetic, and theoretical methods has become a central focus of protein studies.

Recent hydrogen exchange (HX) experiments with cytochrome c (cyt c) significantly extend this capability.^{1,2} Initial results show that some hydrogens in cyt c exchange with solvent hydrogens through transient global unfolding reactions, some through sub-global openings, and others through smaller, more local fluctuations. The sub-global unfolding reactions identify a set of four cooperative substructural units that together compose the entire cyt c molecule (Fig. 1). Combinations of these units

may produce the intermediate structures that define the folding and unfolding pathways of cyt c. Here we briefly illustrate this new capability for studying the multistate nature of protein structures and then consider some lessons that can be useful for future structural studies in equilibrium, kinetic, and hydrogen exchange modes.

THEORETICAL BASES

Under native conditions the vast majority of protein molecules exist in their unique native state (N). A tiny fraction must also occupy all possible higher energy states as dictated by the Boltzmann relationship, $P_n/N = \exp(-\Delta G^\circ/RT)$. Even under native conditions, protein molecules continually unfold and re-fold, cycling through the fully unfolded state (U) and all the partially unfolded forms that protein chemists have tried so hard to characterize. Because protein molecules are highly cooperative structures, most partially unfolded states will exist only at free energy levels higher than U, but, since cooperativity is not infinite, some may well exist at free energy levels between the native and globally unfolded forms. A simple, non-continuous distribution of higher energy states is pictured in Figure 2. More complex free energy surfaces have been the subject of active theoretical investigations.^{3–7}

The minor protein forms that exist at elevated energy levels are invisible to most experimental measurements, which are swamped by signals from the abundant native state. Hydrogen exchange is exceptional in that the native state makes no contribution to the measurement. Slowly exchanging hydrogens protected by a protein's native structure can exchange with solvent hydrogens only during some transient, high energy opening reaction:

Received September 7, 1995; accepted September 11, 1995.

Address reprint requests to S. Walter Englander, the Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6059.

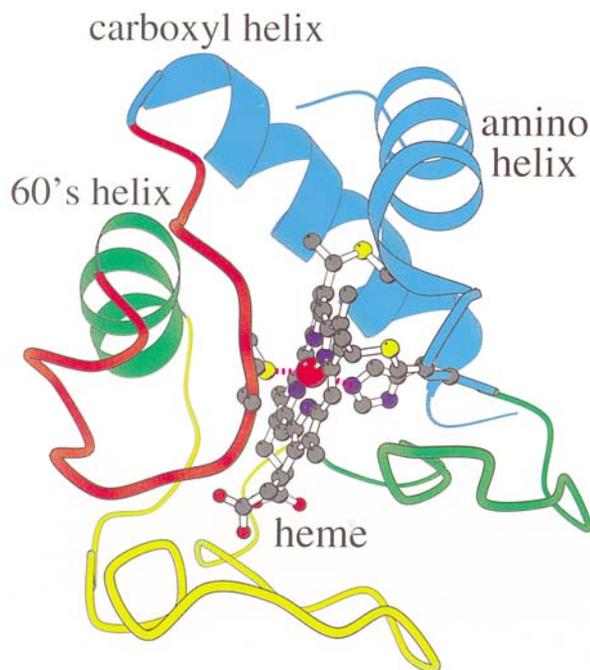
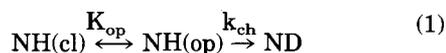


Fig. 1. A set of four cooperative unfolding units constitute the cyt c molecule and may define the intermediates in cyt c unfolding and refolding pathways. The colors code for increasing free energy of unfolding in the order red, yellow, green, blue. Side chains of the two green segments make contact in the central core at the far edge of the heme.



Therefore states within the excited state manifold that include broken hydrogen bonds may be accessible to hydrogen exchange investigation.

Eq.(2) gives the HX rate generated by this reaction scheme. Eq.(3) translates the exchange rate to the free energy of the opening reaction.

$$k_{\text{ex}} = K_{\text{op}}/(K_{\text{op}} + 1) \approx K_{\text{op}}K_{\text{ch}} \quad (2)$$

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

These equations, due originally to Linderstrom-Lang,⁸⁻¹⁰ define the almost universally observed EX2 limit (bimolecular exchange) of hydrogen exchange behavior,⁹ in which structure is stable ($K_{\text{op}} \ll 1$) and structural reclosing rate is fast compared with k_{ch} , the chemical exchange rate for the fully solvent-exposed hydrogen.^{11,12} The exchange rate, k_{ex} , is then proportional to K_{op} , essentially the fraction of time that the productive opening exists [Eq.(2)], and Eq.(3) follows.

Equations 1-3 are independent of the structural details of the opening reactions. Any given hydrogen may be exposed to exchange by the separation of a single hydrogen bond^{9,13} or more sizeable openings¹⁴ ranging even up to transient whole molecule unfolding,¹⁵ as in Equation 4.

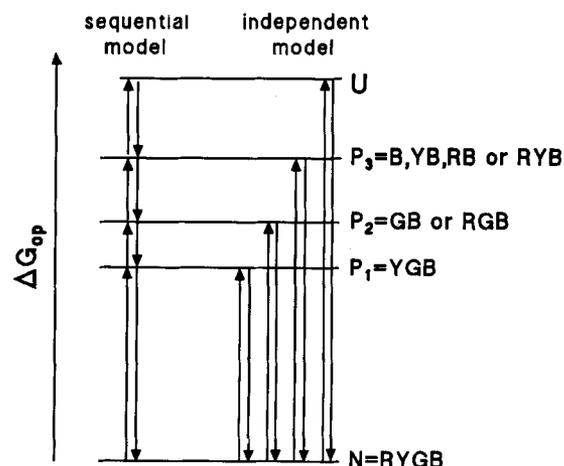


Fig. 2. The energy manifold of states found for cyt c, including the native (N), unfolded (U), and conformationally excited, partially unfolded states (Pi). The different possible identities for the PUFs detected in the HX experiment are indicated in terms of the cooperative units still remaining folded, identified by the color code in Figure 1 (R, Y, G, B = red, yellow, green, blue). The unfolding of the blue unit is known to represent full global unfolding (see text). Microscopic reversibility requires the pairing of unfolding and refolding reactions, as shown. The arrows show the extreme unfolding models; additional opening-closing reactions can be envisioned.

$$k_{\text{ex}} = K_{\text{op,eff}}k_{\text{ch}} = [K_{\text{op}}(\text{local}) + K_{\text{op}}(\text{subglobal} + K_{\text{op}}(\text{global})]k_{\text{ch}} \quad (4)$$

Eq.(4) connects the hydrogen exchange behavior of a protein with subcategories of conformationally excited states that break hydrogen bonding interactions. For any given exchangeable hydrogen, the unfolding intermediate that attains the greatest K_{op} will dominate the exchange. One may investigate the different kinds of states in Eq.(4) by manipulating their free energy levels, and thus the relative populations of unfolding intermediates, and then appropriately interpreting the HX signals that result.

GLOBAL UNFOLDING

The melting transition of cyt c, measured in the conventional way as a function of guanidinium chloride (GdmCl) by changes in intrinsic fluorescence, is shown in Figure 3A (open squares). The usual lengthy extrapolation to zero denaturant concentration gives an estimate of the free energy for global unfolding. Figure 3A also shows H-D exchange results measured by two-dimensional nuclear magnetic resonance (NMR) for several very slowly exchanging peptide group hydrogens in the COOH-terminal helix of cyt c. The exchange rates together with Eqs.(2) and (3) yield the correct value for the free energy of global cyt c unfolding.¹ Similar success has been found for staphylococcal nuclease,¹⁶ ribonuclease A,^{1,17} barnase,¹⁸ staphylococcal protein A domain B (Y. Bai and P. Wright, personal communication), and turkey ovomucoid third do-

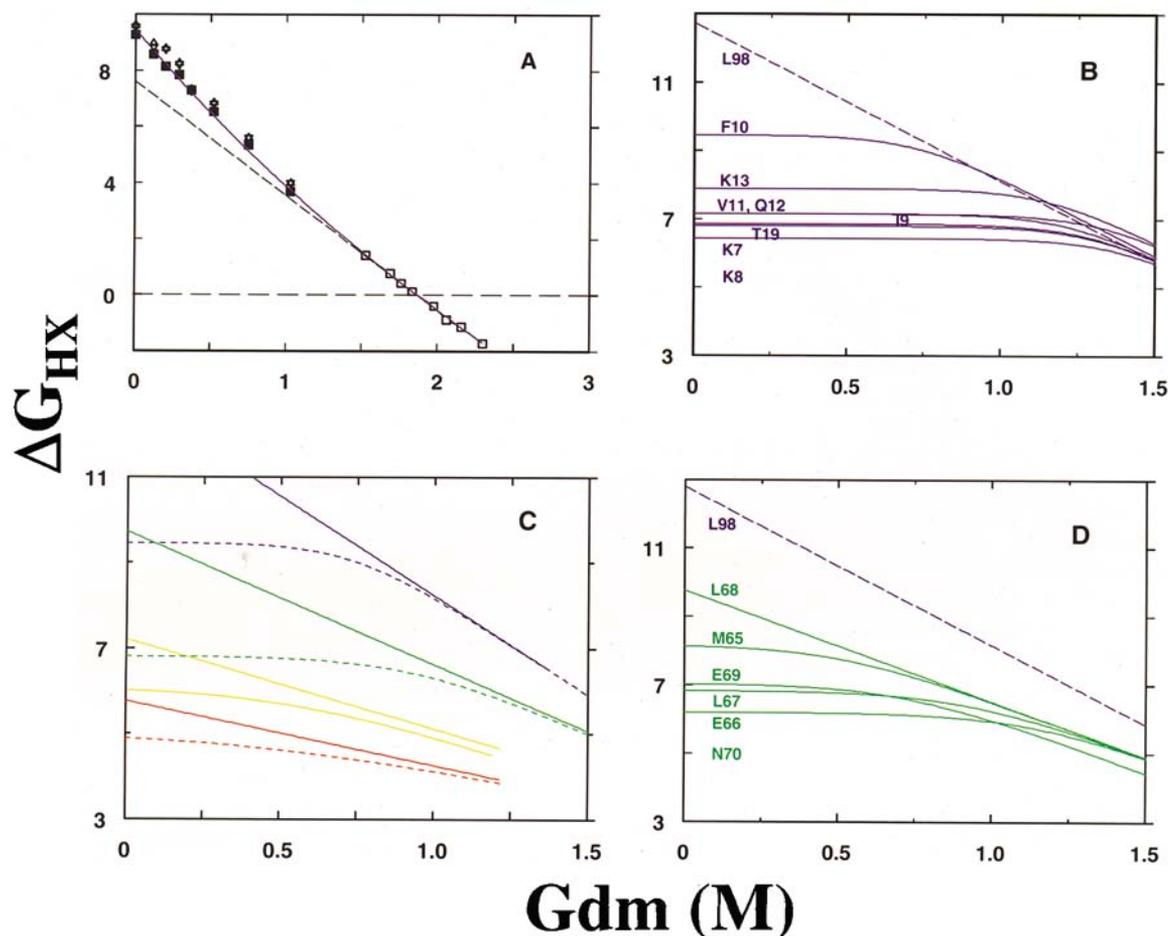


Fig. 3. The dependence on denaturant of H-exchange due to global, sub-global, and local unfolding in cyt c. **A:** Open symbols show classical melting data through the two-state transition. The symbols at lower GdmCl are from HX data for six residues in the $COOH$ -terminal helix, which can only exchange by way of global unfolding. **B:** Residues in the NH_2 -terminal helix exchange by way

of small fluctuations at low GdmCl but become dominated by the global unfolding when it is promoted by denaturant. **C:** Sub-global isotherms, with only two amino acid residues shown per isotherm to minimize clutter. **D:** Exchanging hydrogens in the 60's helix merge to form the green isotherm. Data were measured at 50°C for A and 30°C for B–D, all at pD 7.²

main (L. Swint-Kruse and A. Robertson, personal communication). Some of the slowest hydrogens in these proteins exchange with solvent hydrogens only during the small fraction of time when the protein experiences the globally unfolded state.

The dependence of unfolding free energy on denaturant concentration is usually expressed as in Equation 5.

$$\Delta G(\text{den}) = \Delta G(0) - m[\text{den}] \quad (5)$$

The slope, m , depends on the denaturant binding surface newly exposed in the unfolding reaction.^{19,20}

The large m value characteristic of the GdmCl-sensitive surface exposed in the global unfolding reaction is reflected in the exchange behavior of the slow hydrogens in Figure 3A.

Figure 3B identifies some hydrogens that exhibit near zero m values. Their exchange is determined by

unfolding reactions that expose little new surface. At low GdmCl, global unfolding still occurs but only at a low level that makes no contribution to the measured exchange (Eq. 4). When GdmCl is increased, the global unfolding is sharply promoted (large m) and comes to dominate the exchange of progressively faster hydrogens, which then merge with the global curve to form an HX isotherm. This behavior clearly distinguishes local and global unfolding reactions and provides a method for measuring protein thermodynamic parameters far below the melting transition.

SUBGLOBAL UNFOLDING

One can expect that all the faster exchanging hydrogens will ultimately become dominated by the global unfolding reaction when it is sufficiently enhanced. This does occur, but first a more interesting

intermediate behavior is seen (Fig. 3C). Faster exchanging hydrogens merge into a sequence of three lower lying isotherms, each one analogous to the global unfolding isotherm but with progressively smaller ΔG_{HX} and m . Figure 3C provides an overall view with only two hydrogens for each isotherm. Each isotherm contains many hydrogens, as illustrated in Figure 3D for one of them. Just as the highest energy isotherm portrays the global unfolding equilibrium, the lower lying isotherms reflect partially unfolded forms (PUFs), with smaller unfolding free energies and less surface exposure. That this behavior is not a denaturant-dependent artifact is shown by the fact that analogous data can be obtained as a function of temperature.² The temperature-dependent HX data respond to different parameters (enthalpy and entropy of unfolding) but reveal the same subglobal unfolding units as the GdmCl experiments.

The hydrogen exchange data reveal the identity of the individual cooperative units (from the residues that join each HX isotherm; e.g., Fig. 3B, D), the free energy of each subglobal opening reaction (from the HX rates; Eqs. 2, 3), and the additional GdmCl-sensitive surface exposed in each unfolding [from the m value; Eq.(5)]. The results portray four cooperative unfolding units that together make up the cytochrome *c* molecule (Fig. 1). The red and yellow units in Figure 1 represent entire omega loops, previously defined by Leszczynski and Rose.²¹ The green unit is composed of an omega loop (green-a) and the 60's helix (green-b). The blue unit includes the NH₂-terminal (blue-a) and COOH-terminal (blue-b) helices of cyt *c*, the high energy unfolding of which marks the final transition to the U state (Fig. 3B).

The HX results further show that these units unfold cooperatively in all-or-none transitions. Structural forms between the identified PUFs must in principle exist, but evidently they exist only at higher free energy levels as invisible transitional forms between the cooperatively unfolding PUFs that are observed. Many hydrogens that ultimately join an isotherm can exchange faster than the group isotherm rate at low GdmCl (e.g., Fig. 3B,D), but these do not represent partial PUFs. The near-zero m value shows that this faster exchange represents local unfoldings rather than the unfolding of a large fraction of a cooperative unit.

UNFOLDING AND REFOLDING INTERMEDIATES

The HX data fail to specify the full identity of the cooperative unfolding units. One cannot tell whether each structural unit in Figure 1 unfolds and exchanges independently or whether the unfolding of a given unit occurs together with one or more of the lower energy, faster exchanging units. These alternatives are suggested in Figure 2 (additional

possibilities exist). For example, when the green unit cooperatively unfolds to form P₃, the unfolded form may or may not include also the red and/or the yellow units. One cannot tell because hydrogens in the green isotherm are measured only after the hydrogens that reveal the behavior of the red and yellow cooperative units have already exchanged. The HX data show only that the green unit is open and the blue unit is not.

If the openings occur in a sequentially dependent manner—red alone, then red + yellow, then red + yellow + green, and finally all the units together in the global unfolding—then the PUFs represent steps in a kinetic unfolding sequence, as in Equation 6 and Figure 2.



The same PUFs must then also define the major refolding pathway of cyt *c*, because the HX experiment at each GdmCl concentration is done at equilibrium conditions. To maintain a constant equilibrium concentration of each species, any given unfolding reaction must be matched by an equal and opposite refolding reaction, as in Eq.(6) (microscopic reversibility).

Available evidence² supports the reality of the sequential unfolding model (Eq. 6; Fig. 2), but this is by no means definitive. Alternative opening models (Fig. 2) would produce distinctly different PUF structures. Either conclusion will be interesting in respect to protein substructure but the implications for kinetic folding will be very different. This central issue remains as a challenge for future investigation.

DETECTION RULES

The results summarized here document the underlying multi-state nature of the cyt *c* molecule. We suppose that other proteins are similarly constructed, yet protein molecules often appear to be highly cooperative, two-state structures. The present results help to explain this contradiction and illuminate some general detection rules for observing folding intermediates in equilibrium and kinetic experiments.

Equilibrium Unfolding Experiments

Figure 4 illustrates the equilibrium free energy relationships in cyt *c* that connect the native state (N), the fully unfolded state (U), the intermediate PUFs (P_n), and also some hypothetical intermediates (dashed lines). To distinguish a partially folded intermediate in an equilibrium unfolding experiment, the intermediate must be populated to a significant degree relative to both the N and U forms. In most cases this does not occur,²² for reasons that Figure 4 helps to make clear. The arrows in Figure 4 indicate the range of conditions within which un-

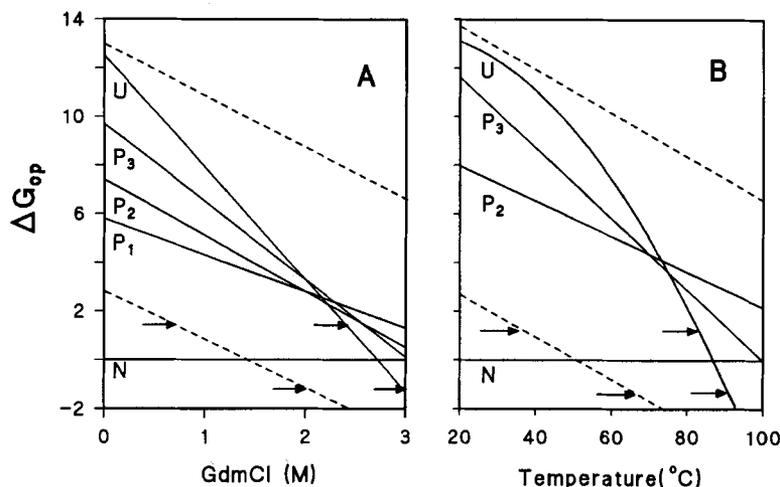


Fig. 4. Crossover curves from data like that in Figure 3. The native state (N) is taken as a reference ($\Delta G = 0$; free energies are in kcal/mol). The global unfolding curve is shown by U and some hypothetical intermediates by the dashed lines. Arrows indicate the range accessible to classical equilibrium melting experiments.

HX unfolding isotherms for the cyt c intermediates, indicated by P_n , are extrapolated from GdmCl results (A) (e.g., Fig. 3; pD7, 30 $^{\circ}$ C) and from some more limited data for HX as a function of temperature (B).² From this behavior some general rules for the detection of folding intermediates can be inferred (see text).

folding can be accurately measured by spectroscopy or calorimetry in equilibrium unfolding experiments ($U/N \sim 0.05$ to 0.95). The imaginary intermediate shown by the lower dashed line, which might represent some molten globule form²³ under mildly destabilizing conditions, achieves significant population and could be directly studied, but the upper level hypothetical intermediate could not.

The metastable PUFs of cyt c (P_n in Fig. 4) occupy free energy wells that are always higher than one or the other of the N and U forms. They are more stable than U at low denaturant but are overtaken by U before they approach N, and therefore will not be detectable in equilibrium unfolding experiments. Conditions might be found that favorably modify these free energy relationships. Mutations that selectively destabilize N relative to some intermediate can be useful²⁸ but denaturants and temperature will not since they are likely to promote global more than subglobal unfolding.

Native State Hydrogen Exchange

The native state HX experiment can detect equilibrium unfolding intermediates that are only infinitesimally populated under native conditions because, unlike other equilibrium experiments, N does not compete. It is only necessary that the intermediate achieve significant population relative to U and to any other unfolded state that allows the same hydrogens to exchange. The native state HX experiment is based on the ability to manipulate this competition by use of denaturant or temperature to selectively promote large scale openings [Eq.(5)].

PUF detection in an HX experiment does not depend on the kinetic barriers between N, U, and the

intermediates. Conversely the HX experiment gives no information on the kinetic folding and unfolding barriers that separate these states. The only rate-based condition stems from the fact that these measurements utilize HX kinetics. The cycling of protein molecules through a PUF must not be slower than the HX rate of the hydrogens that define it. PUFs that are kinetically isolated and equilibrate too slowly with the native state will escape HX detection.

The ability to resolve and identify different PUFs benefits from a large number of exchangeable hydrogen probes and large dynamic range in ΔG and m . Disadvantageous characteristics are small size, which reduces m , and destabilizing conditions, which will cause separate isotherms to merge. Cyt c is not unusually rich in slowly exchanging peptide NH probes but has high stability (~ 13 kcal/mol at the condition studied) and is sizeable, which provides a favorable dynamic range in both ΔG and m . The ability to assign measured amino acid residues to one cooperative unit or another depends also on how precisely their individual hydrogens appear to join a common HX isotherm. This is conditioned by the accuracy of the HX measurement and the ideality of exchange behavior in the open form.

Low-lying HX isotherms may be only poorly definable. Their m values are small and their hydrogens merge with the isotherm only where the different isotherms approach each other and become difficult to distinguish. This restricted detection of low-energy, fast-exchanging PUFs may account for the large energy gap between the native and the lowest energy PUF found here (though see ref. 6).

Kinetic Folding and Unfolding

To detect an intermediate as an independent species in kinetic folding or unfolding experiments, it must occupy a free energy well that is lower than all prior wells in the pathway and must be blocked by a barrier that is higher (trough to peak) than all previous barriers. This restrictive condition accounts for the observation of two at most, often one, and frequently zero intermediates in kinetic experiments.

Several correlates can be noted. Suppose that folding from U to N carries cyt c down the sequential ladder of intermediate PUFs, as suggested in Figure 2. If large kinetic barriers are not encountered, the individual PUFs will not be significantly populated and detected. If a late barrier does cause an intermediate to accumulate, the intermediate detected will incorporate all the prior steps. The usual interpretation of such results has been that intermediates do not exist, or that only the single intermediate that is observed exists. Again, kinetic folding intermediates can be blocked and caused to accumulate by adventitious barriers that are due to non-inherent, off-pathway features, such as non-native proline isomers²⁴ or incorrect prosthetic group ligands.² When this occurs, the forms populated may still represent true folding intermediates even though the barriers are in a sense artifactual. For example, an adventitious barrier in cyt c causes the population of an early intermediate that strongly resembles the blue cooperative unit in Figure 1.^{25,26} However, equilibrium HX experiments with barnase do not find the intermediate populated in kinetic folding experiments.¹⁸ It may also be noted that when an intermediate is populated due to an adventitious barrier, it is not the observed intermediate that slows the folding, as has often been suggested, but rather the presence of the barrier.

Similar restrictions govern the observation of intermediates in kinetic unfolding experiments. Figure 4 shows that when a protein is jumped to unfolding conditions where the free energy levels of U and N have crossed over, the intermediates may have also crossed. In this range, hydrogen exchange will become dominated by the global unfolding reaction.²⁷ The behavior to be expected for the subglobal unfoldings is not clear; they may merge or they may maintain their individual identity. In the observable N to U reaction, the order of unfolding steps and the rate-limiting step itself may be different from the folding sequence. Whether the order of structural steps is determined by their energetic order or is encoded in the protein structure independently of the native-state energy ladder (Fig. 2) remains to be seen. Also it should be appreciated that the isotherms in Figure 4 indicate the level of free energy wells. The activation free energies of the transition state forms that determine kinetic rates

must lie somewhat higher. These are likely to respond to denaturant (m value) and temperature (activation entropy) in a manner essentially parallel to the energy wells indicated by the HX isotherms.

LOOKING FORWARD

We have briefly reviewed the new-found ability to distinguish and assign hydrogen exchange signals that depend on local, subglobal, and global unfolding and considered the requirements for studying protein substructure in equilibrium and kinetic experiments. The discovery of cooperative subglobal structural units uncovers a new dimension of protein structural behavior that may help to expose fundamental structural principles. What can the substructural PUFs tell us about the design, construction, and cooperativity of protein molecules and their biological evolution? Are the PUFs themselves plastic, with boundaries that can change with ambient conditions, or are they rigidly defined by the protein structures that they compose? Do the PUFs reveal true kinetic folding intermediates? The capability of defining and studying PUFs in proteins, within the bounds just described, appears to supply a potent means of approaching these challenging issues.

ACKNOWLEDGMENTS

This paper profited from numerous discussions with Tobin Sosnick and Leland Mayne. The work was supported by NIH research grant GM31847.

REFERENCES

- Bai, Y., Milne, J.S., Mayne, L., Englander, S.W. Protein stability parameters measured by hydrogen exchange. *Proteins* 20:4–14, 1994.
- Bai, Y., Sosnick, T., Mayne, L., Englander, S.W. Protein folding intermediates: Native-state hydrogen exchange. *Science* 269:192–197, 1995.
- Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21:167–195, 1995.
- Chan, H.S., Dill, K.A. Transition states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.* 100:9238–9257, 1994.
- Karplus, M., Sali, A. Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* 5:58–73, 1995.
- Sali, A., Shakhnovich, E., Karplus, M. How does a protein fold? *Nature* 369:248–251, 1994.
- Wolynes, P.G., Onuchic, J.N., Thirumalai, D. Navigating the folding routes. *Science* 267:1619–1620, 1995.
- Linderstrom-Lang, K.U. Deuterium exchange between peptides and water. In: "Symposium on Peptide Chemistry." *Chem. Soc. Spec. Publ.* 2:1–20, 1955.
- Hvidt, A., Nielsen, S.O. Hydrogen exchange in proteins. *Adv. Protein Chem.* 21:287–386, 1966.
- Englander, S.W., Kallenbach, N.R. Hydrogen exchange and structural dynamics of proteins and nucleic-acids. *Q. Rev. Biophys.* 16:521–655, 1984.
- Bai, Y., Milne, J.S., Mayne, L., Englander, S.W. Primary structure effects on peptide group hydrogen exchange. *Proteins* 17:75–86, 1993.
- Connelly, G.P., Bai, Y., Jeng, M.F., Mayne, L., Englander, S.W. Isotope effects in peptide group hydrogen exchange. *Proteins* 17:87–92, 1993.
- Linderstrom-Lang, K.U. Deuterium exchange and protein structure. In: "Symposium on Protein Structure." Neuberger, A. (ed). London: Methuen, 1958.

14. Englander, S.W. Measurement of structural and free energy changes in hemoglobin by hydrogen exchange methods. *Ann. N.Y. Acad. Sci.* 244:10-27, 1975.
15. Woodward, C.K., Simon, I., Tuchsén, E. Hydrogen exchange and the dynamic structure of proteins. *Mol. Cell. Biochem.* 48:135-160, 1982.
16. Loh, S.L., Prehoda, K.E., Wang, J., Markley, J.L. Hydrogen exchange in ligated and unligated staphylococcal nuclease. *Biochemistry* 32:11022-11028, 1993.
17. Mayo, S.L., Baldwin, R.L. Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science* 262:873-876, 1993.
18. Perrett, S., Clarke, J., Hounslow, A.M., Fersht, A.R. Relationship between equilibrium amide proton exchange behavior and the folding pathway of barnase. *Biochemistry* 34:9288-9298, 1995.
19. Pace, C.N. The stability of globular proteins. *CRC Crit. Rev. Biochem.* 3:1-43, 1975.
20. Schellman, J.A. The thermodynamic stability of proteins. *Annu. Rev. Biophys. Biophys. Chem.* 16:115-137, 1987.
21. Leszczynski, J.F., Rose, G.D. Loops in globular proteins: A novel category of secondary structure. *Science* 234:849-855, 1986.
22. Privalov, P.L. Stability of proteins. Small globular proteins. *Adv. Protein Chem.* 33:167-241, 1979.
23. Ptitsyn, O.B. Structures of folding intermediates. *Curr. Opin. Struct. Biol* 5:74-78, 1995.
24. Nall, B.T. Proline isomerization and protein folding. *Comments Mol. Cell. Biophys.* 3:123-143, 1985.
25. Sosnick, T.R., Mayne, L., Hiller, R., Englander, S.W. The barriers in protein folding. *Nature Struct. Biol.* 1:149-156, 1994.
26. Roder, H., Elöve, G.A., Englander, S.W. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335:700-704, 1988.
27. Kiefhaber, T., Baldwin, R.L. Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A by pulsed hydrogen exchange. *Proc. Natl. Acad. Sci. USA* 92:2657-2661, 1995.
28. Sanz, J.M., Fersht, A.R. Rationally designing the accumulation of a folding intermediate of barnase by protein engineering. *Biochemistry* 32:13584-13592, 1993.