A statistical mechanical model for hydrogen exchange in globular proteins

DAVID W. MILLER¹ AND KEN A. DILL²

¹ Graduate Group in Biophysics, and ² Department of Pharmaceutical Chemistry,
University of California at San Francisco, San Francisco, California 94143-1204

(RECEIVED April 17, 1995; ACCEPTED June 14, 1995)

Abstract

We develop a statistical mechanical theory for the mechanism of hydrogen exchange in globular proteins. Using the HP lattice model, we explore how the solvent accessibilities of chain monomers vary as proteins fluctuate from their stable native conformations. The model explains why hydrogen exchange appears to involve two mechanisms under different conditions of protein stability: (1) a “global unfolding” mechanism by which all protons exchange at a similar rate, approaching that of the denatured protein, and (2) a “stable-state” mechanism by which protons exchange at rates that can differ by many orders of magnitude. There has been some controversy about the stable-state mechanism: does exchange take place inside the protein by solvent penetration, or outside the protein by the local unfolding of a subregion? The present model indicates that the stable-state mechanism of exchange occurs through an ensemble of conformations, some of which may bear very little resemblance to the native structure. Although most fluctuations are small-amplitude motions involving solvent penetration or local unfolding, other fluctuations (the *conformational distant relatives*) can involve much larger transient excursions to completely different chain folds.

Keywords: energy landscape; HP lattice model; hydrogen exchange; protein fluctuations

Hydrogen exchange (HX) is a powerful tool in protein research, most notably in studies of structure, folding kinetics, folding intermediates, and ligand binding (Roder et al., 1988; Udgaonkar & Baldwin, 1988; Roder, 1989; Hughson et al., 1990; Englander et al., 1992). The chemical mechanism of exchange is well understood from experiments with extended, unstructured peptides highly solvated in solution (Berger & Linderstrom-Lang, 1957; Eigen, 1964). The reaction involves the acid- or base-catalyzed exchange of backbone amide-group hydrogens with deuterium or tritium from bulk solvent. Studies of poly-DL-alanine and other polypeptides have yielded a quantitative description of the effects of temperature, pH, and solvent character on the exchange rates of solvent-exposed protons (Englander & Poulsen, 1969; Englander et al., 1972; Bai et al., 1993), and rules have been developed to calculate the sequence-specific inductive effects of neighboring amino acid side chains (Molday et al., 1972; Bai et al., 1993). These studies now make it possible to accurately predict the exchange rates of amide protons in extended oligopeptides (Molday et al., 1972; Bai et al., 1993) and in solvent-exposed, unstructured regions of native proteins (Englander & Staley, 1969; Molday et al., 1972; Yee et al., 1974).

In contrast, the mechanism of exchange in *folded* proteins is complicated by the inhomogeneous structural environments of core and surface protons. Factors such as steric inaccessibility to solvent, local charge distributions, and internal hydrogen bonding, which inhibits the chemical catalysis step, can slow the rate of amide proton exchange considerably. Nevertheless, experiments show that most protons, even those buried deep within the protein, exchange at measurable rates under physiological conditions (Hvidt & Linderstrom-Lang, 1954). How do catalysts from the bulk solvent gain access to the hydrophobic core? The mechanism has been studied by measuring amide proton HX rates for whole proteins and, using NMR, for individual protons, under varying degrees of protein stability (Woodward & Rosenberg, 1971; Woodward et al., 1975; Hilton & Woodward, 1979; Woodward & Hilton, 1980). Two such experiments are shown in Figure 1. Figure 1A shows an Arrhenius plot of the HX rate for several slowly exchanging protons in the β -sheet subregion of BPTI. Figure 1B shows the exchange rate as a function of denaturant concentration for several slowly exchanging protons of cytochrome *c*. Both proteins are folded over the ranges of temperature and denaturant concentration shown. The results in Figure 1, which are typical of the slowly exchanging protons of several proteins, have led to the following conclusions regarding the mechanism of hydrogen exchange:

1. The negative slopes indicate that exchange slows exponentially as the stability of the protein is increased via the removal

Reprint requests to: Ken A. Dill, Department of Pharmaceutical Chemistry, Box 1204, University of California at San Francisco, San Francisco, California 94143-1204; e-mail: dill@maxwell.ucsf.edu.

Fig. 1. Experiments imply two exchange mechanisms. **A:** Temperature dependence of the normalized exchange-rate constants for β -sheet protons of BPTI (adapted from Kim & Woodward, 1993). **B:** Denaturant dependence of exchange-rate constants for protons of Cyt *c* (adapted from Bai et al., 1994). In B, $\Delta G = -RT \log(k_{obs}/k_0)$. Under destabilizing conditions, all rates become equal and resemble those of the denatured state. Under stabilizing conditions, curve slopes are smaller, and rates differ by orders of magnitude.

of denaturants or the lowering of temperature. This suggests that there are protons whose exchange from the native conformation would be immeasurably slow in the absence of fluctuations, implying that exchange takes place from conformations involving an opening of the native structure. The native state is said to be “exchange incompetent” for these slow-exchanging protons.

2. The different limiting behaviors in the low- and high-stability regions suggest that two different types of fluctuations are responsible for exchange. At the high temperature (or high denaturant) limit, all of the buried protons exchange at roughly the same rate, suggesting that a common type of fluctuation governs exchange when protein stability is marginal. Conversely, at low temperature (or denaturant) limits, the curves have uniformly smaller slopes, and exchange rates become strongly dependent on spatial position: within a given protein they can differ by more than eight orders of magnitude (Englander et al.,

1972). This suggests that a different type of fluctuation may be responsible for exchange under highly stabilizing conditions. Clare Woodward and her colleagues were first to interpret the biphasic curves as suggesting two distinct exchange mechanisms (the “two-process model”; see Woodward & Rosenberg, 1971).

Kinetic model of exchange and the two-process model

A simple kinetic model (Hvidt, 1964; Hvidt & Nielsen, 1966) has been helpful in analyzing HX data from folded proteins under the various conditions of stability. The model involves two components: (1) a structural factor accounting for the rate at which the protein becomes “exchange competent” by losing any structural barriers to exchange, and (2) a chemical factor accounting for the exchange rate of protons that are fully exposed and unstructured; these are known from reference data in well-characterized systems (see references above). According to this model, the exchange rate of a proton that is intrinsically incompetent to exchange from the native state is described by:

where k_{op} and k_{cl} are the rate constants for opening and closing of the native conformation, respectively, and k_0 is the rate constant for exchange from a solvent-exposed, unstructured peptide. Under steady-state conditions, the overall rate constant for Equation 1 is

$$k_{obs} = \frac{k_{op}k_0}{k_{cl} + k_0} \quad (2)$$

Exchange generally follows one of two types of kinetics, referred to as EX1 and EX2 (Hvidt & Nielsen, 1966). Under EX2 conditions, $k_{cl} \gg k_0$, meaning that exchange-competent and -incompetent structures reach a fast equilibrium prior to the slower exchange step. EX2 exchange is characterized by a linear dependence on the pH. Under EX1 conditions, $k_{cl} \ll k_0$. Here exchange depends only on k_{op} and is pH independent. Most experimentally measured rates show EX2 behavior (Segal & Harrington, 1967; Roder et al., 1985), for which the rate constant in Equation 2 reduces to:

$$k_{obs} = \frac{k_{op}}{k_{cl}} k_0 = K_{op} k_0 \quad (3)$$

where $K_{op} = k_{op}/k_{cl}$ describes the equilibrium between the native state and the exchange-competent state. Thus, the exchange rate is the product of the extended-peptide rate multiplied by the fraction of time the protein is transiently unfolded.

Measurement of the ratio of exchange rates, k_{obs}/k_0 , which is the reciprocal of the protection factor, has been used to determine k_{op} , the stability and relative population of the exchange-competent conformations, through Equation 3. Additional information is gained from the Arrhenius plot slopes, which are obtained by expressing Equation 3 in terms of energies:

$$k_{obs} = C_1 e^{-\Delta G_{op}/kT} e^{-E^\ddagger/kT} \quad (4)$$

where ΔG_{op} is the free energy of opening the native state, $E^\ddagger$ is the activation energy for exchange from a solvent-exposed,

unstructured peptide, and C_1 is a constant. The activation energy, $E^\ddagger$, of k_0 is observed to be temperature independent (Englander et al., 1979), and the free energy, $\Delta G_{op} = \Delta H_{op} - T\Delta S_{op}$, is often assumed to have temperature-independent values of ΔH_{op} and ΔS_{op} . Thus, the slope, m , of the Arrhenius plot for a given proton is obtained by differentiating $\log k_{obs}$ with respect to $1/T$:

$$m = -C_2(\Delta H_{op} + E^\ddagger) \quad (5)$$

where C_2 is assumed to be temperature independent. Hence, the slope is proportional to the enthalpy difference between the native state and the most highly populated exchange-competent structure. Comparisons of these rate-based enthalpies with experimentally measured unfolding enthalpies have been used to infer the types and magnitudes of the fluctuations responsible for exchange.

On this basis, it has been possible to provide a structural interpretation for the two different exchange mechanisms proposed by Woodward et al. to occur in the limits of high and low stability (e.g., Ellis et al., 1975; Woodward et al., 1982; Englander & Kallenbach, 1984). (1) Under destabilizing conditions, it has been proposed that exchange occurs by "global unfolding." That is, because the rate-based enthalpies coincide with equilibrium unfolding enthalpies (Woodward et al., 1975), it has been concluded that the exchange-competent states are the denatured conformations. This is consistent with the observation that all protons exchange at roughly the same rate under these conditions. (2) Under stabilizing conditions (low temperature or low denaturant), exchange rates are slower, and rate-based enthalpies are smaller than for global unfolding. This has been taken as evidence that the conformational fluctuations from the native state are correspondingly smaller, and HX sites more protected, than in denatured states. In this limit, protection is highly site specific, and HX rates vary with the amide proton environment. This second mechanism of exchange has alternately been called "exchange from the folded state" (Woodward & Rosenberg, 1971), the "low activation energy" process (Woodward et al., 1975), and "process b" (Hilton & Woodward, 1979). We refer to it as the "stable-state" mechanism.

Models for the stable-state mechanism

Two types of structural models have been developed to describe the stable-state mechanism: "solvent penetration" models (e.g., Woodward & Rosenberg, 1971; Lumry & Rosenberg, 1975; Richards, 1979) and the "local unfolding" model (Hvidt & Nielsen, 1966; Englander, 1975). The models are reviewed fully in Englander and Englander (1978), Gurd and Rothgeb (1979), Hilton et al. (1981), Woodward et al. (1982), and Englander and Kallenbach (1984), and are discussed briefly below.

Solvent penetration models

In penetration models, protons exchange within the protein interior: catalysts enter the protein core through transiently formed channels and cavities. One penetration model proposes that such channels arise from small, rapid fluctuations of tenths to several angstroms of interior atoms. Other penetration models have proposed that the channels arise from the redistribution of interior hydrogen bonds (Nakanishi et al., 1973; Lumry

& Rosenberg, 1975), or from the random association of preexisting interior cavities (Richards, 1979). In penetration models, the exchange rate of a proton depends both on its average accessibility to solvent, which is a function of its depth of burial and the local mobility of the protein in its vicinity, and on its reactivity with the catalyst, which is affected by local structural features such as packing density and hydrogen bonding. Some penetration models postulate that exchange is highly localized; i.e., protons on adjacent amino acids exchange independently rather than in a correlated way.

Local unfolding model

In the local unfolding model, exchange occurs outside the protein when a subregion, such as an α -helix, transiently unfolds into the bulk solvent. Protein segments of no more than about 10 residues are expected to unfold at one time (the unfolding region is "local"), consistent with the small enthalpies observed for exchange under stable conditions. In this model, the main barrier to exchange is considered to be the interior hydrogen bonding of amide protons, rather than their depth of burial. Adjacent protons within subregions are predicted to exchange at roughly the same rate, because protons belonging to the same structural unit are unfolded simultaneously. Such correlated exchange behavior has sometimes been called "cooperativity."

Experimental evidence

Evidence that has been taken as support for solvent penetration models includes the following:

1. The magnitudes of fluctuations inferred from X-ray crystallography and NMR spectroscopy under native conditions are thought to be large enough to form the solvent channels envisioned by the penetration model (Wüthrich & Wagner, 1978; Frauenfelder et al., 1979).

2. Molecular dynamics (MD) calculations suggest that small, picosecond atomic fluctuations provide penetration pathways for O_2 in myoglobin (Karplus & McCammon, 1981).

3. Proton NMR studies in BPTI show that exchange of water molecules between the protein core and bulk solvent can occur on a subsecond time scale under physiological conditions (Tuchsen et al., 1987; Otting et al., 1991).

4. The pK_a 's of two histidine residues in ribonuclease S during C-2 proton exchange are more characteristic of a buried than exposed environment (Bradbury et al., 1977, 1980).

5. Denaturants at low concentrations do not accelerate stable-mechanism HX rates, implying that regional unfolding is not required for exchange (Woodward et al., 1975; Woodward & Hilton, 1979; Hilton et al., 1981; Mayo & Baldwin, 1993).

6. Roughly 95% of amide protons in myoglobin crystals exchange with deuterium after perfusion with deuterium solvent (Schoenborn et al., 1978), suggesting that exchange does not require unfolding.

7. Exchange kinetics in lysozyme are identical in solution and in the crystal, where it is presumed that local unfolding should not occur (Tuchsen & Ottesen, 1979; Tuchsen et al., 1980).

8. Protons at the interface between trypsin and trypsin inhibitor exchange without dimer dissociation, implying an absence of local unfolding, and the pH and temperature dependence of exchange are characteristic of buried protons in globular proteins (Pershina & Hvidt, 1974; Woodward, 1977).

9. For the slowest exchanging amides in BPTI, next-neighbor residues can have very different exchange rates, implying a lack of correlation (Hilton & Woodward, 1979).

10. For the 15 slow-exchanging amides in BPTI, there is no obvious correlation between H-bond length and the rates of exchange by the stable-state mechanism (Levitt, 1981a; Woodward et al., 1982).

11. In lysozyme (Radford et al., 1992; Pedersen et al., 1993) and leucine zipper peptides (Goodman & Kim, 1991), there are helices in which buried protons exchange more slowly than solvent-exposed protons.

12. Studies of the effects of local structure on tyrosine ring flips in folded proteins suggest that the formation of solvent channels, as envisioned by the penetration model, could give rise to the large range of enthalpies observed for the stable-state mechanism of exchange (Karplus & McCammon, 1981).

Evidence taken as support for the local unfolding model includes the following:

1. In support of neighbor correlations, like-exchanging sets of protons in hemoglobin undergo uniform rate shifts upon the allosteric structural change from deoxy-Hb to oxy-Hb (Englander & Mauel, 1972; Englander, 1975) and these sets are localized within regions of secondary structure (Englander et al., 1983).

2. As an indication that HX rates depend not only on the degree of burial, but also on hydrogen bonding, small molecules that form intramolecular hydrogen bonds, such as salicylate, exchange up to six orders of magnitude slower than their diffusion-limited rates, even though they have otherwise good solvent accessibility (Haslam & Eyring, 1967; Rose & Stuehr, 1968).

3. For intermediate and fast (non- β -sheet) protons in BPTI, trypsin, and ribonuclease, there is no correlation between crystallographic B factors and exchange rates, suggesting that small internal fluctuations are not the primary mechanism of exchange (Levitt, 1981b; Kossiakoff, 1982; Wlodawer & Sjolín, 1982).

4. There is no obvious correlation between exchange rate and depth of burial in trypsin and ribonuclease (Kossiakoff, 1982; Wlodawer & Sjolín, 1982).

5. Englander and Kallenbach (1984) have argued that the free energies would be too high for the catalytic OH^- ions to penetrate a protein core.

6. The observed shift in pH_{min} for exchange by the stable-state mechanism is argued to be caused by the separation of interior charged groups, which requires a structural change greater in magnitude than the small fluctuations envisioned by the penetration model (Englander & Kallenbach, 1984).

7. In α -helical segments of ribonuclease S (Kuwajima & Baldwin, 1983), BPTI (Wagner & Wüthrich, 1982), and apamin (Wemmer & Kallenbach, 1983), amides exchange at similar rates despite very different solvent accessibilities in the native protein.

8. Amides in α -helical and β -sheet regions of both synthetic polypeptides and native proteins exchange slowly despite being solvent accessible (Leichtling & Klotz, 1966; Ikegami & Kono, 1967; Nakanishi et al., 1972; Welch & Fasman, 1974; Rohl et al., 1992; Bai et al., 1994).

Both penetration and local unfolding models view the exchange-competent conformations as being small excursions from the native state because the rate-based enthalpies are smaller than those for unfolding the protein. The main distinctions between the two mechanisms are: (1) whether the exchange reaction is external in bulk solvent (local unfolding model) or

internal within the protein interior (solvent penetration model), and (2) whether exchange involves neighbor correlations or whether neighboring residues can exchange independently. Neither model attempts to make quantitative predictions for individual proton exchange rates.

Our aim here is to avoid making assumptions about the various mechanisms of exchange and to see instead if a unified viewpoint can be derived from an underlying statistical mechanical theory of how amino acids develop access to solvent through conformational fluctuations from native states. Using a simple lattice representation of proteins, we model the nonchemical part of the hydrogen exchange process and explore the mechanism by which buried monomers gain access to solvent under varying conditions of chain stability. Our model involves two components: a general statistical mechanical treatment of monomer accessibility, and a simple lattice model to relate structure to accessibility. Although the lattice representation of structure is very simple, our statistical mechanical treatment is more complete and has fewer approximations than previous models. We make no assumptions about mechanisms of fluctuations, about whether the equilibrium is two state or three state, about the rates of exchange from compact or unfolded conformations, or about how different conformations are affected by external conditions. Rather, these properties are derived from the theory.

We find that the two limiting mechanisms of exchange, under stable and unstable conditions respectively, arise as a natural consequence of different limiting behaviors of native fluctuations. Exchange by the stable-state mechanism is found to resemble both solvent penetration and local unfolding models in some respects, but there are also novel aspects, namely that: (1) fluctuations are usually ensembles of many different conformations, and (2) the exchange-competent conformations under stable conditions can occasionally involve large conformational deviations from the native structure. We explain these results in terms of the shape of the folding energy landscape and discuss possible implications for biological proteins.

The lattice model

We model proteins using the two-dimensional HP lattice model (Lau & Dill, 1989, 1990; Chan & Dill, 1991; Dill et al., 1995). A protein is represented by a sequence of H (hydrophobic) and P (other) monomers configured on a two-dimensional lattice. Amino acids are represented by beads, and the protein backbone by lines connecting the beads. The background lattice serves to divide space into amino acid-sized units. Lattice sites may be either empty or filled by a single bead (the excluded-volume constraint), and empty lattice sites are assumed to contain a solvent molecule. In the HP model, each HH contact, formed when two nonsequential H monomers occupy adjacent lattice sites, is favored by a contact free energy ϵ ($\epsilon < 0$). This energy is a simple way of capturing the importance of hydrophobic interactions, whatever their underlying atomic origins in hydrogen bonding, van der Waals interactions, and solvent ordering (Dill, 1990; Chan & Dill, 1991). HP and PP contacts are neither favored nor disfavored energetically. Hence the free energy of a conformation is $h\epsilon$, where h is the number of HH contacts. The conformation in Figure 2 has 8 HH contacts, and therefore has a free energy of 8ϵ . The magnitude of ϵ determines the degree of stability imparted by external conditions; large and negative ϵ is more stabilizing. Conformational entropy, the driv-

Fig. 2. Sample conformation for the two-dimensional HP lattice model. Black and white beads represent H (hydrophobic) and P (polar) monomers, respectively. Noncovalent HH contacts are favored by a free energy, ϵ ($\epsilon < 0$). This conformation has 8 HH contacts, so the free energy is 8ϵ .

ing force for unfolding, enters the model through the exhaustive enumeration of all the possible chain configurations (see discussion below).

The disadvantages of the model are clear: atomic resolution is lost; conformations are restricted to a lattice; it is in two dimensions; the energy function is simplified; and chains are unrealistically short. Yet, despite these disadvantages, the model has been found useful for modeling protein properties (Chan & Dill, 1989, 1990, 1994; Lau & Dill, 1989, 1990; Shortle et al., 1991; Dill et al., 1995) because it shows several protein-like features, including cooperative collapse, a nonpolar core, secondary and tertiary structures, and multistage folding kinetics. And because it characterizes the full conformational space in an exact and unbiased way, we are able to consider all fluctuations of all sizes, without being restricted to small-amplitude motions, short time scales, or mathematical approximations. Most importantly, we believe the model captures the main physics of protein folding—the hydrophobic interactions, conformational freedom of the chain, and the steric restrictions imposed by excluded volume.

This model does not include explicit hydrogen bonding, but we believe this is not a serious limitation. There are three possible roles of hydrogen bonding in hydrogen exchange. First, hydrogen bonding is required to catalyze the exchange reaction, but as in past literature on exchange mechanisms, we calculate ratios (observed rates divided by rates for solvent-exposed, unstructured sites) and thus avoid the need to consider the chemical details of the exchange process. Second, to the extent that hydrogen bonding contributes to protein stability, it is an implicit component of the HH, HP, and PP interactions of the HP model. Because the model produces hydrophobic cores, and monomer burial, we consider it a simplest model of how the accessibilities of amino acids in proteins depend on stability. Third, interresidue hydrogen bonds present a structural barrier to exchange because they must be broken before the exchange process can be initiated. The breaking of these bonds can also give rise to correlated exchange, as in the local unfolding model. In the present model the proximity of an amino acid to solvent determines its ability to exchange, as described below. In this regard, all microscopic barriers to exchange, including hydrogen

bonding, are treated as a single factor implicitly accounted for by monomer burial. However, neighbor correlations can occur naturally in a sequence-specific way without introducing the additional energy parameter a hydrogen bond would require.

We study HP sequences of 16 monomers. For any 16-mer chain there are exactly 802,075 possible conformations that can be configured on a two-dimensional lattice. These conformations are generated by computer, and each is weighted by a Boltzmann factor according to the number of HH contacts made. Figure 3 shows an energy diagram for a sample sequence. The native structure (ground state) is the conformation with the largest possible number of HH contacts, and thus the lowest contact free energy. We study only those sequences (“nondegenerate”) that have a single native conformation, because we believe that they best represent biological proteins, which fold to unique structures. All higher-energy conformations comprise the “denatured” states, and are grouped by energy into “first-excited” states, “second-excited” states, etc., corresponding to successively fewer HH contacts (see Fig. 3). Figure 3C shows the density of states, $g(h)$, the distribution of conformations having h HH contacts. For all HP sequences, $g(h)$ diminishes rapidly as h increases. That is, there are far more open, high-energy conformations than compact, low-energy conformations.

Our treatment of hydrogen exchange is divided into two parts. First we develop a statistical mechanical theory to describe how monomer sites that are exchange incompetent in the native state become competent to exchange through Boltzmann populations of non-native conformations. This theory is completely general and does not depend on lattices or on any other particular model. Second, we introduce the HP model to give a concrete basis for relating protein structure and stability to exchange competence.

Statistical mechanical model of exchange competence

At a temperature T , any conformation, c , of a protein will be populated according to its Boltzmann probability, P_c :

$$P_c = \frac{e^{-E_c/kT}}{\sum_{c=1}^N e^{-E_c/kT}} \quad (6)$$

where E_c is the free energy of conformation c , k is Boltzmann’s constant, and N is the number of conformations available to the protein. Equation 6 is a general, model-independent equation for any equilibrium system. At low temperatures, only the native conformation will be populated because its Boltzmann factor will be dominant. At high temperatures, the denatured conformations will be populated because there are so many more non-native than native conformations, and the entropically favored states will dominate.

The average value of any observable conformational property A_c can be computed as a sum over Boltzmann probabilities:

$$\langle A \rangle = \sum_{c=1}^N A_c P_c = \frac{\sum_{c=1}^N A_c e^{-E_c/kT}}{\sum_{c=1}^N e^{-E_c/kT}} \quad (7)$$

Fig. 3. A,B: The “energy ladder” diagram for the 16-mer HP sequence in Figure 2. The native conformation is at the bottom of the ladder; this single lowest energy conformation has 8 HH contacts. Increasing steps up the energy ladder represent successively fewer HH contacts; these levels are the first-excited states, second-excited, etc. Sample conformations of each energy level are shown in A. C: Density of states distribution, $g(h)$, is plotted against the number of conformations of this HP sequence that have the given number, h , of HH contacts.

In particular, the observable property of interest here is the exchange competence of an individual proton site. If $A_{c,i}$ characterizes the exchange competence of proton i in conformation c , Equation 7 gives the average exchange competence of that proton as a function of temperature.

Equation 7 is completely general and model independent. Now we introduce the HP model. The conformational energy depends only on the number of HH contacts: $E_c = h\epsilon$, where h is the number of HH contacts in conformation c , and ϵ is the HH contact free energy. Written in terms of h , Equation 7 becomes (see Appendix)

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_N} A_{h,i} g(h) e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h) e^{-h\epsilon/kT}} \quad (8)$$

where h_N is the number of HH contacts in the native conformation, $g(h)$ is the density of states, and $A_{h,i}$ is the average value of $A_{c,i}$ for the $g(h)$ conformations having h HH contacts.

In the present model, the exchange competence of a monomer is given by its accessibility to solvent: we define $A_{c,i} = 0$ if monomer i is completely surrounded by four other monomers

in conformation c , and $A_{c,i} = 1$ if monomer i is adjacent to a lattice site containing a solvent molecule. The model is illustrated in Figure 4, which shows the native and three first-excited states of the sequence in Figure 3. The five monomers buried in the native state are exchange incompetent, and four of them become competent to exchange in at least one of the first-excited states. In the present model, buried monomers do not exchange, while those adjacent to a solvent-occupied lattice site exchange at a generalized rate, k_0 , corresponding to a solvated, unstructured peptide. Therefore the exchange rate of any monomer, i , is

$$k_i = \langle A \rangle_i k_0 \quad (9)$$

where k_0 is the extended-peptide exchange rate for monomer i . Equation 9, which can be determined by exact enumeration for monomers of any short HP sequence, is analogous to Equation 3 of the Hvidt model for EX2-type exchange and gives the exchange rate in terms of average monomer solvent accessibility.

Because $\langle A \rangle_i$ of Equation 8 depends on ϵ , monomer exchange rates can be determined under various conditions of chain stability. We use the single parameter $|\epsilon/kT|$ as a measure of the degree of stabilization due to external conditions. When $|\epsilon/kT|$ is large (corresponding to strongly stabilizing conditions such as low temperatures or denaturants), only the na-

Fig. 4. Native and fluctuational conformations. **A:** Native state of the sequence in Figure 3. Monomers 8, 11, 14, 15, and 16 are each surrounded by four other monomers and so are "buried," and hence exchange incompetent, in the native state. **B,C,D:** The three first-excited states. Monomers 8, 11, 14, and 16 become exchange competent in the first-excited states because each of those monomers makes a contact with a solvent lattice site in at least one of the three conformations. Monomer 15 cannot exchange in any of the three first-excited states but can exchange in the second-excited states (not shown). Boxed regions represent the smallest section of chain that must reconfigure to reach the native state. Dissimilarities (see text) between the first-excited states in B, C, and D and the native state are 0.275, 0.274, and 0.219, respectively.

tive state is populated to any appreciable degree, so the accessibilities of monomers, and their corresponding exchange rates, can be determined directly from the native structure. In the other limit, when $|\epsilon/kT|$ is small (high temperature or denaturants), most molecules are denatured, and the average solvent accessibility and exchange rate of a monomer will reflect the distribution of accessibilities in the ensemble of denatured conformations. Thus, for a monomer buried in the native state, average solvent accessibility and the exchange rate increase as $|\epsilon/kT|$ becomes smaller, because the Boltzmann populations of non-native exchange-competent states increase.

Below we show results of calculations of the ratio $k_i/k_{0,i}$, the reciprocal of the protection factor, as a function of $|\epsilon/kT|$, for monomers of the sample 16-mer sequence shown in Figures 3 and 4.

Results

Model exchange curves

Figure 5 shows the predicted exchange rate, $k_i/k_{0,i}$, versus $|\epsilon/kT|$ for the five monomers buried in the native conformation of Figure 4A. These five monomers are exchange incompetent in the native state, so this calculation corresponds to the experiments shown in Figure 1. Because the parameter $|\epsilon/kT|$ is equivalent to $1/T$ for fixed ϵ , the plot in Figure 5A is analogous to Arrhenius plots of the type shown in Figure 1A. Figure 5B shows the same curves with both axes reversed, so that it corresponds to the denaturant plot of Figure 1B. The theoretical results in Figure 5 show the same general features as the experimental results in Figure 1, with exchange rates that converge to a single value under destabilizing conditions and diverge under stabilizing conditions, and a transition region sugges-

Fig. 5. HX curves computed from the model versus HH-contact energy, ϵ , for the five buried monomers of Figure 4A. Curves correspond to monomers 8 (∇), 11 (Δ), 14 ($\diamond$), 15 ($\square$), and 16 ($\times$). **A:** Analog of the Arrhenius plots of Figure 1A. **B:** Analog of the denaturant curves of Figure 1B. Both figures show converging exchange rates under destabilizing conditions, diverging rates under stable conditions, and a transition region suggesting two exchange mechanisms.

tive of the two limiting exchange mechanisms proposed by Woodward.

The basis for these results is illustrated in Figure 6, which shows how the populations of different conformations of the 16-mer change with temperature. Under destabilizing conditions (Fig. 6A,B), the chain populates an ensemble of conformations having large energies and correspondingly unfolded structures. Exchange under these conditions occurs from the denatured states, i.e., by global unfolding. Because solvent exposure is high in these unfolded conformations, exchange rates are similar and converge toward the maximum rate as native stability is decreased. Conversely, under stabilizing conditions (Fig. 6C,D), only the lowest-energy excited states contribute to the fluctuations. Exchange of buried monomers therefore occurs from a restricted set of conformations with native-like energy and compactness. For monomers 8, 11, 14, and 16, stable-state exchange is from the first-excited states, because exchange competence is achieved in these conformations (see Fig. 4). For monomer 15, which is buried in all three first-excited states, exchange first occurs only when the chain fluctuates to the second-excited states.

The two apparent exchange mechanisms in Figures 1 and 5 are explained as a shift in the class of conformations from which exchange occurs: from the broad set of unfolded conformations to a much smaller set of highly compact conformations. This result is consistent with the "global unfolding" and "stable state" mechanisms of Woodward's two-process model and supports the belief that the shift in mechanisms results from changes in protein structure rather than from complex exchange chemistry or a shift from EX2- to EX1-type exchange. The slopes of the curves are obtained from Equation 8 (see Appendix) and are

proportional to the energy required for a fluctuation from the native state to the most highly populated exchange-competent conformation: $m(\epsilon)_i = \langle h \rangle_{acc,i} - h_N$, where $\langle h \rangle_{acc,i}$ is the weighted average of h for all conformations in which monomer i is solvent accessible. Under destabilizing conditions, the most highly populated exchange-competent conformations are unfolded and are the $h = 6$ conformations (Fig. 6A) for the sample 16-mer of Figure 4. Hence $\langle h \rangle_{acc,i} = 6$ in the limit of small values of $|\epsilon/kT|$, and the limiting slope is -2 . Under these conditions, monomer exchange rates are similar and approach the maximum value, $k_{0,i}$, as in the global unfolding model. Conversely, under stable conditions only the exchange conformations of lowest energy contribute to exchange, so $\langle h \rangle_{acc,i}$ is equal to h_i^* , the number of HH contacts in the lowest-energy exchange conformations of monomer i (see Appendix). For monomers 8, 10, 14, and 16, which can exchange from the first-excited states (Fig. 4), $h_i^* = 7$, and the limiting slope for large $|\epsilon/kT|$ is -1 . For monomer 15, $h_i^* = 6$, so the limiting slope is -2 . The diversity of monomer exchange rates in this stable limit arises from differences in the energy required for the minimum fluctuation to an exchange-competent structure. This result is evident from Equation 8 in the limit of large $|\epsilon/kT|$, where the average solvent accessibility becomes:

$$\langle A \rangle_i \rightarrow A_{h_i^*,i} \frac{g(h_i^*)}{g(h_N)} e^{-(h_i^* - h_N)\epsilon/kT}. \quad (10)$$

Equation 10 is a single exponential, so the exchange curve of a given monomer is a straight line with slope $m(\epsilon)_i = h_i^* - h_N$.

Fig. 6. Populations of conformations. Populations at $|\epsilon/kT| =$ (A) 2.65, (B) 4.00, (C) 5.00, and (D) 8.00, respectively. The melting temperature, where the native state is 50% populated, is $|\epsilon/kT| = 2.65$.

Monomer 15 requires greater fluctuational energy to exchange (2ϵ versus ϵ), so as stability is increased, its exchange rate decreases more rapidly than the other buried monomers.

The approximate linearity of the curves in the two regions arises from the temperature independence of $\langle h \rangle_{acc,i}$ in the limits of large and small $|\epsilon/kT|$. For large $|\epsilon/kT|$, this result follows directly from Equation 10. For small $|\epsilon/kT|$, conformations from many energy levels contribute to exchange, so a constant slope is not necessarily expected. Whether or not conformations of a single energy level will dominate exchange, as is the case with the sequence shown here, depends on the density of states, $g(h)$, which is highly sequence specific. The density of states also affects the curve shapes.

Although the model denaturant curves in Figure 5B do not become as flat as the experimental curves of Figure 1B, this is undoubtedly due to the relatively large size of the energy spacing in our short-chain model compared to real proteins, which have many more monomer contacts. Short chains also restrict the possible range of stable-limit HX rates among monomers. The high rate diversity observed in experiments (Fig. 1A,B) would be obtained in our model if more energy levels were available, as is the case for real proteins.

Structures of exchange-competent states

Figure 4B, C, and D shows the three first-excited states of the given 16-mer sequence. Under stable conditions, these are the predominant exchange-competent conformations for four of the five buried monomers. The structural fluctuations from the native state represented by these conformations are consistent with both solvent penetration and local unfolding models. In Figure 4B for example, monomer 8 has become solvent accessible by a small shift in its local subregion (monomers 7–13) that has forced monomer 8 out into the solvent, much like a local unfolding event. In Figure 4D, monomers 14 and 16 have both become accessible by the movement of a sequentially distant chain region (monomers 1–5), resulting in the exposure of both monomers to bulk solvent. This latter example could be considered consistent with solvent penetration, because monomers 14 and 16 are part of an intact core of residues that become exposed by removal of some chain surface. In our model examples, the chains are too short to have enough “core” for us to interpret either penetration or local unfolding unambiguously. Nevertheless, from this and other examples, we believe this model supports the general types of protein motions characteristic of both solvent penetration and local unfolding mechanisms. As for neighbor correlations, the sequence in Figure 3 shows independent, rather than correlated exchange; monomers 14, 15, and 16 form a buried subregion in the native conformations, but monomer 15 exchanges at a much slower rate. In general, both correlated and independent exchange are possible depending on the monomer sequence.

“Conformational distant relatives”

How different are the hydrogen-exchanging conformations compared to the native state in the stable limit? To answer this question, we must explore the relatedness of native and first-excited conformations. Figure 4B, C, and D shows examples of relatively small fluctuations in which the native chain fold and hydrophobic core are largely preserved in the first-excited con-

formations. At the other extreme, Figure 7 shows an example of a native state (Fig. 7A) and one of its first-excited states (Fig. 7B) that are completely different, sharing no common residue contacts. This is an example of a fluctuation in which a very small change in energy corresponds to a very large change in structure.

Are such large structural fluctuations rare or common? Figure 8A shows the average magnitude of structural change under stable-state conditions, obtained by comparing all possible 16-mer native states to their first-excited states and measuring structural differences between them. We use the dissimilarity measure of Yee and Dill (1993), which scores the conformational relatedness of two structures: 0 means two structures are identical. Figure 8A shows that fluctuations from native to first-excited states display a wide range of magnitudes. For comparison, Figure 8B shows a similar distribution in which each of the possible 16-mer native states is compared to each other native state rather than to its first-excited states. This second distribution involves unrelated conformations and so should represent approximately the complete range of dissimilarity that is possible for compact 16-mer structures. The significant overlap of the two distributions shows that although most first-excited states represent small structural excursions from the native state, occasional low-energy fluctuations result in conformations that show near-maximal dissimilarity from the native. We call such states “conformational distant relatives” (CDR) of the native

Fig. 7. Example of conformational distant relatives. **A:** Sample 16-mer native state ($h = 6$). **B:** One of its first-excited states. No native contacts are conserved. The dissimilarity is 0.809.

Fig. 8. Dissimilarity distributions. **A:** Comparison of each nondegenerate 16-mer native state with each of its first-excited states. **B:** Pairwise comparison of all nondegenerate 16-mer native states. Horizontal axis shows the dissimilarity measure of Yee and Dill (1993). The average dissimilarities for A and B are 0.50 and 0.65, respectively. This figure shows that first-excited states can have a broad distribution of structures, including very non-native ones.

structure, being only slightly higher in free energy, but very different in conformation. Although the percentage of CDRs among first-excited states is small and highly sequence dependent (data not shown), we find that their relative population increases with the number of first-excited states, $g(h_N - 1)$ (Fig. 9). That is, proteins that have a greater number of first-excited states also have more CDRs. This result is to be expected, because typically only a few first-excited states can be configured by small-amplitude fluctuations of the native state.

Figure 10 shows an unexpected prediction from the model. In the studies described above, we have considered protons that are exchange incompetent in the native state: for example, they may be interior protons buried within the core, or surface protons that are involved in amide hydrogen bonds. As conditions become increasingly denaturing, such protons become increasingly competent to exchange, and their HX rates increase. Now what happens to protons that are exchange competent in the native state, such as those in exposed loop regions where hydrogen bonding is absent? In our lattice model of HX, such protons

Fig. 9. Average dissimilarities between native and their first-excited states for all nondegenerate 16-mers versus the number of first-excited states, $g(h_N - 1)$. Fluctuations are more distant from the native structure for HP sequences that have higher numbers of first-excited conformations.

are represented by the monomers that are solvent accessible in the native state. Figure 10 indicates the interesting possibility that although such protons should exchange at the maximum rate under both stabilizing and destabilizing conditions, under inter-

Fig. 10. HX rates for monomers that are solvent accessible in the native state of Figure 4A. Curves are for monomers 2 ($\square$), 5 ($\times$), 6 ($\diamond$), 7 (∇), and 12 ($\triangle$). As $|\epsilon/kT|$ decreases toward the denaturation midpoint, HX rates first decrease due to burial in compact denatured conformations and then increase as the chain becomes unfolded.

mediate conditions of stability they may exchange at a slightly *slower* rate. This is because surface protons that are locked into an exposed and unstructured position in the native state can become buried (or H-bonded) in the compact non-native states that are increasingly populated as native stability is slightly decreased. As stability is decreased further, all proton sites eventually become fully exposed and unstructured, and exchange rates increase toward the maximum. Hence, just as small amounts of denaturing agents tend to expose hydrophobic residues, they may also tend to bury polar monomers. Figure 10 shows that such effects correspond only to small rate differences, perhaps accounting for why they have not been previously reported, as far as we know.

Discussion

One of the main issues addressed in the present study concerns the magnitude of protein fluctuations. Two common assumptions, implicit in both solvent penetration and local unfolding models are that: (1) such fluctuations must, under stable conditions, involve only small changes in energy relative to the native state, and (2) conformations similar in energy to the native state must also be similar in structure. We agree with (1), but the present study suggests that (2), although true most of the time, may not be required. The fluctuations observed in our model can involve small changes in energy relative to the native state but large changes in conformation.

Are conformational distant relatives important for biological proteins? Two fundamental issues are: (1) whether low-energy states can be structurally dissimilar to the native state, and (2) whether such non-native structures, if they exist, can be visited on the time scales relevant to hydrogen exchange. We contend that (1) is not an artifact of the lattice model, but follows from any model based on the idea that proteins have "rugged" energy landscapes (Bryngelson & Wolynes, 1989; Frauenfelder et al., 1991; Bryngelson et al., 1995; Dill et al., 1995; Wolynes et al., 1995), i.e., that very non-native structures can have very low energies (Fig. 11A). The HP lattice model has such ruggedness. At the other extreme, protein folding might be guided by a smooth "funnel-like" energy landscape (Fig. 11B), whereby conformational dissimilarity with the native state implies a large difference in free energy. In that case, fluctuations involving small differences in energy relative to the native state must involve small conformational changes. This is the implicit assumption in existing local unfolding and solvent penetration models.

There is considerable experimental evidence that some proteins have energy landscapes that involve at least some degree of ruggedness. Some proteins fold on time scales of seconds and longer, rather than the much shorter time scales (perhaps nanoseconds to microseconds) that might be expected for smooth funnels. Both prion protein PrPC (Pan et al., 1993; Cohen et al., 1994; Huang et al., 1994) and α -lytic protease (Baker & Agard, 1994) have been found to exist in both a native and highly non-native conformation under physiological conditions, separated by a large kinetic barrier. Computer threading experiments (Novotny et al., 1984) have found that a single amino acid sequence can be constrained to form two entirely different structures (one α -helix and the other β -sheet) with negligible differences in the *in vacuo* CHARMM free energy.

Fig. 11. Two possible landscapes for protein folding. The y-axis is conformational free energy, and x is one "reaction coordinate" for folding. N denotes the native state. **A:** On rugged landscapes, there are multiple minima in free energy, so a small energy difference can correspond to a large conformational change. This is just a schematic to emphasize this point and is not meant to imply that helical and sheet structures interconvert. **B:** On funnel-like landscapes, large distances from the native state imply a large increase in free energy.

Although protein landscapes are not fully understood, they can be studied using kinetic refolding experiments, because it is the shape of the landscape that determines folding rates under native conditions. On this basis, evidence suggests that small, single-domain proteins, which typically show relatively simple kinetics, may have landscapes that look more like funnels. Proteins with more complex folding kinetics, such as large, multi-domain proteins, probably have more rugged landscapes. It is not clear whether any biological protein has as much ruggedness as the HP lattice model. However, if the amount of ruggedness is significant (i.e., if there are multiple deep minima in free energy), then it follows, irrespective of the model used to describe it, that some fluctuations may bear little resemblance to the native structure.

The second issue is kinetic accessibility. Can the native structure of a protein fluctuate to very non-native low-energy conformations on the time scale of HX? The time scales of experiments dictate the heights of kinetic barriers that can be crossed. The large range of time scales observed for EX2 exchange—picoseconds to several months at low pH—suggests that some non-native conformations could be kinetically accessible, although the evidence is currently limited.

How can we determine if native protein fluctuations involve distant relatives? Such fluctuations are probably very difficult

to observe experimentally because the native structure is so dominant under stabilizing conditions, and the denatured conformations are so dominant under unfolding conditions. Moreover, these conformations need not be on the kinetic folding pathway and so may be undetectable in refolding experiments. HX in the high-stability limit can detect native fluctuations, but as we noted before, slopes of Arrhenius plots tell us only that the energy changes (from the native state) are small; they cannot tell us how big the structural changes are. Native fluctuations can also be studied by NMR or by X-ray crystallographic *B* factors, but, in both cases, the measured fluctuations involve averages over multiple non-native states, so no one such conformation can be studied in isolation.

The present model appears consistent with data cited as support for local unfolding and solvent penetration models because each of these models describes a subset of the fluctuations indicated by the present model. A main point of our study is that local unfolding and solvent penetration may not be mutually exclusive: the native structures in our model show fluctuations of both types. At the same time, we find that structural fluctuations may sometimes be to conformations that differ considerably from the native state. Our model also predicts that knowledge of the native structure is not sufficient in itself to predict monomer exchange rates; these rates are strongly affected by the structures of the first-excited states, which cannot be predicted in a simple way from knowledge of the native structure. This result seems consistent with previously reported difficulties in correlating HX rates with features of native structures.

An interesting question raised by these results is whether MD calculations, which sample only small excursions from native structures, could detect CDR conformations.

Acknowledgments

The idea to do these calculations arose from conversations with Clare Woodward and Sarina Bromberg, to whom we are most grateful. We thank Robert L. Baldwin, Walter Englander, Tobin Sosnick, and Clare Woodward for teaching us about HX and for many very helpful discussions; Karen Tang for comments on the manuscript; and NIH for financial support. D.W. Miller is supported by a National Science Foundation Fellowship GER-9255688 (1994–1995).

References

- Bai Y, Milne JS, Mayne L, Englander SW. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct Funct Genet* 17:75–86.
- Bai Y, Milne JS, Mayne L, Englander SW. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins Struct Funct Genet* 20:4–14.
- Baker D, Agard DA. 1994. Kinetics versus thermodynamics in protein folding. *Biochemistry* 33:7505–7509.
- Berger A, Linderstrom-Lang K. 1957. Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* 69:106–118.
- Bradbury JH, Chapman B, Crompton MW, Norton R, Teh J. 1980. Hydrogen-deuterium exchange of the C-2 protons of histidine and histidine peptides and proteins. *J Chem Soc Perkin Trans* 2:693–699.
- Bradbury JH, Crompton MW, Teh JS. 1977. Nuclear-magnetic-resonance study of the histidine residues of S-peptide and S-protein and kinetics of 1H–2H exchange of ribonuclease A. *Eur J Biochem* 81:411–422.
- Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct Funct Genet* 21:167–195.
- Bryngelson JD, Wolynes PG. 1989. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J Phys Chem* 93:6902–6915.
- Chan HS, Dill KA. 1989. Compact polymers. *Macromolecules* 22:4559–4573.
- Chan HS, Dill KA. 1990. Origins of structure in globular proteins. *Proc Natl Acad Sci USA* 87:6388–6392.
- Chan HS, Dill KA. 1991. “Sequence space soup” of proteins and copolymers. *J Chem Phys* 95:3775–3787.
- Chan HS, Dill KA. 1994. Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys* 100:9238–9257.
- Cohen FE, Pan KM, Huang Z, Baldwin M, Fletterick RJ, Prusiner SB. 1994. Structural clues to prion replication. *Science* 264:530–531.
- Dill KA. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
- Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding. A perspective from simple exact models. *Protein Sci* 4:561–602.
- Eigen M. 1964. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. *Angew Chem Int Ed Engl* 3:1–19.
- Ellis LM, Bloomfield VA, Woodward CK. 1975. Hydrogen-tritium exchange kinetics of soybean trypsin inhibitor. Solvent accessibility in the folded conformation. *Biochemistry* 14:3413–3419.
- Englander JJ, Calhoun DB, Englander SW. 1979. Measurement and calibration of peptide group hydrogen-deuterium exchange by ultraviolet spectrophotometry. *Anal Biochem* 92:517–524.
- Englander JJ, Rogero JR, Englander SW. 1983. Identification of an allosterically sensitive unfolding unit in hemoglobin. *J Mol Biol* 169:325–344.
- Englander SW. 1975. Measurement of structural and free energy changes in hemoglobin by hydrogen exchange methods. *Ann NY Acad Sci* 244:10–27.
- Englander SW, Downer NW, Teitelbaum H. 1972. Hydrogen exchange. *Annu Rev Biochem* 41:903–924.
- Englander SW, Englander JJ. 1978. Hydrogen-tritium exchange. *Methods Enzymol* 49G:24–39.
- Englander SW, Englander JJ, McKinnie RE, Ackers GK, Turner GJ, Westrick JA, Gill SJ. 1992. Hydrogen exchange measurement of the free energy of structural and allosteric change in hemoglobin. *Science* 256:1684–1687.
- Englander SW, Kallenbach NR. 1984. Structural dynamics of proteins and nucleic acids. *Q Rev Biophys* 16:521–655.
- Englander SW, Mauel C. 1972. Hydrogen exchange detection of discrete ligand-induced changes in hemoglobin. *J Biol Chem* 247:2387–2394.
- Englander SW, Poulsen A. 1969. Hydrogen-tritium exchange of the random chain polypeptide. *Biopolymers* 7:329–393.
- Englander SW, Staley R. 1969. Measurement of the free and the H-bonded amides of myoglobin. *J Mol Biol* 45:277–295.
- Frauenfelder H, Petsko GA, Tsernoglou D. 1979. Temperature dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* 280:558–563.
- Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscapes and motions of proteins. *Science* 254:1598–1603.
- Goodman E, Kim P. 1991. Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide. *Biochemistry* 30:11615–11620.
- Gurd FRN, Rothgeb M. 1979. Motions in proteins. *Adv Protein Chem* 33:73–165.
- Haslam JL, Eyring EM. 1967. Deuterium oxide solvent isotope effects on N-H···O, O-H···N, and N-H···N intramolecular hydrogen bonds. *J Phys Chem* 71:4470–4475.
- Hilton BD, Trudeau K, Woodward CK. 1981. Hydrogen exchange rates in pancreatic trypsin inhibitor are not correlated to thermal stability in urea. *Biochemistry* 20:4697–4703.
- Hilton BD, Woodward CK. 1979. On the mechanism of isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biochemistry* 18:5834–5841.
- Huang Z, Gabriel JM, Baldwin MA, Fletterick RJ, Prusiner SB, Cohen FE. 1994. Proposed three-dimensional structure for the cellular prion protein. *Proc Natl Acad Sci USA* 91:7139–7143.
- Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249:1544–1548.
- Hvidt A. 1964. Discussion of the pH dependence of the H–D exchange of proteins. *CR Trav Lab Carlsberg* 34:299–317.
- Hvidt A, Linderstrom-Lang K. 1954. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochim Biophys Acta* 14:574–575.
- Hvidt A, Nielsen SO. 1966. Hydrogen exchange in proteins. *Adv Protein Chem* 21:287–386.
- Ikegami A, Kono N. 1967. Tritium-hydrogen exchange of polypeptides in aqueous solutions. *J Mol Biol* 29:251–274.
- Karplus M, McCammon JA. 1981. The internal dynamics of globular proteins. *CRC Crit Rev Biochem* 9:293–349.
- Kim KS, Woodward C. 1993. Protein internal flexibility and global stabil-

- ity: Effect of urea on hydrogen exchange rates of bovine pancreatic trypsin inhibitor. *Biochemistry* 32:9609–9613.
- Kossiakoff AA. 1982. Protein dynamics investigated by the neutron diffraction-hydrogen exchange technique. *Nature* 296:713–721.
- Kuwajima K, Baldwin RL. 1983. Exchange behavior of the H-bonded amide protons in the 3-13 helix of ribonuclease S. *J Mol Biol* 169:299–324.
- Lau KF, Dill KA. 1989. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* 22:3986–3997.
- Lau KF, Dill KA. 1990. Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA* 87:638–642.
- Leichtling BH, Klotz IM. 1966. Catalysis of H–D exchange in polypeptides. *Biochemistry* 5:4026–4036.
- Levitt M. 1981a. Hydrogen bond and internal solvent dynamics of BPTI protein. *Ann NY Acad Sci* 367:162–181.
- Levitt M. 1981b. Molecular dynamics of hydrogen bonds in bovine pancreatic trypsin inhibitor protein. *Nature* 294:379–380.
- Lumry R, Rosenberg A. 1975. The mobile defect hypothesis of protein function. *Colloq Int CNLRS* 246:55–63.
- Mayo SL, Baldwin RL. 1993. Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science* 262:873–876.
- Molday RS, Englander SW, Kallen RG. 1972. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* 11:150–158.
- Nakanishi M, Tsuboi M, Ikegami A. 1972. Fluctuations of the lysozyme structure. *J Mol Biol* 70:351–361.
- Nakanishi M, Tsuboi M, Ikegami A. 1973. Fluctuations of the lysozyme structure. II. Effects of temperature and binding of inhibitors. *J Mol Biol* 75:673–682.
- Novotny J, Brucoleri R, Karplus M. 1984. An analysis of incorrectly folded protein models. *J Mol Biol* 177:787–818.
- Otting G, Liepinsh E, Wüthrich K. 1991. Protein hydration in aqueous solution. *Science* 254:974–980.
- Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE. 1993. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA* 90:10962–10966.
- Pedersen T, Thomsen N, Andersen K, Madsen J, Poulsen F. 1993. Determination of the rate constants k_1 and k_2 of the Linderstrom-Lang model for protein amide hydrogen exchange. *J Mol Biol* 230:651–660.
- Pershina L, Hvidt A. 1974. A study by the hydrogen-exchange method of the complex formed between the basic pancreatic trypsin inhibitor and trypsin. *Eur J Biochem* 48:339–344.
- Radford S, Dobson C, Evans P. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358:302–307.
- Richards FM. 1979. Packing defects, cavities, volume fluctuations, and access to the interior of proteins. *Carlsberg Res Commun* 44:47–63.
- Roder H. 1989. Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol* 176:446–473.
- Roder H, Elove GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335:700–704.
- Roder H, Wagner G, Wüthrich K. 1985. Amide proton exchange in proteins by EX1 kinetics: Studies of the basic pancreatic trypsin inhibitor at variable pH and temperature. *Biochemistry* 24:7396–7407.
- Rohl CA, Scholtz JM, York EJ, Stewart JM, Baldwin RL. 1992. Kinetics of amide proton exchange in helical peptides of varying chain lengths. Interpretation by the Lifson-Roig equation. *Biochemistry* 31:1263–1269.
- Rose MC, Stuehr J. 1968. Kinetics of proton transfer reactions in aqueous solution: Rates of internally hydrogen-bonded systems. *J Am Chem Soc* 90:7205–7209.
- Schoenborn BP, Hanson JC, Darling GD, Norvell JC. 1978. Real space refinement of neutron diffraction data from carbon monoxide sperm whale myoglobin. *Acta Crystallogr A* 34:65.
- Segal DM, Harrington WF. 1967. The tritium-hydrogen exchange of myosin and its proteolytic fragments. *Biochemistry* 6:768–787.
- Shortle D, Chan HS, Dill KA. 1991. Modeling the effects of mutations on the denatured states of proteins. *Protein Sci* 1:201–215.
- Tuchsen E, Hayes J, Ramaprasad S, Copie V, Woodward C. 1987. Solvent exchange of buried water and hydrogen exchange of peptide NH groups hydrogen bonded to buried waters in bovine pancreatic trypsin inhibitor. *Biochemistry* 26:5163–5172.
- Tuchsen E, Hvidt A, Ottesen M. 1980. Enzymes immobilized as crystals. Hydrogen isotope exchange of crystalline lysozyme. *Biochimie* 62:563–566.
- Tuchsen E, Ottesen M. 1979. A simple hydrogen exchange method for cross-linked protein crystals. *Carlsberg Res Commun* 44:1–10.
- Udgaonkar JB, Baldwin RL. 1988. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335:694–699.
- Wagner G, Wüthrich K. 1982. Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor (BPTI) in solution: Studies with two dimensional NMR. *J Mol Biol* 160:343–361.
- Welch WH, Fasman GD. 1974. Hydrogen-tritium exchange in polypeptides. Models of alpha-helical and beta conformations. *Biochemistry* 13:2455–2466.
- Wemmer DE, Kallenbach NR. 1983. The structure of apamin in solution: 2D NMR study. *Biochemistry* 22:1901–1906.
- Wlodawer A, Sjolín L. 1982. Hydrogen exchange in ribonuclease A: Neutron diffraction study. *Proc Natl Acad Sci USA* 79:1418–1422.
- Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science* 267:1619–1620.
- Woodward CK. 1977. Dynamic solvent accessibility in the soybean trypsin inhibitor-trypsin complex. *J Mol Biol* 11:509–515.
- Woodward CK, Ellis LM, Rosenberg A. 1975. Solvent accessibility in folded proteins: Studies of hydrogen exchange in trypsin. *J Biol Chem* 250:432–439.
- Woodward CK, Hilton BD. 1979. Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annu Rev Biophys Bioeng* 8:99–127.
- Woodward CK, Hilton BD. 1980. Hydrogen isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biophys J* 32:561–576.
- Woodward CK, Rosenberg A. 1971. Studies of hydrogen exchange in proteins. VI. Urea effects on RNase hydrogen exchange kinetics leading to a general model for hydrogen exchange from folded proteins. *J Biol Chem* 246:4114–4121.
- Woodward CK, Simon I, Tuchsén E. 1982. Hydrogen exchange and the dynamic structure of proteins. *Mol Cell Biochem* 48:135–160.
- Wüthrich K, Wagner G. 1978. Internal motion in globular proteins. *Trends Biochem Sci* 3:227–230.
- Yee DP, Dill KA. 1993. Families and the structural relatedness among globular proteins. *Protein Sci* 2:884–899.
- Yee RY, Englander SW, von Hippel PH. 1974. Native collagen has a two-bonded structure. *J Mol Biol* 83:1–16.

Appendix

Accessibilities and Arrhenius slopes in the HP model

In this Appendix, we first convert Equation 7 into a form more convenient for the HP model. Equation 7 gives the Boltzmann-averaged solvent accessibility, $\langle A \rangle_i$, of monomer i , for a set of N conformations, c , of free energy E_c . It is more convenient to group the N conformations by their numbers of HH contacts, h . Because the number of conformations with h HH contacts is the density of states, $g(h)$, we can express Equation 7 as:

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_N} \sum_{c=1}^{g(h)} A_{c,i} e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} \sum_{c=1}^{g(h)} e^{-h\epsilon/kT}} \quad (11)$$

where h_N is the number of HH contacts in the native conformation. Because the Boltzmann exponential factors depend only on h and not on particular conformations, c , Equation 11 becomes

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_N} e^{-h\epsilon/kT} \sum_{c=1}^{g(h)} A_{c,i}}{\sum_{h=0}^{h_N} g(h) e^{-h\epsilon/kT}} \quad (12)$$

Equation 12 can be further simplified by defining a quantity, $A_{h,i}$, which is the average solvent accessibility of monomer i for all $g(h)$ conformations having h HH contacts:

$$A_{h,i} \equiv \frac{1}{g(h)} \sum_{c=1}^{g(h)} A_{c,i} \quad (13)$$

For example, Figure 4B, C, and D shows that monomer 14 has solvent accessibilities of 1, 0, and 1 in the three conformations with $h = 7$, so its average solvent accessibility at this energy level is $A_{7,14} = 2/3$. Substituting Equation 13 into Equation 12 the average solvent accessibility becomes

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_N} A_{h,i} g(h) e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h) e^{-h\epsilon/kT}}, \quad (14)$$

which is the result given in Equation 8. Because some monomers are solvent inaccessible in the native ($A_{h_N,i} = 0$) and other compact states, we define h_i^* as the largest value of h for which $A_{h,i} \neq 0$. Thus, for any monomer, i , buried in the native state, $h_i^* = h_{N-1}$ if it becomes solvent accessible in any of the first-excited states, $h_i^* = h_{N-2}$ if it remains buried until the second-excited states, and so on. Figure 4 shows that $h_i^* = 7$ for monomers 8, 11, 14, and 16, all of which are buried in the native state but solvent accessible in the first-excited states. For monomer 15, which is buried in the native state and all three first-excited states, $h_i^* = 6$. Because $A_{h,i} = 0$ for $h > h_i^*$, the upper summation limit in the numerator becomes h_i^* , and Equation 14 becomes

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_i^*} A_{h,i} g(h) e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h) e^{-h\epsilon/kT}}. \quad (15)$$

The slope of the temperature plot in Figure 5A is obtained by differentiating $\log \langle A \rangle_i$ with respect to $|\epsilon/kT|$. From Equation 15:

$$m(\epsilon)_i = \frac{\sum_{h=0}^{h_i^*} h A_{h,i} g(h) e^{-h\epsilon/kT}}{\sum_{h=0}^{h_i^*} A_{h,i} g(h) e^{-h\epsilon/kT}} - \frac{\sum_{h=0}^{h_N} h g(h) e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h) e^{-h\epsilon/kT}}. \quad (16)$$

The second term in Equation 16 is simply the average number of HH contacts for the chain, $\langle h \rangle$. To first approximation, $\langle h \rangle = h_N$ under native conditions, because Figure 6 shows that the native population dominates over the experimental range in $|\epsilon/kT|$. The first term in Equation 16 can be simplified by assuming that solvent accessibility is two-state for each free energy level, i.e., that $A_{h,i}$ must equal 0 or 1. In other words, once a buried monomer becomes solvent accessible (in one or more of the $h = h_i^*$ conformations), it is assumed to be completely solvent accessible in every conformation of equal or more positive contact energy. This approximation does not significantly affect the slope of the exchange curve, although it does have a small effect on the two-slope transition temperature (data not shown). With the approximations given above, Equation 16 becomes

$$m(\epsilon)_i = \frac{\sum_{h=0}^{h_i^*} h g(h) e^{-h\epsilon/kT}}{\sum_{h=0}^{h_i^*} g(h) e^{-h\epsilon/kT}} - h_N = \langle h \rangle_{acc,i} - h_N \quad (17)$$

where $\langle h \rangle_{acc,i}$ is the average value of h for only those conformations that are solvent accessible for monomer i , i.e., those for which $h \leq h_i^*$. Many exchange-competent conformations can contribute to exchange, so $\langle h \rangle_{acc,i}$ is determined by those that are most highly populated for a given $|\epsilon/kT|$. Because the contact free energy is $h\epsilon$, Equation 17 shows that to first approximation, the slope of the exchange curve is proportional to the free energy difference between the native state and predominant exchange-competent state.