Correlation between Native−State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model †

ARTICLE *in* BIOCHEMISTRY · FEBRUARY 1998

Impact Factor: 3.02 · DOI: 10.1021/bi9720641 · Source: PubMed

Correlation between Native-State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model†

I. Bahar,‡,§ A. Wallqvist,||,⊥ D. G. Covell,|| and R. L. Jernigan*,‡

Molecular Structure Section, Laboratory of Experimental and Computational Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-5677, Polymer Research Center and Chemical Engineering Department, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey, and Frederick Cancer Research and Development Center, National Cancer Institute, Science Applications International Corporation, Frederick, Maryland 21702

Received August 20, 1997; Revised Manuscript Received November 14, 1997

ABSTRACT: Recently, we developed a simple analytical model based on local residue packing densities and the distribution of tertiary contacts for describing the conformational fluctuations of proteins in their folded state. This so-called Gaussian network model (GNM) is applied here to the interpretation of experimental hydrogen exchange (HX) behavior of proteins in their native state or under weakly denaturing conditions. Calculations are performed for five proteins: bovine pancreatic trypsin inhibitor, cytochrome *c*, plastocyanin, staphylococcal nuclease, and ribonuclease H. The results are significant in two respects. First, a good agreement is reached between calculated fluctuations and experimental measurements of HX despite the simplicity of the model and within computational times 2 or 3 orders of magnitude faster than earlier, more complex simulations. Second, the success of a theory, based on the coupled conformational fluctuations of residues near the native state, to satisfactorily describe the native-state HX behavior indicates the significant contribution of local, but cooperative, fluctuations to protein conformational dynamics. The correlation between the HX data and the unfolding kinetics of individual residues further suggests that local conformational susceptibilities as revealed by the GNM approach may have implications relevant to the global dynamics of proteins.

Hydrogen exchange (HX) data for proteins directly indicate the relative accessibility of various protons to solvent, typically under conditions that can be denaturing to different extents. Observed HX times can vary over extremely broad ranges, from instantaneous to months. Generally these various times have been interpreted and ascribed to several kinds of processes—local fluctuations for the fastest, local unfolding for intermediate, and global unfolding for the slowest exchanges *(1)*. Some disagreements have arisen between the descriptions of the underlying kinetic processes and the interpretation of HX mechanisms because of differences in the operational definitions of these processes *(2)*. HX from native proteins was recently pointed out to monitor with a remarkable precision rapid conformational fluctuations of the order of microseconds to milliseconds, while no correlation could be observed between HX free energies of individual residues and their unfolding rates which are several orders of magnitude slower than local fluctuations *(3)*. Global unfolding rates are expected to be the same for all residues, while the superimposed exchange rates due to residue fluctuations vary depending on local interactions *(4, 1, 5)*. Here we will avoid ascribing rates to specific processes but instead investigate directly the extent of protection in the native state and compare this with the measured exchange rates. This analysis has become possible with the development of a new simple model of protein structure and dynamics, recently proposed for analyzing the conformational fluctuations of folded proteins *(6, 7)*.

Two limits, or two mechanisms of exchange, have been proposed for the HX behavior of proteins: the EX2 limit, which generally describes the exchange in the native state or under relatively mild denaturing conditions *(8−10)*, and the EX1 limit, which takes place at high temperature, high pH (≥ 7), or high concentration of denaturants *(2)*. In the former case, a preequilibrium between open (exchange-susceptible) and closed (protected) conformations of amide protons exists; the observed exchange rates, k_{obs}, reflect the rate constant k_x for the base-catalyzed chemical exchange step H ↔ D, subject to the thermodynamic equilibrium between open and closed states. The pre-equilibrium is achieved by rapid changes between exchange-incompetent and exchange-competent conformations prior to the chemical exchange step *(9)*. Such rapid motions have been attributed to both local fluctuations and global unfolding *(11, 12)*. In the EX1 case, on the other hand, the exchange is fast compared to the conformational motions that expose amide protons to solvent; and the slow process of conformational opening, or global unfolding, controls the observed rate. Adopting a kinetic scheme of the form *(8−11, 5)*

$$\text{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{open} \xrightarrow{k_x} \text{exchange}$$

† Partial support from NATO CRG Project 951240 is acknowledged.
* Corresponding author.
‡ NIH.
§ Bogazici University and TUBITAK.
|| Science Applications International Corp.
⊥ Present address: Department of Chemistry, Rutgers University, P.O. Box 939, Piscataway, NJ 08855-0939.

S0006-2960(97)02064-3
Published on Web 01/06/1998

where k_{op} and k_{cl} are the rate constants for the opening, and closing of protein conformations, the observed exchange rate $k_{obs} = k_{op}\, k_x/(k_{cl} + k_x)$ for $k_{op} << k_{cl}$ reduces to $k_{obs} = k_{op}$ in the EX1 limit, given that $k_{cl} << k_x$. In the EX2 limit, on the other hand, $k_x << k_{cl}$, which leads to $k_{obs} = K_{op}\, k_x$. Here K_{op} is the equilibrium constant of the opening process, $K_{op} = k_{op}/k_{cl}$. In this limit, the free energy change for the opening/closing equilibrium measured in HX experiments becomes

$$\Delta G_{HX} \equiv -RT \ln (k_{obs}/k_x) = -RT \ln K_{op} \qquad (1)$$

The ratio k_x/k_{obs}, also referred to as the protection factor, provides a measure of the resistance of amide protons to exchange in the folded structures relative to their intrinsic exchange rates (k_x) in a random polypeptide *(13)*. In the most general description of HX behavior, both EX1 and EX2 mechanisms would be considered, although one or the other may dominate depending on the experimental conditions *(5, 2, 14, 3)*.

In the present study, we show that a theory which recently proved successful for describing the fluctuation behavior of folded proteins near native-state coordinates *(6, 7)* may be equally well extended to the interpretation of H/D exchange data. Our methodology is identical to that of classical treatments of the vibrational behavior of junctions in networks *(15, 16)*. In our model, residues are viewed as the network junctions, and a single parameter harmonic potential is assumed to exist between all near neighbor (≤ 7.0 Å) residue pairs, whether or not they are sequential neighbors. As a consequence, each residue is subject to Gaussian fluctuations about its mean position, hence the name *Gaussian network model* (GNM). A major simplification is that the same interaction force constant is adopted for all residue pairs. This has been inspired by a recent normal mode analysis of Tirion *(17)* in which a single-parameter harmonic potential was shown to be as satisfactory as detailed residue-specific potentials for describing the dynamics of protein crystals. Application of GNM to a series of proteins, varying in structural class and size, indicated an excellent agreement between theoretical mean-square fluctuations of residues and crystallographic *B* factors *(6)*.

The GNM is applied here to five proteins: bovine pancreatic trypsin inhibitor (BPTI), cytochrome *c* (cyt *c*), plastocyanin, staphylococcal nuclease (SNase), and ribonuclease H (RNase H). The results, obtained within computation times of the order of seconds, are in excellent agreement with (i) the protection factor measurements from H/D exchange experiments carried out for native-state or weakly denaturing conditions *(18−20, 11, 12)*, and (ii) more detailed theoretical approaches employing residue-specific potentials and computationally expensive combinatorial methods *(21−23)*.

Thus, the GNM lends itself to an estimation of the HX susceptibilities/resistances of individual residues by providing a precise measure of the amplitudes of conformational fluctuations in the restrictive tertiary context of the native folded structure. In a strict sense, the theory calculates the free energy cost of distorting residues on a local scale. Whether this energy cost also reflects the unfolding free energy of individual residues, given the high cooperativity between residue fluctuations in the model and the observed strong correlation between theory and HX data, will be discussed in detail. Yet, the close agreement between the HX data for individual residues taken under conditions where the EX2 mechanism dominates on the one hand, and the fluctuations of residues predicted by a simple model based on local packing densities and topology of tertiary contacts in crystal structures on the other, suggests that cooperative local fluctuations near the folded state have a direct effect, more important than previously thought, on the observed exchange rates. More significantly, we show that there is no need to invoke the additional contribution of global unfolding for a first-order interpretation of the native-state HX data, but instead we simply evaluate the free energy change associated with the cooperative conformational fluctuations of residues in their native compact tertiary context. In a sense this is consistent with the fact that the HX experiments presently considered were performed either under native-state conditions or under weakly unfolding conditions in which the protein appears to be completely folded when analyzed by probes such as circular dichroism and fluorescence. The further correlation between the calculated or experimental HX data and the unfolding kinetics of individual residues suggests that information relevant to the global conformational susceptibilities may be included in the spectra from X-ray or neutron diffraction techniques, as pointed out in previous reviews *(24, 25)*.

THEORY

The GNM of proteins is based on the following assumptions: (i) The native structure is viewed as a network of homogeneous interactions between all residue pairs [*i*, *j*], bonded or nonbonded, separated by a distance $R_{ij} \leq r_c = 7$ Å. Here, the residue positions in space are conveniently identified with those of the α-carbons, and the cutoff distance r_c is selected on the basis of the radius of the first coordination shell around residues in native folds, as indicated by knowledge-based analyses of databank structures *(26−29)*. (ii) The interaction energy between all pairs is assumed to obey a harmonic potential. This form of the pairwise interaction potential ensures Gaussian fluctuations of residues around their mean coordinates. (iii) A single parameter γ, the force constant of the harmonic potential, is adopted for all residue pairs regardless of residue type. This assumption is justifiable in part by recent observations that a significant fraction (above 80%) of the effective empirical interresidue potentials in native folds is contributed by the homogeneous interactions stabilizing compact globular forms *(28)*. The good agreement reached between theoretical mean-square fluctuations and crystallographic *B* factors lends support to the validity of the model and its assumptions *(6)*.

Under these assumptions, the potential energy of the protein is expressed as

$$V = \Sigma_i \Sigma_j V(\mathbf{R}_i, \mathbf{R}_j) = {}^1/_2\gamma \text{ tr } \{[\Delta\mathbf{R}]^T\, \Gamma\, \Delta\mathbf{R}\} \qquad (2)$$

where $\Delta\mathbf{R}$ is the *n*-dimensional hypervector, the elements of which are the fluctuations vectors $\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$ in the positions $\mathbf{R}_i$ of residues $1 \leq i \leq n$ around mean (or crystallographic) coordinates $\mathbf{R}_i^0$, tr designates the trace of the 3×3 matrix enclosed in braces, $V(\mathbf{R}_i, \mathbf{R}_j)$ is the harmonic potential governing the fluctuations $\Delta\mathbf{R}_{ij} = (\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)$ in interresidue distances given by

$$V(\mathbf{R}_i, \mathbf{R}_j) = {}^1/_2\gamma\,(\Delta\mathbf{R}_{ij})^2 = {}^1/_2\gamma\,(\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)(\Delta\mathbf{R}_j - \Delta\mathbf{R}_i) \quad (3)$$

and Γ is the so-called Kirchhoff matrix, describing the pairs of contacting residues: the off-diagonal elements of Γ are assigned the value $\Gamma_{ij} = \Gamma_{ji} = -1$ if $R_{ij} \leq r_c$ and zero otherwise; and the diagonal elements are equal to the negative sum of the off-diagonal elements in the same row (or column). Thus, Γ_{ii} represents simply the coordination number of the ith residue and thereby reflects the *local packing density* in the neighborhood of residue i, $1 \leq i \leq n$, in a protein of n residues.

The equilibrium correlations $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$ between fluctuations of residues i and j are proportional to the ijth element of the inverse of Γ, $[\Gamma^{-1}]_{ij}$, following the identity *(15, 16, 30, 6)*

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{1}{Z}\int \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \exp\{-V/k_BT\}\, d\{\Delta\mathbf{R}\} = (3k_BT/\gamma)[\Gamma^{-1}]_{ij} \quad (4)$$

Here Z is the partition function $Z = \int \exp\{-V/k_BT\}\, d\{\Delta\mathbf{R}\}$, k_B is the Boltzmann constant, T is the absolute temperature, and $d\{\Delta\mathbf{R}\}$ designates the integration over all residue fluctuations. The mean-square fluctuations $\langle(\Delta\mathbf{R}_i)^2\rangle$ of residues are then readily found from the diagonal elements $[\Gamma^{-1}]_{ii}$ of Γ^{-1}.

In the present model, the fluctuations of residues about their mean positions obey the Gaussian distribution

$$W(\Delta\mathbf{R}_i) = A \exp\{-3(\Delta\mathbf{R}_i)^2/2\,\langle(\Delta\mathbf{R}_i)^2\rangle\} \quad (5)$$

which leads, after use of eq 4 for the case $i = j$, to the conformational entropy change ΔS_i

$$\Delta S_i = k_B \ln W(\Delta\mathbf{R}_i) = -\gamma\,(\Delta\mathbf{R}_i)^2/(2T[\Gamma^{-1}]_{ii}) \quad (6)$$

associated with $\Delta\mathbf{R}_i$, the fluctuation of residue i. Equation 6 also reflects the free energy increase of entropic origin

$$\Delta G_i = -T\Delta S_i = \frac{\gamma}{2}(\Delta\mathbf{R}_i)^2/[\Gamma^{-1}]_{ii} \quad (7)$$

contributed by the ith residue, upon distortion of its coordinates by an amount $\Delta\mathbf{R}_i$. This energy increase is inversely proportional to $[\Gamma^{-1}]_{ii}$, or alternatively $\langle(\Delta\mathbf{R}_i)^2\rangle$. Physically, this signifies a stronger resistance to deformation, including unfolding, of residues subject to lower amplitude fluctuations in the folded state.

In the following, the free energy change associated with the conformational entropy change (eq 6) will be evaluated for a series of proteins. Basically, the energy cost of distorting the residue positions by a unit of $\Delta\mathbf{R}_i$ will be sought and compared with the free energy changes or unfolding penalties revealed by HX experiments. The parameter γ will be suitably adjusted to rescale the theoretical results with respect to experimental data.

RESULTS AND COMPARISON WITH HX EXPERIMENTS

We calculate the mean-square fluctuations of residues by means of the inverse of the connectivity matrix Γ for five proteins, BPTI, cyt *c*, plastocyanin, SNase, and RNase H, and compare the results with the free energy increases observed in H/D exchange experiments, by means of eq 7. The crystal structures *(31−37)* available in the Brookhaven Protein Data Bank (PDB) *(38)* are used for evaluating the matrix Γ for each protein. Detailed information on the structural characteristics of the first four proteins was given in our previous study *(23)* and will be omitted here. RNase H, which is added to our previous set, will be considered below in greater detail.

Results are displayed as separate panels in Figures 1 and 2 for the five proteins examined. In all panels, data from H/D exchange experiments (bars) are displayed together with the theoretical curves evaluated using the GNM (solid curves). The helical and β-strand regions of the proteins are indicated in the lower parts of the panels. Peaks refer to residues exhibiting a strong resistance to conformational change, i.e., slow exchangers, while minima indicate the opposite behavior, i.e., yielding or easily fluctuating regions. H/X data are not reported for solvent-exposed or highly flexible residues whose protons exchange within the dead time of experiments. On the other hand, it is possible to calculate free energy changes for all residues; hence continuous curves are obtained by GNM. Generally, regions participating in secondary structure elements exhibit relatively more pronounced free energy increases, although departures from this trend are also observed *(39, 40)*; and conformational constraints other than intramolecular hydrogen bonds, including surface N−H exchange *(41, 42)* and solvent penetration *(1)*, have been pointed out to be important in determining the HX behavior. Overall, good agreement is observed between H/D exchange data and the present theory. Even relatively sparse, weak peaks observed in experiments can be traced out in the theoretical curves.

For illustrative purposes, the results recently obtained *(23)* from a combinatorial analysis coupled with a knowledge-based scoring function are also displayed as dotted curves in Figures 1 and 2. The approach proposed by Hilser and Freire *(21)* was used therein. The GNM calculations are 2−3 orders of magnitude faster than such combinatorial analyses and do not require knowledge of residue-specific potentials. In fact, the method of calculation in GNM is analytical; no simulations are required. Local packing density, chain connectivity, and number distribution of residue contacts are the only properties that determine the GNM results. Yet GNM provides an even more detailed account of the exchange behavior at the residue level, capturing some of the peaks that had been smoothed out in the previous calculations.

Bovine Pancreatic Trypsin Inhibitor. The experimental results in Figure 1a refer to the H/D exchange protection factors against unfolding measured by Kim, Fuchs, and Woodward *(19)* converted into unfolding free energies, following the procedures outlined by Englander and coworkers *(43, 13, 10)*. A strong correlation between experimental data and theoretical results is observed. The crystal structure determined at 1.0 Å resolution *(33)* (PDB entry 5pti) has been used in the calculations.

A distinctive feature observed in Figure 1a is the relatively high free energy change for Phe22, located in the middle of the first β-sheet, and its close neighbors, which conforms with experiments *(44, 45)*. This highly stable region is

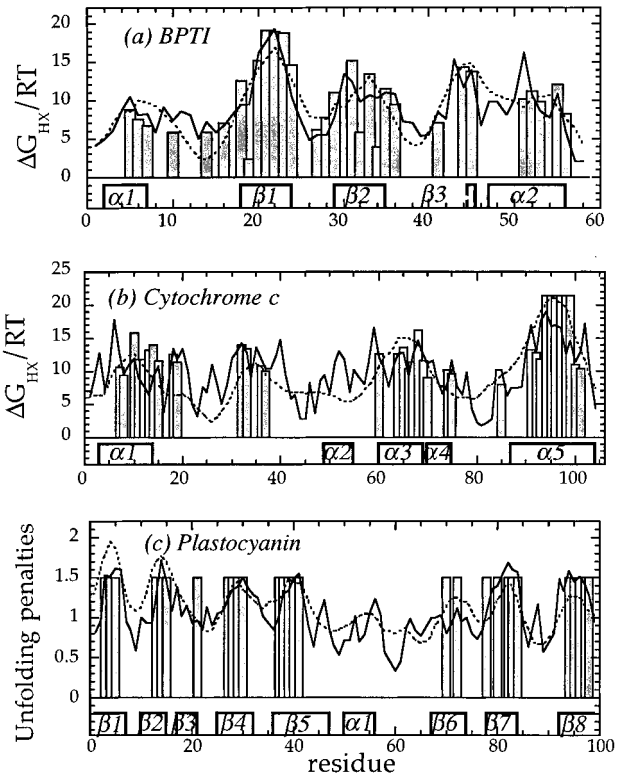

FIGURE 1: Comparison of experimental and theoretical free energy changes for H/D exchange. Results are presented for BPTI, cytochrome *c*, and plastocyanin in panels a−c, respectively. In all three panels, vertical bars refer to experimental data, and solid curves are the theoretical results found using eq 7, based on the Gaussian network model (GNM) of proteins *(6)*. For comparison, the results from the detailed combinatorial analysis with residue-specific potentials recently performed by Wallqvist et al. *(23)* are displayed by the dotted curves. The secondary structure elements of the proteins are indicated along the abscissa. Experimental results for BPTI in panel a refer to the free energy changes evaluated from the exchange rates observed at 30 °C and pH 3.5 *(19)*. The coordinates of the BPTI α-carbons for GNM calculations are taken from the neutron and X-ray refined structure of BPTI *(33)*. In panel b the bars indicate the data from the HX experiments on cyt *c (11)*, and the solid curve represents the results from GNM calculations using the crystal structure of horse heart cyt c at low ionic strength *(37)*. Panel c displays the energy penalties (arbitrary units) for perturbing residue conformations in plastocyanin. The bars indicate the residues determined by NMR spectroscopy *(56)* to be stable or resistant to exchange. The solid curve is obtained by the GNM using the NMR structure of French bean plastocyanin *(57)*.

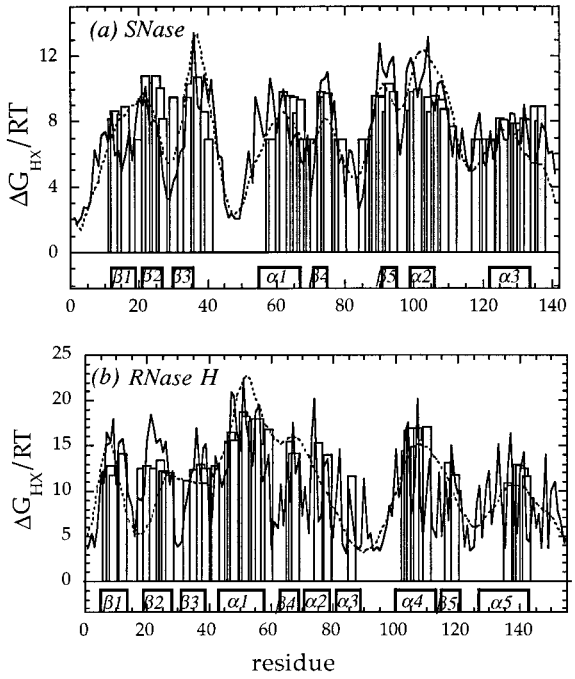

FIGURE 2: Comparison of experimental and theoretical free energy changes for HX of (a) staphylococcal nuclease and (b) ribonuclease H, presented in the same format as Figure 1. Bars in panel a refer to HX measurements of Loh et al. *(20)*, and the corresponding GNM calculations (solid curve) are done using the PDB crystal structure of SNase determined at 1.5 Å resolution *(31)*. Bars in panel b refer to experimental data taken *(12)* at low concentrations of denaturant, under conditions in which the protein appears to be completely folded when examined by circular dichroism or fluorescence. The solid curve representing the GNM results is obtained using the crystal structure of RNase H obtained at 1.48 Å resolution *(36)* (PDB entry 2rn2). Side-chain interaction sites, as defined in our previous studies *(29, 28)*, are adopted here instead of the α-carbons, for improving the accuracy.

complemented by the residues Asn43−Phe45 and by the disulfide bridge Cys30−Cys51. We note that the mutants F22A, Y23A, N43G, and F45A in this region were reported to be highly destabilizing relative to the wild-type protein *(46, 47)*. A strong resistance to HX is observed to take place at β-strands, in conformity with early NMR exchange measurements *(48)*. This behavior, generally attributed to hydrogen bonding of amide protons in the β-sheet, is captured here on the basis of local packing densities and tertiary contact distributions alone.

Among the three disulfide bridges of the protein, Cys14−Cys38 appears to be the least protected. This may be correlated with the experimental observation that this disulfide bond is the last to form on the folding pathway *(49, 50)*. The equilibrium constant for forming each of the three disulfide bonds of BPTI have been measured *(51)* and the Cys14−Cys38 bond is known to be the weakest.

Cytochrome c. The free energy changes for horse-heart cyt *c* in the absence of the heme group are shown in Figure 1, panel b. The bars indicate the data from the hydrogen exchange experiments of Bai et al. *(11)*, taken at a pD of 7.0. These exhibit the same dominant features as the earlier measurements at low pH in strong saline solution *(18)*. We have performed GNM calculations by using two different crystal structures *(34, 37)* with PDB codes 1hrc and 1crc (at low ionic strength), obtained at 1.9 and 2.1 Å resolution, respectively. The heme group was not included in the calculations. The results for 1crc only are displayed in panel b of Figure 1; those of 1hrc show quite similar trends but are not shown.

The middle portions of the two terminal helices (helix A, Asp2−Cys14; helix E, Lys87−Glu104) exhibit the strongest resistance to conformational change. These two helices were observed by time-resolved circular dichroism and fluorescence spectroscopy to form at early folding steps *(52)*, in conformity with earlier HX pulse labeling detection of folding intermediates *(53)*. A more careful examination of the theoretical curve indicates that the highest peaks occur precisely at Gly6 and Glu94. These two residues are highly conserved among members of the cyt *c* family. In particular, the region near Glu94, i.e., the middle portion of helix E, is distinguished by a broad maximum, including peaks at Arg91, Asp93−Tyr97, and Thr102, which may be correlated with the widely accepted identification of this helix as being

crucial to the stability of cyt *c*. The interactions between conserved residues at the interface (e.g., Gly6−Leu94 and Phe10−Tyr97) were also found to stabilize the molten globule, or A state, of the protein *(54)*.

The next highest stability points are Trp59, Leu68, Leu32, and Thr40, located near the hydrophobic heme binding pocket. Although the 60s helix (Lys60−Glu69) is also important for the stability of cyt *c (11)*, this helix appears here to be more similar to the loop Leu32−Thr40 insofar as its conformational fluctuations are concerned, rather than the terminal helices. This observation is in accord with the fact that the highest energy global unfolding isotherm obtained for cyt *c* HX at low concentrations of denaturant corresponds to the terminal helices, while the HX data for the 60s helix, and for the region near the loop Leu32−Thr40 is attributed to a subglobal unfolding *(11, 55)*. Such a hierarchy was indeed suggested by the HX pulse-labeling experiments of Roder and collaborators: the terminal helices were observed to acquire about 60% protection within the 20 ms burst phase of refolding, while the 60s and 70s helices formed at later stages *(53)*.

In the extreme case of residues subject to the largest amplitude motions, which may consequently possess the greatest tendency to unfold, our theory indicates Met79−Phe81, again in accord with H/D unfolding isotherms *(11, 55)*.

Plastocyanin. Figure 1c displays the results for plastocyanin. The bars represent the slowly exchanging regions identified in NMR experiments *(56)*. GNM calculations were performed using both the NMR structure of French bean plastocyanin *(57)* (PDB entry 9pcy, model 1) and the crystal structure of poplar plastocyanin determined *(32)* at 1.6 Å resolution (7pcy). The two structures are superimposable with an rms deviation of 3.7 Å for backbone coordinates. The GNM results obtained for the two structures are nearly indistinguishable.

We note that free energy changes calculated for the different strands of the protein are generally of comparable size. For example, among the four strands present in each of the two β-pleated sheets of plastocyanin, Ala1−Gly7 and Glu25−Asn32 in sheet I and Gly78−Cys84, Met92−Glu99, and the N-terminal part of strand Pro36−Pro47 in sheet II exhibit comparable behavior, mainly a considerable free energy increase manifested by relatively broad peaks. Also, there is a sharp peak of comparable magnitude near Phe14 in 9pcy (or Val15 in 7pcy). Strands Asn17−Val21 and Gly67−Leu74, on the other hand, differ from the others in their weaker responses.

The occurrence of several maxima of comparable magnitude in the GNM theoretical curve signals the presence of a long-range (along the chain sequence) coupling between the different structural elements of the protein. Thus, it is hard to conceive here of a single structural unit acting as a core, in contrast to the results for BPTI and cyt *c*, where a few loci could be distinguished by their significantly higher stability. The more cooperative nature of residue fluctuations observed may be attributed simply to the fact that plastocyanin is a β-class protein, with only one short stretch of helical residues, and the overall tertiary structure is stabilized by interstrand contacts. In fact, attempts to fold short fragments so as to understand the earliest events in folding pathway showed that the peptide fragments derived from French bean plastocyanin were relatively devoid of any folded backbone conformation and adopt random dihedral angles, as observed in solution by NMR, in contrast to myohemerythrin fragments, for example, which exhibit a strong tendency to preform helices *(58)*.

Staphylococcal Nuclease. Figure 2a displays the results for SNase. Experimental free energy changes *(20)* are shown together with the GNM results (solid curve) obtained from the crystal structure (PDB entry 2sns) determined at 1.5 Å resolution *(31)* and results from our recent simulations (dotted curve), similarly to the panels in Figure 1.

The crystal structure 2sns is a complex with Ca^{2+} and a nucleotide inhibitor that blocks the activity (cleavage of sugar−phosphate bonds in DNA or RNA) of SNase. GNM calculations repeated with the unligated SNase crystal structure refined at 1.7 Å resolution *(35)* (PDB entry 1stn) showed that the results remain virtually unchanged in spite of the local conformational differences between the crystal structure of SNase in the ternary complex and in the isolated form.

The most significant feature from an examination of Figure 2a is the occurrence of a broad minimum centered around residues 43−49. These residues are located at the coil region extending between the third strand (Gln30−Leu36) of the β-domain and the first helix (Gly55−Glu67) of the α-domain. We recall that RNase is an $\alpha + \beta$-class protein, comprised of a five-sheet β-barrel and a cluster of three α-helices. The same loop region contains the Ca^{2+} binding residues Asp40 and Thr41, which, in contrast to near neighboring residues, are in a tightly packed environment. The high flexibility of residues 43−49 near the Ca^{2+} attachment site could be required for accommodating the ligand. A similar feature occurs near Lys84, Tyr85, and Arg87, which hydrogen-bond to inhibitor phosphates, and near Tyr113 and Tyr115, which pack against the ligand's nucleotide base.

In the case of residues exhibiting the highest increase in free energy upon distortion of crystal coordinates, we distinguish the highest peaks at Leu36, Val104, and Ala90 (or Tyr91 in 1stn). These are all buried residues, located at the termini of strands or helices, and presumably contribute to the hydrophobic core. We note that the highest peak (Leu36) in our curve lies in the sequence of hydrophobic residues (Leu36−Val39) reported by Wang and Shortle *(59)* to confer stability via nonnative interactions to a three-stranded β-sheet that persists at high urea concentrations.

Hilser and Freire *(22)* used a computational method based on a statistical thermodynamic formalism *(21)* to interpret the HX data of SNase and make predictions on the folding pathway of the protein. Their method, which was also adopted in our previous study *(23)*, yields HX free energies in good agreement with experiments. The extension to predicting a folding pathway, however, requires careful interpretation. From a statistical examination of partially folded (below 10%) structures, the three helices appear to be highly stabilized, while the β-barrel structure becomes more probable after reaching about 40% folding *(22)*. These observations may simply reflect the fact that helices are stabilized by local hydrogen bonds, while stabilization of β-strands requires tertiary contacts.

Unambiguous identification of a folding pathway is possible only by direct examination of the intermediates

formed during unfolding or folding processes, either using time-resolved techniques *(60)* or perhaps from dynamic simulations. For SNase, recent examination of a folding and unfolding mechanism by stopped-flow fluorescence techniques revealed the occurrence of a partially folded state with a stable β-domain and a largely disordered α-helical region *(61)*. Interestingly, the three highest peaks obtained here coincide with the strands β2, β3, and β5, which could lend support to a relationship between early formation and high stability.

Ribonuclease H. Figure 2b displays the results for RNase H. HX measurements of Marqusee and co-workers *(12)* at low concentrations of denaturant are shown (bars) together with two theoretical curves. The crystal structure of RNase H obtained at 1.48 Å resolution *(36)* (PDB entry 2rn2) is used in these calculations. The GNM calculations (solid curve) can trace out the detailed behavior of the individual residues more closely than do the more complex calculations (dotted curve) based on statistical thermodynamic considerations.

The highest peaks in the GNM curve are at Ala51, Met47, Glu48, Ala55, and Leu56 in helix A, Leu107 in helix D, and Val74 in helix B. Helix A was pointed out to be the first secondary structure element to acquire protection from exchange, followed by helices D and B *(12)*. A helical subdomain (helices A, D, and B) was also observed to define the structure of the acid state (molten globule form) of RNase H *(62)*. In particular, the highest peak (Ala51) in the GNM curve coincides exactly with the residue exhibiting the strongest protection in the HX experiments *(12)*. Among other regions indicated by GNM to exhibit a strong ($\geq$15*RT*) free energy increase in response to conformational change, we cite Ile7−Thr9 and Gly11−Ser12 in β1, Gly20−Leu26 in β2, Ser36, Gly38, and Tyr39 in β3, Leu49, Ala52, and Ile53 in helix A, Leu67 in β4, Tyr73 in helix B, Leu103 and Trp104 in helix D, Trp118 in β5, and Cys133 and Ala137 in helix E, most of which are also identified by HX to be strongly protected against exchange.

We note that Marqusee and co-workers *(12)* examined the HX behavior of partially folded conformations of *Escherichia coli* RNase H, as a function of denaturant (GdmCl) concentration. Comparing their HX data with previous circular dichroism, fluorescence, and NMR measurements *(63)* and with the structural characteristics of the molten globule state of RNase H under acidic conditions *(62)*, Marqusee and coworkers concluded that the structures exhibiting the highest ΔG values in HX experiments were also those identified by previous studies to form at an early folding stage or to participate in the molten globule form *(12)*.

The dependence of ΔG_{HX} on GdmCl concentration was examined by Chamberlain et al. (*12*) to estimate the contributions of the two processes, local fluctuations and global unfolding, to the observed HX behavior. The parts of the curves insensitive to GdmCl concentration, in the limit of zero-to-low concentrations of denaturant, were attributed to the contribution of local fluctuations, while the decreasing portions of the curves were associated with global unfolding *(12)*. Thus, ΔG_{HX} data were separated into two contributions of comparable magnitude for the two mechanisms of exchange. In our calculations, on the other hand, the free energy change is determined on the basis of conformational fluctuations exclusively. The fact that the HX data can be explained by this simple model hints that some insights into the global unfolding mechanism of a protein might be gained from a rigorous analysis of the conformational fluctuations of the crystal structure. This important point will be revisited in the discussion.

DISCUSSION

Correlation between Residue Fluctuations and HX Data: A Simple Justification. The close agreement between H/D exchange data and our simple analytical approach based on local packing densities suggests that the local intrinsic flexibility, induced by local packing density and topology of tertiary contacts, is the dominant property probed by these experiments. One may argue that the higher flexibility of a given residue, as manifested by the higher amplitude fluctuations obtained by the GNM, leads to a longer residence in the so-called open state, permitting a more ready exchange of the amide hydrogens with solvent. A justification for this follows.

Under steady-state conditions, $K_{op} = k_{op}/k_{cl} = P_{op}/P_{cl}$, where P_{op} and P_{cl} are the equilibrium probabilities of open and closed forms, respectively. In order to establish the connection between this model and our GNM, it suffices to identify the closed conformation with the one in which all residues are at their mean (crystallographic) positions ($\Delta\mathbf{R}_i = 0$), while the open conformation corresponds to a finite departure $\Delta\mathbf{R}_i$ from mean coordinates. Under this definition K_{op} can be expressed in terms of the probability distribution function given by the GNM model as

$$(K_{op})_i = (P_{op}/P_{cl})_i = W(\Delta\mathbf{R}_i)/W(\Delta\mathbf{R}_i = 0) = \exp\{-3(\Delta\mathbf{R}_i)^2/2\langle(\Delta\mathbf{R}_i)^2\rangle\} \quad (8)$$

Substitution into eq 1, while assuming that the EX2 limit holds, yields

$$(\Delta G_{HX})_i = 3RT(\Delta\mathbf{R}_i)^2/2\langle(\Delta\mathbf{R}_i)^2\rangle = \frac{\gamma}{2}(\Delta\mathbf{R}_i)^2/[\Gamma^{-1}]_{ii} \quad (9)$$

which is identical to our eq 7. The last equality in eq 9 follows directly from eq 4 for the particular case $i = j$.

On the basis of this derivation, the fact that the H/D exchange data presently analyzed are satisfactorily explained by the GNM suggests that (i) the EX2 limit, which underlies the connection between the observed kinetic data and the GNM description, applies to the examined data conforming to the well-established view of native proteins HX kinetics *(8, 10, 11)*; (ii) the H/X data for individual residues in the native state or under mild denaturing conditions are strongly correlated with the amplitudes of conformational fluctuations; and (iii) the observed thermodynamic behavior essentially originates in the contribution of the conformational entropy change to the free energy.

Crystallographic B Factors and HX Data. The validity of the above arguments, i.e., strong correlation between the fluctuation behavior of protein crystals and the native-state H/X data, can be directly tested by comparing the crystallographic *B* factors, or more precisely the reciprocal of the *B* factors, with the observed free energy changes. In fact, the GNM basically yields vibrational amplitudes of residues in the specific context or tertiary structure of the investigated proteins.

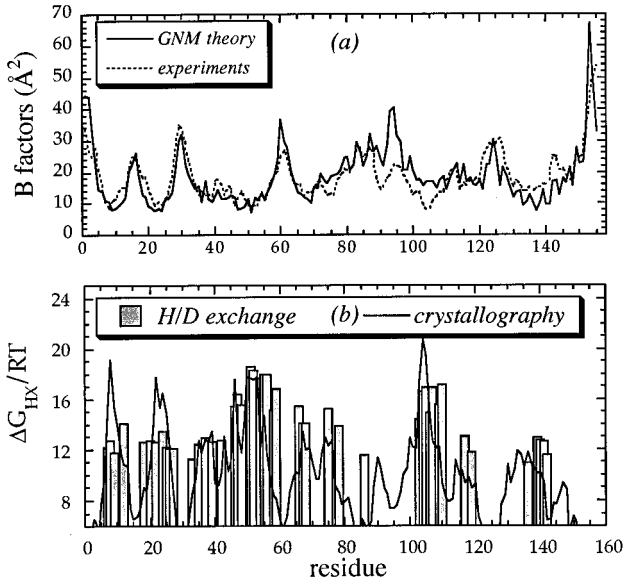

FIGURE 3: (a) Crystallographic *B* factors for RNase H. The solid and dotted curves refer to the theoretical (GNM) and experimental (X-ray crystallography) results. Experimental values are the α-carbon temperature factors reported by Katayanagi et al. *(36)*. The theoretical curve is evaluated from eq 4, for $i = j$, using the α-carbon coordinates of the same structure. The force constant, γ, is suitably rescaled to permit a direct comparison between theory and experiments. (b) Comparison of the crystallographic *B* factors *(36)* and the HX data *(12)* for RNase H, demonstrating the strong correlation between residue fluctuations in the folded state and free energy increases observed in HX experiments.

Among the proteins presently considered, the crystal structure of RNase H, for example, has been determined at high resolution, 1.48 Å *(36)*. The GNM calculated mean-square fluctuations for individual residues in this protein are in excellent agreement with experiments. The results are illustrated in Figure 3, panel a. The crystallographic temperature factors of this protein could thus be directly compared with HX data, after a uniform rescaling to match the absolute free energy changes. The result is shown in panel b of Figure 3. The ordinate scales are suitably chosen so that ΔG values only above a threshold detected by experiments are displayed.

The close agreement between crystallographic temperature factors *(36)* and HX data *(12)* is noteworthy. However, the comparison of the X-ray and HX data for the other four proteins (not displayed) showed that the correlation between the two sets of data is not always as good as that observed in Figure 3b or as good as that obtained with the GNM model in Figures 1 and 2. This is mainly due to inaccuracies in the X-ray measurements of atomic fluctuations, or other effects affecting crystallographic data, such as intermolecular contacts, which are unrelated to the intrinsic conformational properties of the protein. Mean coordinate data in the PDB structures are more precise than are the fluctuations (*B* factors) about mean coordinates; and an analytical estimation of residue fluctuations based on mean coordinates, as in the GNM presently adopted, might be expected to permit a more precise estimation of the conformational behavior of individual residues in folded states.

HX Data and Folding Cores: *Implications on the Significance of Conformational Fluctuations near the Native State*. Woodward and co-workers proposed that the slow exchange region of proteins is also the folding core, on the basis of H/D exchange behavior of BPTI and its mutants *(19, 64)*. That a strong correlation exists between local packing density, and consequently local intrinsic flexibility and H/D exchange measurements, is clear from the above analysis. Whether this implies that those regions, exhibiting small amplitude fluctuations and slow exchange behavior, are also implicated in the early folding pathway of the protein is an issue open to discussion *(65)*. Throughout the above presentation of the five proteins examined, we deliberately pointed out the correspondence between residues exhibiting a high free energy cost of HX and those detected in time-resolved experiments to form at early folding stages. Or, alternatively, residues easily exchanging with the solvent were compared with those formed at late stages of folding. To a certain extent, a correlation between these phenomena is discernible. For example, from their examination of the partially folded structures of RNase H, in connection with several experimental observations, Marqusee and co-workers reached the conclusion that "the thermodynamically most stable regions of a protein are the first to acquire structure" *(12)*. An extensive review of the potential use of HX data for unraveling the structure of folding intermediates was recently presented by Englander et al. *(55)*.

If such a correlation exists, this implies that the distribution of conformational fluctuations near native-state coordinates, demonstrated here to be the major measure probed by the HX study of native proteins, embodies hidden information on the global dynamics of proteins. Thus, one may acquire clues about the folding/unfolding mechanisms or on the global motions relevant to function through a rigorous analysis of the fluctuation behavior in the crystalline state. The potential utility of crystallographic temperature factors for extracting information on the collective dynamics of proteins was emphasized in an extensive review by Lumry *(25)*. It has been advocated by Lumry, for instance, that the temperature factors of the crystal structures include much information about the variability of structures. This is substantially true; however, the extent of agreement between inverse temperature factors and HX data is poorer than that observed between the present GNM kinetics results and the HX data. GNM lends itself to a modal analysis of vibrational spectrum, which may be of fundamental utility in future studies of the global motions in proteins. More importantly, proteins of several hundred residues that cannot be explored fully with atomic molecular dynamics using present computers may be easily analyzed with the GNM approach.

Concluding Remarks. The present calculations simply indicate how easy or difficult are the conformational fluctuations in the vicinity of a given residue, as determined by the extent of residue packing, within the context of the global structure. Hydrogen atoms are small; so fluctuations that provide solvent access are not required to be so large, but they must be concerted somehow with one other to provide ultimate access to the solvent outside the protein. If the protein denatures in total or in part, such a path would be opened. Conceivably, the large free energy losses for some of these exchanges could arise from the requirement for a large number of small, but cooperative, individual changes.

The agreement observed here between the model calculations and the hydrogen exchange free energies is remarkable. The implications are simple. First, residues with the highest

density of surrounding residues have hydrogens that are less easily exchangeable. The slow-exchanging regions are usually located at high number densities of residues. This is not necessarily the same as high atom density. However, these two kinds of densities are likely to be crudely correlated, together even with the hydrogen-bond density in an overall sense. Previously, both the hydrogen-bond density and the atomic density within 7 Å spheres have been implicated in the rate of exchange *(24*, *39)*. Numerically it was reported that 78% of unexchanged and only 5% of exchanged sites for trypsin have an atomic packing density higher than 160 atoms in the 7 Å sphere *(24)*. Second, in addition to local packing density, the chain connectivity and the topology of tertiary contacts in the folded state affect the observed behavior. The GNM approach takes into account the couplings among all residues. Although conformational fluctuations operate on a local scale, the overall cooperativity that underlies these motions might be relevant to global denaturation. The present calculations cannot shed light on the extent of denaturation, local or global. However, it is certainly conceivable that the regions of low residue density and weak interresidue coupling, as determined by the distribution of tertiary contacts in the folded state, might be exactly those that denature more readily than others.

ACKNOWLEDGMENT

We thank R. L. Baldwin for his careful reading of the manuscript and useful suggestions.

REFERENCES

1. Woodward, C., Simon, I., and Tüchsen, E. (1982) *Mol. Cell. Biochem. 48*, 135−160.
2. Loh, S. N., Rohl, C. A., Kiefhaber, T., and Baldwin, R. L. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 1982−1987.
3. Arrington, C. B., and Robertson, A. D. (1997) *Biochemistry 36*, 8686−8691.
4. Woodward, C., and Hilton, B. (1980) *Biophys. J. 32*, 561−576.
5. Kiefhaber, T., and Baldwin, R. L. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 2657−2661.
6. Bahar, I., Atilgan, A. R., and Erman, B. (1997) *Folding Des. 2*, 173−181.
7. Haliloglu, T., Bahar, I., and Erman, B. (1997) *Phys. Rev. Lett. 79*, 3090−3093.
8. Hvidt, A., and Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
9. Qian, H., Mayo, S. L., and Morton, A. (1994) *Biochemistry 33*, 8167−8171.
10. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994) *Proteins*: *Struct., Funct., Genet. 20*, 4−14.
11. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
12. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) *Nat. Struct. Biol. 3*, 782−787.
13. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins*: *Struct., Funct., Genet. 17*, 75−86.
14. Swint-Kruse, L., and Robertson, A. D. (1996) *Biochemistry 35*, 171−180.
15. Flory, P. J. (1976) *Proc. R. Soc. London A 351*, 351−380.
16. Pearson, D. S. (1977) *Macromolecules 10*, 696−701.
17. Tirion, M. M. (1996) *Phys. Rev. Lett. 77*, 1905−1908.
18. Jeng, M. F., and Englander, S. W. (1991) *J. Mol. Biol. 221*, 1045−1061.
19. Kim, K. S., Fuchs, J. A., and Woodward, C. K. (1993) *Biochemistry 32*, 9600−9608.
20. Loh, S. N., Prehoda, K. E., Wang, J., and Markley, J. L. (1993) *Biochemistry 32*, 11022−11028.
21. Hilser, V. J., and Freire, E. (1996) *J. Mol. Biol. 262*, 756−771.
22. Hilser, V. J., and Freire, E. (1997) *Proteins*: *Struct., Funct., Genet. 27*, 171−183.
23. Wallqvist, A., Smythers, G. W., and Covell, D. G. (1997) *Protein Sci. 6*, 1627−1642.
24. Kossiakoff, A. A. (1985) *Annu. Rev. Biochem. 54*, 1195−1227.
25. Lumry, R. (1995) in *Protein*−*Solvent Interactions* (Gregory, R. G., Ed.) pp 1−141, Marcel Dekker, Inc., New York.
26. Miyazawa, S., and Jernigan, R. L. (1985) *Macromolecules 18*, 534−552.
27. Miyazawa, S., and Jernigan, R. L. (1996) *J. Mol. Biol. 256*, 623−644.
28. Bahar, I., and Jernigan, R. L. (1997) *J. Mol. Biol. 266*, 195−214.
29. Bahar, I., and Jernigan, R. L. (1996) *Folding Des. 1*, 357−370.
30. Kloczkowski, A., Mark, J. E., and Erman, B. (1989) *Macromolecules 22*, 1423−1432.
31. Cotton, F. A., Hazen, E. E., Jr., and Legg, M. J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2551−2555.
32. Collyer, C. A., Guss, J. M., Sugimura, Y., Yoshizaki, F., and Freeman, H. C. (1990) *J. Mol. Biol. 211*, 617−632.
33. Wlodawer, A., Walter, H., Huber, R., and Sjolin, L. (1984) *J. Mol. Biol. 180*, 301−329.
34. Bushnell, G. W., Louie, G. V., and Brayer, G. D. (1990) *J. Mol. Biol. 214*, 585−595.
35. Hynes, T. R., and Fox, R. O. (1991) *Proteins*: *Struct., Funct., Genet. 10*, 92−105.
36. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1992) *J. Mol. Biol. 223*, 1029−1052.
37. Sanishvili, R., Volz, K. W., Westbrook, E. M., and Margoliash, E. (1996) *Structure 3*, 707−716.
38. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. J., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
39. Jeffrey, G. A., and Saenger, W. (1994). *Hydrogen bonding in biological structures*. Springer-Verlag, Berlin, Germany.
40. Finucane, M. D., and Jardetzky, O. (1996) *Protein Sci. 5*, 653−662.
41. Tüchsen, E., and Woodward, C. (1985) *J. Mol. Biol. 185*, 405−419.
42. Tüchsen, E., and Woodward, C. (1985) *J. Mol. Biol. 185*, 421−430.
43. Molday, R. S., Englander, S. W., and Kallen, R. G. (1972) *Biochemistry 11*, 150−158.
44. Richarz, R., Sehr, P., Wagner, G., and Wüthrich, K. (1979) *J. Mol. Biol. 130*, 19−30.
45. Wagner, G., Stassinopolou, C. I., and Wütrich, K. (1984) *Eur. J. Biochem. 145*, 431−436.
46. Goldenberg, D. P., and Zhang, J. (1993) *Proteins: Struct., Funct., Genet. 15*, 322−329.
47. Kim, K. S., Tao, F., Fuchs, J. A., Danishevsky, A. T., Housset, D., Wlodawer, A., and Woodward, C. K. (1993) *Protein Sci. 2*, 588−596.
48. Wagner, G., and Wüthrich, K. (1982) *J. Mol. Biol. 160*, 343−361.
49. Kosen, P. A., Creighton, T. E., and Blout, E. R. (1981) *Biochemistry 20*, 5744−5754.
50. Goldenberg, D. P. (1992) *Trends Biochem. Sci. 17*, 257−261.
51. Creighton, T. E., and Goldenberg, D. P. (1984) *J. Mol. Biol. 179*, 497.
52. Elöve, G. A., Chaffotte, A. F., Roder, H., and Goldberg, M. E. (1992) *Biochemistry 31*, 6876−6883.
53. Roder, H., Elöve, G. A., and Englander, S. W. (1988) *Nature 335*, 700−704.
54. Marmorino, J. L., and Pielak, G. J. (1995) *Biochemistry 34*, 3140−3143.
55. Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. (1997) *Protein Sci. 6*, 1101−1109.
56. Chazin, W. J., and Wright, P. E. (1988) *J. Mol. Biol. 202*, 623−636.

57. Moore, J. M., Lepre, C. A., Gippert, G. B., Chazin, W. J., Case, D. A., and Wright, P. E. (1991) *J. Mol. Biol. 221*, 533−555.
58. Dyson, H. J., Sayer, J. R., Merutka, G., Shin, H. C., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol. 226*, 819−835.
59. Wang, Y., and Shortle, D. (1996) *Protein Sci. 5*, 1898−1906.
60. Woodward, C. K. (1994) *Curr. Opin. Struct. Biol. 4*, 112−116.
61. Walkenhorst, W. F., Green, S. M., and Roder, H. (1997) *Biochemistry 36*, 5795−5805.
62. Dabora, J. M., Pelton, J. G., and Marqusee, S. (1996) *Biochemistry 35*, 11951−11958.
63. Yamasaki, K., Ogasahara, K., Yutani, K., Oobatake, M., and Kanaya, S. (1995) *Biochemistry 34*, 16552−16562.
64. Woodward, C. (1993) *Trends Biochem. Sci. 18*, 359−360.
65. Mullins, L., Pace, C. N., and Raushel, F. M. (1997) *Protein Sci. 6*, 1387−1395.

BI9720641