PROTEINS: Structure, Function, and Genetics 36:175–191 (1999)

Conformational Dynamics of Cytochrome c: Correlation to Hydrogen Exchange

Angel E. García* and Gerhard Hummer
Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico

ABSTRACT **We study the dynamical fluctuations of horse heart cytochrome c by molecular dynamics (MD) simulations in aqueous solution, at four temperatures: 300 K, 360 K, 430 K, and 550 K. Each simulation covers a production time of at least 1.5 nanoseconds (ns). The conformational dynamics of the system is analyzed in terms of collective motions that involve the whole protein, and local motions that involve the formation and breaking of intramolecular hydrogen bonds. The character of the MD trajectories can be described within the framework of rugged energy landscape dynamics. The MD trajectories sample multiple conformational minima, with basins in protein conformational space being sampled for a few hundred picoseconds. The trajectories of the system in configurational space can be described in terms of diffusion of a particle in real space with a waiting time distribution due to partial trapping in shallow minima. As a consequence of the hierarchical nature of the dynamics, the mean square displacement autocorrelation function, $\langle |x(t) - x(0)|^2 \rangle$, exhibits a power law dependence on time, with an exponent of around 0.5 for times shorter than 100 ps, and an exponent of 1.75 for longer times. This power law behavior indicates that the system exhibits suppressed diffusion (subdiffusion) in sampling of configurational space at time scales shorter than 100 ps, and enhanced (superdiffusion) at longer time scales. The multi-basin feature of the trajectories is present at all temperatures simulated. Structural changes associated with inter-basin displacements correspond to collective motions of the Ω loops and coiled regions and relative motions of the α-helices as rigid bodies. Similar motions may be involved in experimentally observed amide hydrogen exchange. However, some groups showing large correlated motions do not expose the amino hydrogens to the solvent. We show that large fluctuations are not necessarily correlated to hydrogen exchange. For example, regions of the proteins forming α helices and turns show significant fluctuations, but as rigid bodies, and the hydrogen bonds involved in the formation of these structures do not break in proportion to these fluctuations. Proteins 1999;36:175–191.** **Published 1999 Wiley-Liss, Inc.†**

Key words: nonlinear dynamics; molecular dynamics; protein hydration; hydration dynamics; strange dynamics

INTRODUCTION

Cytochrome c (cyt c) is a heme protein that plays a central role in cell respiration and as a storage and delivery unit for electrons. The overall fold of cyt c is highly conserved in organisms ranging from bacteria to mammals. Cyt c has proven to be a useful model system for the folding kinetics of proteins.[1–6] Recently, Shastry and Roder[6] have described the fluorescence changes associated with refolding of cyt c. Their experiments reveal that folding of cyt c occurs in at least two distinct stages. In the first stage, a collapsed state occurs within 50 μs. This collapse state is separated from the unfolded state by a free energy barrier, independent of solution conditions (e.g., pH, GuHCl, imidazole). The formation of this state is consistent with the energy landscape theory.[7] The first stage of folding is not the rate limiting step for folding of cyt c, which requires several milliseconds under the most favorable conditions. The second stage of folding is characterized by the formation of native tertiary contacts in a hydrophobic core. Shastry and Roder[6] have suggested that this two-step mechanism applies to most proteins, with the exception of some small proteins that can reach the folded state without going through compact intermediates.[8] In addition to protein folding studies, cyt c has also served as a model system in theoretical studies of the dynamics,[9] dielectric properties,[10] reorganization energy,[11] and electron transport[12] in proteins.

Recent experiments on oxidized horse heart cyt c by Bai et al.[13–17] have identified partially unfolded structures in the hydrogen exchange process. Those structures occur as fluctuations covering the range from the folded to the fully unfolded state. From these experiments, the following scenario emerges.[18] The native state of cyt c is stabilized overall by free energies of 10–15 Kcal/mol. The free energy differences between folded and unfolded states can be adjusted experimentally by changing the temperature or solvent conditions (i.e., adding denaturants like guanidinium HCl). These large-scale fluctuations involve the partial unfolding of cooperative units associated with the Ω loops and α helical regions of the protein. Bai et al.[16] have identified hydrogen exchange in the 70–85 and the 36–61 loops to be associated with low-energy (5–7 Kcal/mol)

Grant sponsor: U.S. Department of Energy, Los Alamos National Laboratory; Grant number: LDRD-IP grant ER-99006.
*Correspondence to: Angel E. García, Theoretical Biology and Biophysics Group, T10, MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545. E-mail: angel@t10.lanl.gov
Received 1 December 1998; Accepted 26 March 1999

partially unfolded states, while the α helices in the N and C termini, and the 60's helix exchange slowly, characterized by high-energy (9–11 Kcal/mol) fluctuations to the unfolded state. Similar experiments on reduced cyt c have confirmed these findings.[19]

Hydrogen exchange in proteins depends on fluctuations in their structure[20] and on their interactions with the aqueous solvent. Studies of the dynamics and relaxation of proteins have shown that transitions among conformational substates are important in protein function.[21] Molecular dynamics simulations of proteins at room temperature have revealed that protein fluctuations can be described by a small set of collective motions involving large regions of the protein. Those regions typically involve intact elements of the secondary structure present in the native state. Dynamical modes that describe most of the Cartesian coordinate fluctuations of proteins and other biomolecules in MD simulations[22–24] have been shown to be nonlinear in the following sense: trajectories sample multiple conformational minima, with each basin in protein conformational space being visited for periods of about 100 ps.

The dynamics of a protein near its native conformation can be classified into three types:[24] *unimodal, multimodal fluctuation*, and *relaxational*. *Unimodal dynamics* describes small fluctuations within a single local minimum. Unimodal motions can be described by harmonic and quasi-harmonic dynamics. Simulations of native proteins well below room temperature may exhibit this behavior. *Multimodal fluctuation dynamics* describes motions close to equilibrium involving many local minima that are close in configurational space and energy. The energy barriers separating these states are of the same order of magnitude as k_BT, where k_B is the Boltzmann constant and T is the temperature in Kelvin. This is the dynamical behavior expected from a folded protein below and close to the denaturation temperature. These motions cannot be described by harmonic dynamics. Details of the dynamics and kinetics will depend on the roughness of the energy landscape.[7,21] *Relaxational dynamics* have the characteristics of multimodal fluctuation dynamics, but, when the initial state is far from equilibrium, will also exhibit a *drift* from the initial metastable state to states closer to equilibrium. We expect to observe this dynamical behavior in proteins well above their denaturation temperature. The protein should unfold at these temperatures but in a time scale longer than simulation times of 1 ns.[25–27] Given the short time of typical simulations in comparison with the unfolding time, higher temperature simulations will allow only a glance at higher energy fluctuations starting from the folded state.

The goal of this work is to explore the dynamical fluctuations of cyt c, near the folded state at room temperature and in the early unfolding stages at higher temperatures, and to correlate the character of these fluctuations to experimentally observed hydrogen exchange patterns. For this purpose, molecular dynamics simulations for periods larger than 1.5 ns are performed at four temperatures, 300, 360, 430, and 550 K, that represent the protein motions under native state conditions, slightly above the experimentally measured melting temperature, at a temperature above melting where the hydrophobic effect is large and mostly enthalpic,[28–30] and at a temperature where unfolding may occur in the nanosecond time scale, respectively.

METHODS

Description of the System and Simulations

Horse heart cytochrome c in its oxidized state is simulated in the presence of water and excess salt. We use the structure of horse heart cyt c in solution and in its reduced state (pdb code: 1frc)[31] as the initial configuration of oxidized cyt c near its folded state. The NMR structure for oxidized cyt c was reported by Qi et al.[32] but coordinates were not available. The crystal structures of reduced and oxidized yeast cyt c show a C_α root mean square distance (*rmsd*) of 0.34 Å.[33] However, the solution structure of reduced cyt c, determined by NMR, was quite different from the corresponding crystal structure, showing a C_α rmsd of 2.37 Å. Small angle x-ray scattering studies have shown that the differences between the reduced and oxidized forms in solution are small in the presence of 200 mM sodium chloride.[34] Therefore, differences in structures due to crystal versus solution conditions are larger than differences between the oxidized and reduced states in the crystals. This suggested that the available NMR structure supplies a better model for the solution structure of oxidized cyt c than the crystal form of the oxidized protein. During the progress of this work, Banci et al.[35] published a structure of the oxidized state and found their structure to be different from that of Qi et al.[32] and quite similar to the x-ray structure[33] and the NMR structure of the reduced state.[31] From a comparison between the Protein-Data-Bank-available NMR structure[31] with their structure, Banci et al.[35] found very similar secondary structures, but with a major difference in the backbone conformation in the 32–38 region of the Ω_1 loop.

A covalent bond between Met-80 and the heme iron is maintained in all the simulations. Time resolved spectroscopy experiments have estimated that the rate of breaking of this bond is 10^4–10^5 (at T = 313 K, under denaturing conditions). That is, the lifetime of this bond is 40 μs,[26,36] and therefore, it can be assumed that it remains formed at all times during our ns simulations. His-33 is modeled in its protonated state;[10,37] His-18 and His-26 are modeled as neutral with hydrogens at the δ position. Six internal water molecules found in the solution NMR structure are included in the initial configuration.[31] The total charge of the protein in the oxidized state is +9e. We also include 20 Cl^- and 11 Na^+ ions in the system. Ions are initially located at random within the aqueous solution, but at least 5 Å away from other ions or protein atoms. Water molecules are added around the protein to fill a cubic box with ≅ 57 Å on the side. The walls of this box are at least 11 Å away from any protein atom to allow the protein system to sample noncompact (partially unfolded) conformations at high T. The resulting system has 5,760 water molecules, 31 ions, 1,746 atoms in the protein, and a total of 19,057

atoms. The resulting concentrations for the system are $[Cl^-] = 0.18$ M, [cyt c] = 0.8 mM, and $[Na^+] = 0.10$ M.

We use the all-atom force field of Cornell et al.[38] Electrostatic interactions are modeled with the particle mesh Ewald (PME)[39] algorithm implemented in Amber 4.1.[40] Direct interactions are cut-off at 9 Å. The Fourier space part of the Ewald potential is calculated on a cubic grid of 48 points on the side and interpolated over all space with a cubic spline. For the T = 300 K trajectory, we simulate the system at constant (N,T,P; P = 1 atm). For T = 360 K and 430 K trajectories, we scale the density proportionally to the density of water at the liquid-vapor coexistence temperature.[30] For the T = 550 K trajectory we maintain the density as in the T = 430 K simulation. These calculations are performed at constant (N,T,V). The solvated system is subjected to 1,000 steepest descent energy minimization cycles. The resulting configuration is simulated at T = 100 K for 100 ps. The temperature is then gradually increased to 300 K over a 50 ps time interval. This simulation is extended to 2.5 ns. The configuration at t = 1 ns of the T = 300 K simulation is used to start the T = 360 K calculation. At this time, the volume of the box is increased by 2%. The system is slowly heated to T = 360 K over 100 ps and further equilibrated for another 200 ps. This simulation is extended to 1.75 ns. The configuration at t = 100 ps of the T = 360 K simulation is used to continue heating the system to 430 K over a period of 100 ps, followed by a further equilibration for another 100 ps. The volume of the system is increased by 3.1% over the volume at T = 300 K. The simulation is then extended to 1.6 ns at T = 430 K. The configuration at t =100 ps of the T = 430 K simulation is used to continue heating the system to 550 K over a period of 100 ps, followed by a constant temperature run of 1.6 ns. The last 1.5 ns of each trajectory are used in all the analyses described below. Configurations are saved for further analysis at a rate of two per ps.

Optimal Dynamic Coordinates

To study the system fluctuations we use a set of directions, **m**, in the 3*N*-dimensional space of the protein, that best represent (in a least-square sense) the ensemble of protein structures observed in the simulations. The construction of these *optimal dynamic coordinates*, referred to as optimal modes, has been previously described.[22–24,41] Motions along these directions show multi-centered oscillations, rapid transitions from one center to another, and damped quasi-harmonic oscillations around each center.

Optimal coordinate directions, *m*, are obtained by solving an eigenvalue equation

$$\sigma \cdot \mathbf{m} = \lambda \mathbf{m}, \tag{1}$$

where the positive semi-definite matrix $\sigma_{\alpha,\beta}$ is defined by

$$\sigma_{\alpha,\beta} = \frac{1}{S}\sum_{i=1}^{S} (r_i - y_0)_\alpha (r_i - y_0)_\beta. \tag{2}$$

Here, $y_0 = 1/S\ \Sigma^S\ \mathbf{r}_i$ is the average configuration, S is the number of configurations, $\mathbf{r}_i$, and α and β refer to Cartesian components of the 3N protein atomic coordinates. The eigenvectors $\mathbf{m}_k$ that minimize the mean square distance (MSD), d^2 of configurations from a vector are given by

$$d^2(\mathbf{m}_k) = Tr(\sigma) - \lambda_k. \tag{3}$$

The eigenvector, $\mathbf{m}_k$, corresponding to the largest eigenvalue, λ_k, can be regarded as the direction of the line that passes through the average conformation, $\mathbf{y}_0$, which best represents the predominant motions in the protein. $Tr(\sigma) = \Sigma_{i=1}^{3N} \sigma_{i,i}$ is the trace of σ. Optimal coordinates are systematically ranked in the order of decreasing eigenvalues. The mean square fluctuations are given by $(1/N)\ Tr(\sigma) = (1/N)\Sigma_i\lambda_i$.

A generalization of this method to represent 2-dimensional and 3-dimensional projections of the configurational space, as planes and volumes that better represent the dynamics of the system, has been presented previously.[41,42] Similar methods have been introduced by Amadei, Linssen and Berendsen[43] and by Hayward and Go.[44] These coordinates are specific to the molecule and trajectory sampled during a MD simulation and can be seen as principal component axes of the protein configurational space along a given trajectory.

Mean Square Displacement Autocorrelation Function

With the above formalism, we can describe the MD trajectory of the protein in configurational space as a particle diffusing in real space. The protein configuration is considered a stochastic variable, *x(t)*, moving with a random velocity, *ζ(t)*, in the reduced space of the optimal modes. (Here *x(t)* refers to the projection of the protein coordinates on one of the optimal coordinates. All indices labeling the directions are omitted.) These modes are coupled strongly to a heat bath consisting of the other, possibly quasi-harmonic, modes and the solvent. Defining the displacements

$$x(t) = \int_0^t \zeta(t')\, dt' + x(0), \tag{4}$$

the ensemble average of the mean square displacement (*MSD*) is

$$\langle x^2(t)\rangle = \int_0^t dt' \int_0^{t'} dt'' \langle \zeta(t')\zeta(t'')\rangle, \tag{5}$$

where we assume no correlations between *x(0)* and *ζ(t)* (i.e., $\int_0^{t'} \langle \zeta(t')x(0)\rangle = 0$). For a stationary distribution, the velocity–velocity correlation function satisfies

$$\langle \zeta(t')\zeta(t'')\rangle = \langle \zeta(|t' - t''|)\zeta(0)\rangle = \tilde{C}(|t' - t''|). \tag{6}$$

The time derivative of the *MSD* is

$$\frac{d}{dt}\langle x^2(t)\rangle = 2\int_0^t \langle \zeta(t')\zeta(0)\rangle\, dt'. \tag{7}$$

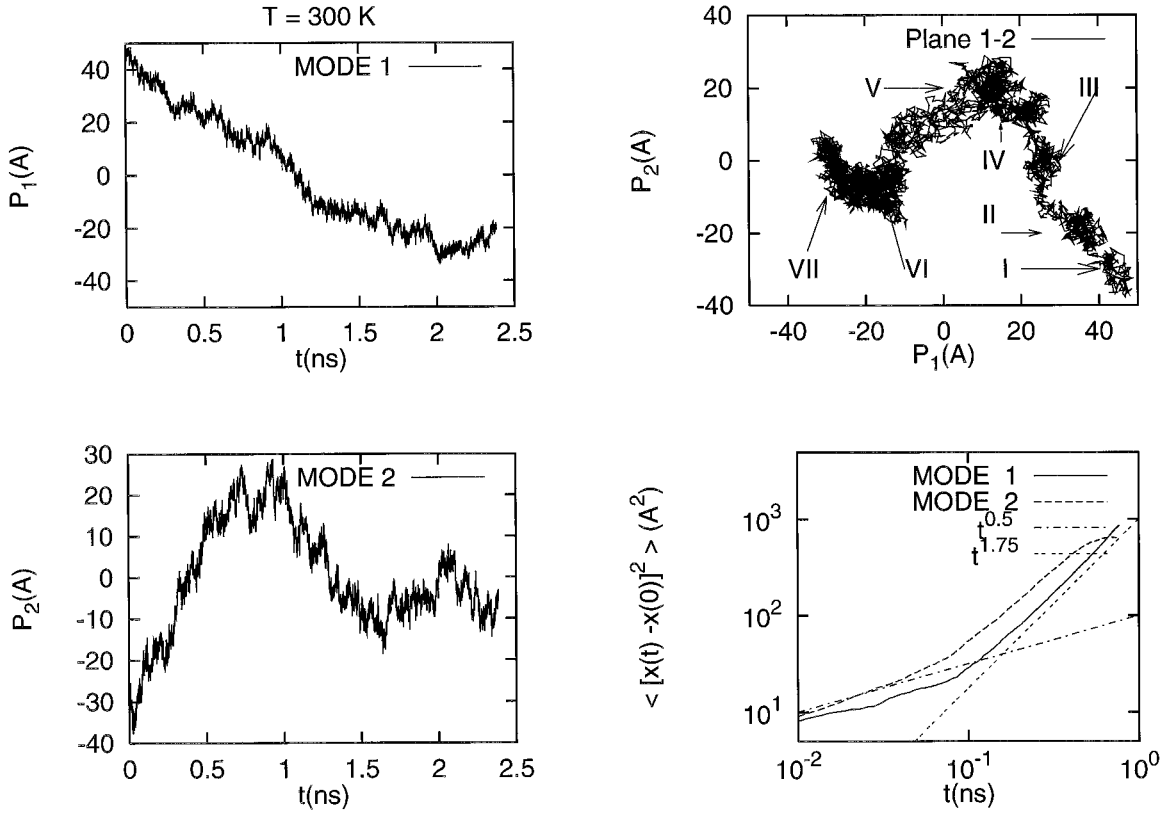

Fig. 1. Optimal dynamic coordinates for cyt c at T = 300 K. The left-hand-side plots show the projections of the 2.5 ns MD trajectory along the two modes that describe most of the system fluctuations (the 1 ns equilibration period is included in this analysis, but not in the others). The top right-hand-side plot shows the projection on the principal plane spanned by the best two optimal modes. This projection establishes that cyt c samples multiple minima. Transitions from one basin to another occur on a time scale of 0.1 ns. The bottom right-hand-side plot shows a log–log plot of the mean square displacement autocorrelation function, and two simple power law curves with exponents 0.5 and 1.75. For a normal diffusive process, the *MSD* autocorrelation function should exhibit a linear time dependence in the long time limit. From this curve we see that motions in the sub 0.1 ns time scale sampled a smaller space than in normal diffusion, while at longer time scale the system samples a larger configurational space than in normal diffusion.

Thus, the integrand on the r.h.s. of Equation (7) is exactly $\tilde{C}(t')$, the correlation between the velocities of the system at time t and $t + t'$ averaged over a long period of time t. For regular diffusion the r.h.s. of Equation (7) is a constant that is usually identified as being proportional to the diffusion coefficient, D. Equation (7) illustrates the direct relation between the diffusive behavior and the correlation function, $\tilde{C}(t)$. When this function decays exponentially or faster for large values of t, the integral is finite and the particle (or protein) exhibits a regular diffusive behavior. In contrast, for a system diffusing anomalously, $\tilde{C}(t)$ may have a long algebraic tail of the form

$$\tilde{C}(t) \simeq \kappa/t^{\alpha}. \tag{8}$$

In such a system the MSD is characterized by

$$\langle x^2(t)\rangle \sim t^{2H_D}, \tag{9}$$

where $H_D = 1 - \alpha/2$. The exponent H_D is the Hölder exponent, which in the case of simple Brownian motion, has the value ½. Values of $H_D > ½$ and $H_D < ½$ correspond to *super-diffusion* and *sub-diffusion*, respectively. Trapping in configurational space gives rise to sub-diffusion, whereas occasional long jumps of the system in configurational space give rise to super-diffusion.

Hydrogen Bond Formation

To study the correlation between thermal fluctuations and hydrogen exchange we monitor the probability of formation and mean lifetime of hydrogen bonds. We use a geometrical criterion for hydrogen bonds, where bonds are considered to be formed when the distance between the donor and the acceptor is ≤ 3.2 Å and the angle between donor, hydrogen, and acceptor is $\geq 120°$. We monitor time series of the instantaneous formation of hydrogen bonds, as well as 10 ps block averages of the hydrogen bond formation. We monitor alpha helices (i, $i + 4$), 3–10 helices (i, $i + 3$), π helices (i, $i + 5$), and (i,$i + 2$) backbone hydrogen bonding patterns for all trajectories.

Hydration Structure

The local hydration of each amino group is monitored by the radial distribution functions of water molecules

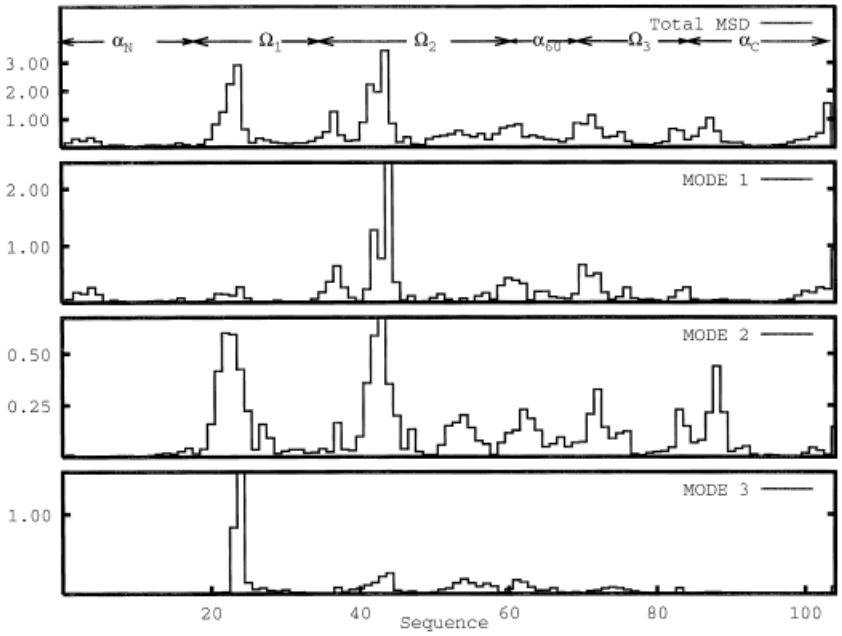

Fig. 2. Total mean square displacements (*MSD*) along the three optimal modes (modes 1, 2, and 3) for each of the C_α atoms in cyt c at 300 K. Most of the motions occur in the loop region before the C-terminal helix and after the N-terminal helix. The N-terminal helix shows the smallest fluctuations. The *MSD* are given in $Å^2$.

Fig. 3. Superposition of the structure attained at three major basins sampled during the T = 300 K simulations and represented by displacements along the two optimal modes from the average sampled configuration. The protein secondary structure elements have been color-coded following Bai et al.:[16] the 70–85 loop (red), the 36–61 loop (yellow), the 20–35 loop (green), the 60's helix (green), and the amino-terminal helix and the carboxyl-terminal helix (blue). Bai et al.[16] used the same color code to identify regions showing similar rates of hydrogen exchange. Regions exchanging faster were colored red, followed by yellow, green, and blue for the slowest.

around the hydrogen atoms. Amino groups involved in intramolecular hydrogen bonds do not show water molecules closer than 3.0 Å from the hydrogen atoms. However, when intramolecular hydrogen bonds open, water molecules will form hydrogen bonds to the amide hydrogen (HN) of amino groups with water oxygen (OW) distances around 2 Å. The radial distribution functions (rdf) are calculated from histograms of the number of water molecules within spheres of increasing radius centered at the reference atoms. These distributions are then normalized by the number of waters uniformly distributed within the spherical shells around a reference point. The coordination

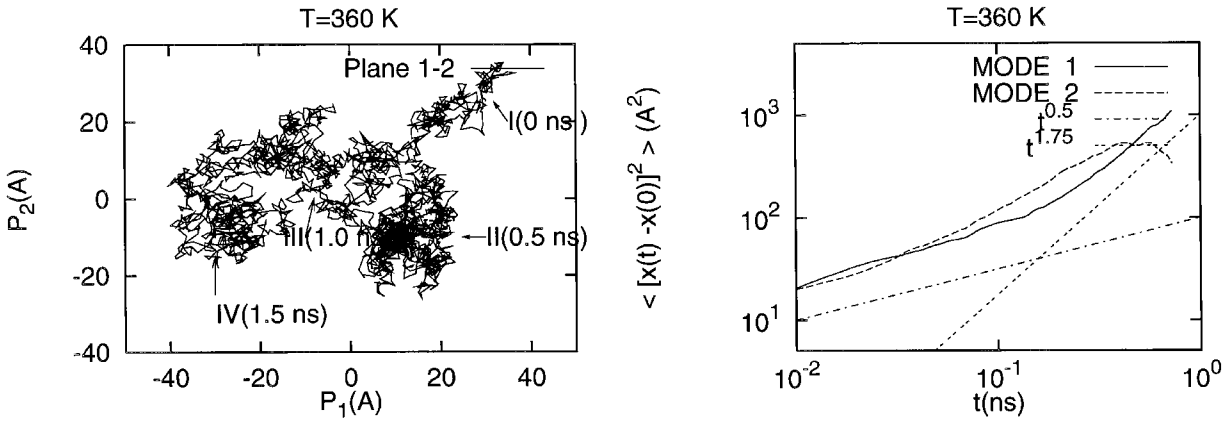

Fig. 4. Projection of the cyt c MD trajectory at T = 360 K on a plane spanned by the two optimal modes. This projection establishes that cyt c samples multiple minima. The right-hand-side plot shows a log-log plot of the mean square displacement autocorrelation function, and two simple power law curves with exponents 0.5 and 1.75.

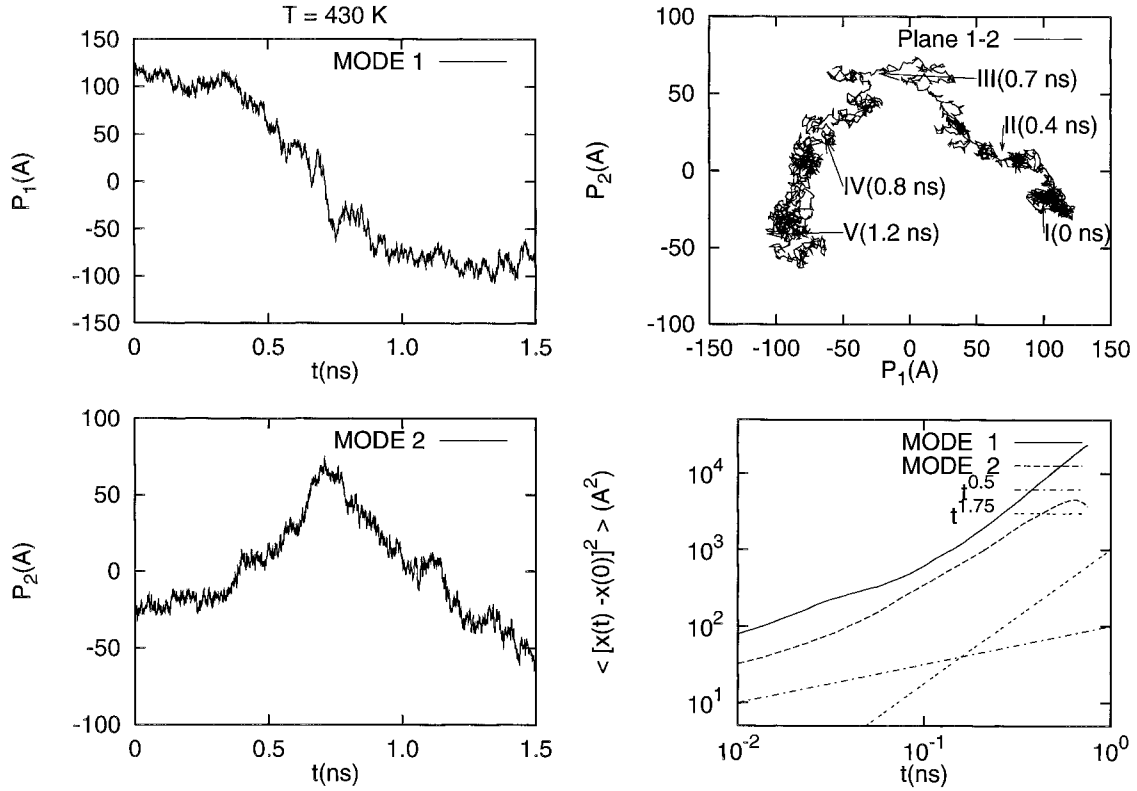

Fig. 5. Optimal modes for cyt c at T = 430 K. The left-hand-side plots show the projections of the 1.6 ns MD trajectory along two optimal modes. the top right-hand side plots show the projections onto the two principal planes. The left-hand-side plots show a large concerted motion (almost 0.6 Å *rmsd*) of modes 1 and 2. This motion results in a jump from one basin (see plane 1–2, r.h.s.) at t = 0 to another at t = 1.2 ns. This jump occurs via an intermediate state at t = 0.7–0.8 ns. The bottom right-hand-side plot shows a log–log plot of the mean square displacement autocorrelation function and two simple power law curves with exponents 0.5 and 1.75.

number of water molecules around a group is obtained by integrating the radial distribution function within a sphere that encompasses the first hydration shell. We choose the radius of this shell such that the corresponding coordination number of water around the charged N-terminal domain group is one at 300 K (i.e., r = 2.5 Å). The probability of other backbone amide hydrogens forming hydrogen bonds with water molecules is calculated relative to the exposure of this reference group. We report water coordination numbers for all backbone amide hydrogens. The coordination number of water in the first hydration shell of amide groups hydrogen can be used as a

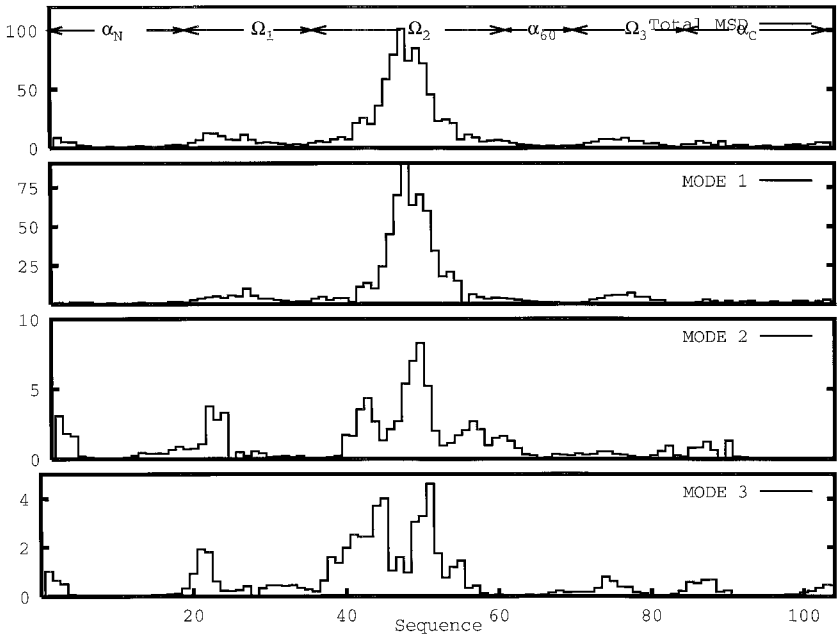

Fig. 6. Total mean square displacements, in Å^2, sampled at T = 430 K (Total *MSD*) and along three optimal modes (modes 1, 2, and 3) for each of the C_α atoms in cyt c. Most of the motions occur in the 36–61 Ω loop. The first mode shows a motion with a *rmsd* of 9 Å, the second mode shows a motion with a *rmsd* of 3.0 Å.

measure of the exposure factor, where

$$P_\alpha(T) = 1 - E_\alpha(T) = 1 - \frac{N_\alpha(R, T)}{N_1(R, T)}. \qquad (10)$$

Here α refers to the amide proton index, $N_\alpha(R,T)$ is the coordination number for the site α within a shell of radius R at temperature T; E_α is the exposure factor relative to the exposure of the (fully exposed) N-terminal group, and $P_\alpha(T)$ is the protection factor. Within this model, the relative protection factor described above is assumed to be comparable to the measured hydrogen exchange protection factors, but not identical.

RESULTS and DISCUSSION

Collective Motions

The displacements along the first two optimal modes of the MD trajectory at 300 K are shown on the left-hand-side panels of Figure 1. The motions along these modes are strongly correlated, as can be seen from the projection of these motions on a plane spanned by these two modes, on the top right-hand-side panel. We observe that the trajectories sample multiple minima (labeled I to VII), with each basin being sampled for a few hundred picoseconds. The bottom right hand side panel shows the log–log plots of the *MSD* function, $\langle |x(t) - x(0)|^2 \rangle$, along the two optimal modes and power law curves with exponents of 0.5 and 1.75. This autocorrelation function identifies the motion along the optimal modes as one dimensional random (but not Brownian) walks. Two features, within this random walk context, must be emphasized from these curves. First, for times shorter than 100 ps this curve shows a power law with an exponent less than unity (0.5). An exponent less than unity implies that the random walk covers less configurational space than a Brownian motion, indicating a strong suppression of diffusion (sub-diffusion). Second, for longer times, the curve shows a power law with exponent larger than unity (1.75). An exponent larger than unity implies that the random walk covers more configurational space than a Brownian walk (super-diffusion), indicating a faster, well-concerted motion that occurs in a longer time scale.

The observed inter-basin hopping motions are better understood in terms of the protein structure by plotting the mean square displacement of every C_α atom along the chain for all modes (labeled total *MSD*) and the first three optimal modes at T = 300 K (Fig. 2). The first three optimal modes describe 28%, 16%, and 7% (a total of 51%) of the system mean square fluctuations. The total *MSD* of the system during the simulation is 1.17 Å^2. Structural changes associated with inter-basin displacements are collective changes (with a *MSD* of 3 Å^2) at the Ω_1 and Ω_2 loops, in the coiled regions, and small relative motions of the α-helices as rigid bodies (with *MSD* smaller than 1 Å^2). A superposition of the conformations adopted around the three major basins sampled during the simulation at 300 K are shown in Figure 3. *Notice that the α-helical regions are displaced as rigid bodies (i.e., without distortion), while the geometry of the Ω loop regions is distorted.*

The fluctuations of cyt c during the T = 360 K MD trajectory are similar in character to those of the T = 300 K MD trajectory. The projection of the 360 K MD simulation trajectory on the plane spanned by the two optimal modes is shown in Figure 4. At 360 K, the protein also samples multiple minima, with transitions occurring in the 100 ps time scale. The MSD autocorrelation function shows a power law behavior, similar to the one observed at T = 300 K.

The *MSD* for the protein at 360 K is 1.64 Å^2 and the first three optimal modes describe 24% , 11%, and 9% of the

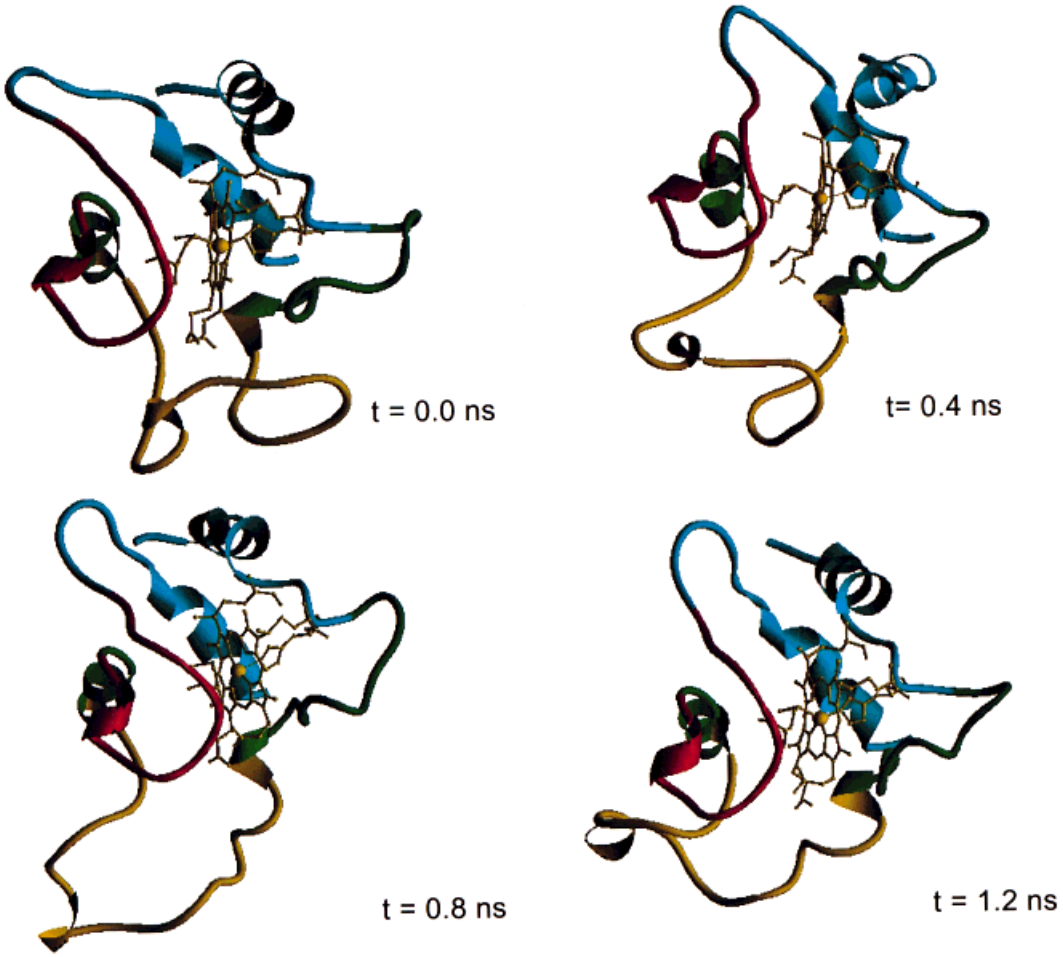

Fig. 7. Configurations sampled at times t = 0, 0.4, 0.8, and 1.2 ns along the T = 430 K MD simulation. The intermediate basin sampled in the transition from the initial basin (t = 0.0 ns) to another basin (t = 1.2 ns) has the 36–61 Ω loop in an extended conformation. Most hydrogen bonds at the tip of the loop are maintained. The loop returns to a packed conformation in contact with the rest of the protein, but in a different configuration. The *rmsd* displacement between the heme and the three α helices is 1.7 Å.

fluctuations. Changes associated with these modes are a 4 Å^2 *MSD* for the Ω_3 loop, mostly described by mode 1, and 2 Å^2 *MSD* for the Ω_1 and Ω_2 loops, distributed equally among the three modes. *At 360 K, as in the simulation at 300 K, the α-helical regions are displaced as rigid bodies (i.e., without distortion), while the geometry of the Ω loops regions is distorted.*

The projection of the T = 430 K MD trajectory on the optimal coordinates, shown in Figure 5, is different from the dynamics at lower temperatures (300 K and 360 K), in terms of the amplitude of the structural changes described, but similar in the sense that multiple basins are sampled over time scales longer than 100 ps, and small overdamped oscillations are sampled at short times. At this temperature, a basin near the folded state is sampled for over 400 ps. During the next 500 ps, a transition from the native state basin to another basin occurs, resulting in a partial unfolding of the protein. The protein remains near this basin for the remaining 600 ps of the simulation. The *MSD* of every C_α atom along the chain for all modes and the first three optimal modes are shown in Figure 6. The total *MSD* is 12.3 Å^2, with each one of the first three modes describing 62%, 9%, and 5% of the fluctuations. The *MSD* at T = 430 K is almost an order of magnitude larger than the *MSD* observed at the 300 K and 360 K simulations. The conformational changes described by these modes occur via an intermediate state where the 36–61 Ω_2 loop opens and later folds into a different conformation. Structural plots of cyt c along this trajectory are shown in Figure 7.

The largest fluctuations observed, in all our simulations, occur in the Ω loops region, color-coded yellow (Y), and not in the red (R) region, as suggested by the hydrogen exchange experiments.[16] However, our simulations agree with the proteolysis experiments of Wang et al.,[45] where proteinase K initially attacks a Y region, rather than the R region. Fontana et al.[46] have also reported that thermolysin cleaves cyt c at a bond in the Y region, in the presence of the helix inducer, TFE. The influence of the Met80–Heme iron covalent bond on the fluctuations of the red region needs to be explored before more definite conclusions are drawn in this respect. Notice that all α helices in the protein remain intact

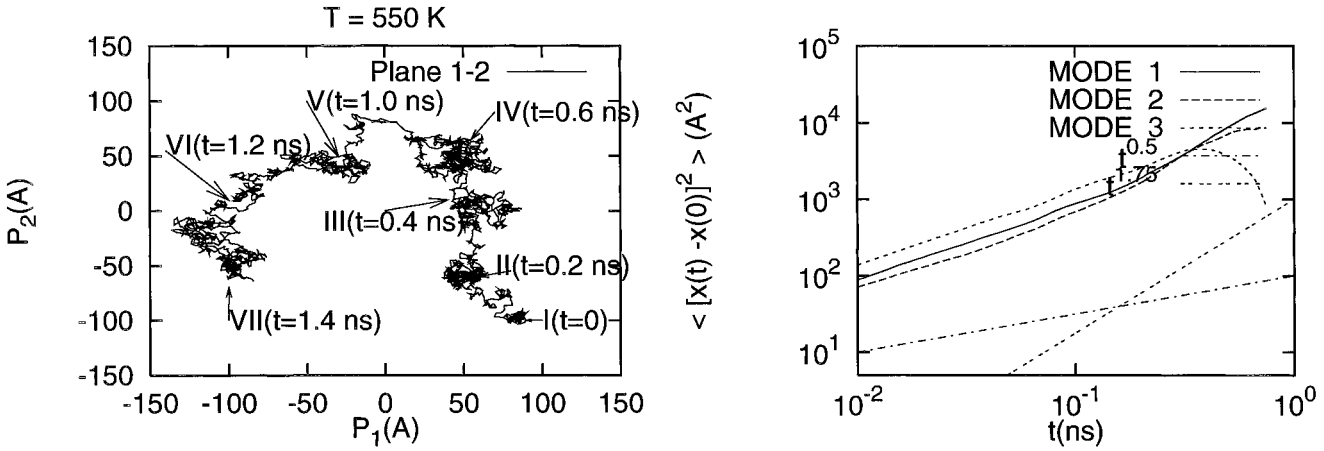

Fig. 8. Optimal modes for cyt c at T = 550 K. The left-hand-side plots show the projections on the principal plane spanned by the first and second optimal mode. The right-hand-side plot shows a log–log plot of the mean square displacement autocorrelation function, and two simple power law curves with exponents 0.5 and 1.75.

during the T = 430 K simulation. We should expect the protein to unfold at this temperature, but in a longer time scale.

To explore the possibility of unfolding cyt c in a nanosecond time scale we simulate the system at 550 K. This choice of temperature is motivated by previous studies of temperature-induced unfolding of proteins by MD simulations where unfolding was observed.[47–49]

The projection of the T = 550 K MD trajectory on the optimal coordinates is shown in Figure 8. This trajectory is similar to the T = 430 K trajectory and different from the dynamics at lower temperatures (300 K and 360 K). At 550 K, a basin near the folded state is sampled for a short period of about 100 ps. Multiple transitions to local basins occur during the 1.5 ns simulations. The total *MSD* is 16.9 Å^2, with each one of the first three modes describing 32% , 15%, and 8% of the fluctuations. The alpha helices are disrupted. Description of the dynamics of hydrogen bonds involved in secondary structure formation is described next.

Hydrogen Bond Opening and Closing and Backbone Hydration

Amide groups hydration

Molecular dynamics simulations of proteins with explicit treatment of water molecules offer an opportunity to study the protein hydration structure and dynamics. We use radial distribution functions and coordination numbers as references for describing the hydration structure around all amino hydrogens in the protein backbone. Figure 9 shows the coordination number and radial distribution function of water averaged about the hydrogen atoms of the N-terminal amino group in cyt c (at T = 300 K) as a function of the radial distance from the reference atom center. The radial distribution functions show that water is organized around the hydrogen atoms by forming a hydrogen bond with a mean distance of 2 Å. The running integral shows that the number of water molecules within a 2.4 shell surrounding the hydrogen atoms contains, on average, one water molecule. The exposure of the amino group hydrogens to water depends on the participation of the amino groups in hydrogen bonding, and different groups exhibit different behavior as a function of temperature. The radial distribution functions for selected groups in cyt c, as a function of temperature, are shown in Figure 10.

The coordination numbers, *N(R)*, of the first coordination shell of all backbone amide hydrogens in cyt c obtained from the 1.5 ns molecular dynamics simulations at T = 300, 360, 430, and 550 K are shown in Table I. The secondary structure assignments are listed for reference. The calculated coordination numbers of water molecules in the first hydration shell of backbone amide hydrogens at 300 K show little or no exposure ($N(R) \leq 0.001$) for amino acids 5–15, 63–70, 89–90, and 96–102 identified with α helices, and amino acids 18, 35–36, 49, 52, 55–57, 73–75, and 80, identified with Ω loops or turns. With the exception of Gly 6, all these amino hydrogens were experimentally observed to exchange with protection factors between 10^3 and 10^9.[16,50] However, other NHs with measured protection factors within this range were exposed to solvent, with $N(r) \geq 0.1$, during the T = 300 K simulation. These are the backbone NHs in amino acids 19, 29–33, 37–39, 42–43, 53–54, 59–60, 79, 91–93, and 95. Some of these variations may be a consequence of the starting structure. For example, Banci et al.[35] found major backbone conformational differences between their structure and the structure of Qi et al.,[31] (which is used in this work), to be in the amino acid stretch 32–38—one of the regions where we find discrepancy between the measured hydrogen exchange data and our calculated solvent exposure probability. However, for some of these hydrogen atoms, the residence time of water molecules[51] may also influence the protection factor.

Figure 11 shows the cyt c–water proximity correlation function and coordination number, calculated during the 1.5 ns production period of MD simulations at T = 300,

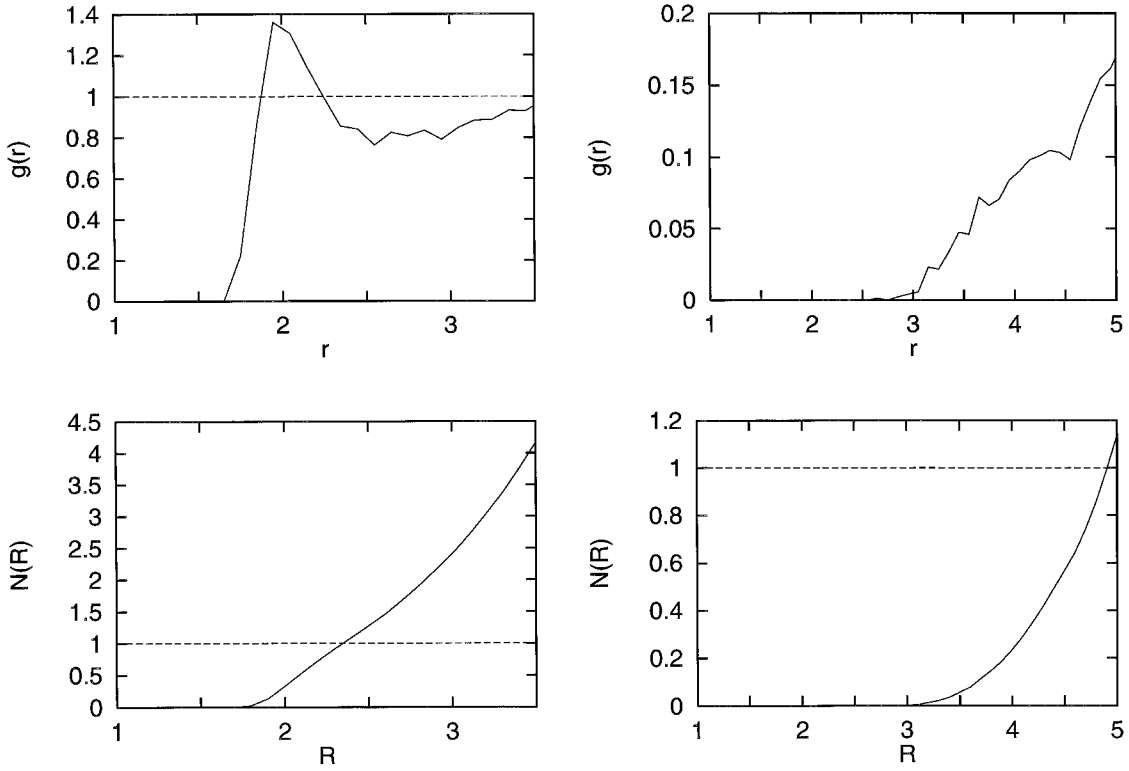

Fig. 9. Radial distribution functions (**top**) and running integrals of the coordination number (**bottom**) of water molecules averaged about the hydrogen atoms of the charged N-terminal amino group (**left**) and one amide group that does not coordinate to water, Gln 12 (**right**) in cyt c, obtained from 1.5 ns simulation at 300 K. The charged N-terminal amino groups are exposed to the solvent and make hydrogen bonds to water molecules. Gln 12 is involved in an intramolecular hydrogen bond with Lys 8.

360, 430, and 550 K.[52] The proximity correlation function is plotted as a function of the distance between water and the closest protein atom. The protein–water radial distribution functions show two peaks at r = 1.75 Å and at r = 2.75 Å, indicating ordering of water molecules on the protein surface. The peak at 1.75 Å represents the correlation between water molecules and protein hydrogen bond donors. The height of this peak decreases with increasing temperature, with the peak practically disappearing at T = 550 K. The second peak at 2.75 Å represents the correlation between water molecules and hydrogen bond acceptors in the protein. The height of this peak decreases slightly with increasing temperature but a drastic change in this peak is observed for T = 550 K, where the peak is much wider. Figure 11 also shows the proximity water coordination number (N(r)) for cyt c, at various temperatures, as a function of the distance of closest approach between water and protein atoms. The y-axis shows the total number of water molecules contained within this layer of thickness r surrounding the protein surface. The total number of water molecules accepting hydrogen bonds (integrated to r = 2.25 Å) with the protein are 99, 90, 82, and 77, for T= 300, 360, 430, and 550 K, respectively. The number of water molecules participating in direct hydrogen bonds with the protein (integrated to 3.0 Å) are 408, 383, 368, and 386, for T = 300, 360, 430, and 550 K, respectively. The total number of coordinated water molecules at T = 550 K is larger than at T = 430 K, due to partial unfolding of the protein. The correlation between the protein overall hydration and the protein dynamics will be discussed below.

Backbone hydrogen bonding

Figure 12 shows the time series of the formation of hydrogen bonds for α and 3_{10} helices in the N-terminal helix of cyt c during the 300 K simulation. Local fluctuations allow hydrogen bonds to transiently break and reform. We see that, although some hydrogen bonds are formed up to 98% of the time, the longest time that a particular hydrogen bond is present continuously varies from 19 to 800 ps. However, the number of water molecules forming hydrogen bonds with these amino group hydrogens is *zero*. The detailed dynamics of individual hydrogen bonds can be complicated, but in general, hydrogen bond breaking does not imply that the amide hydrogen is exposed to water molecules. The dynamics of hydrogen bond breaking and formation follows a similar pattern at higher temperatures (T = 360 K and T = 430 K). Therefore, we conclude that geometrical definitions of hydrogen bond opening/closing are not a good measure of hydrogen exchange probability. The geometrical criteria used to define a good hydrogen bond can be violated for periods of time short relative to the time it takes water to diffuse into the first hydration shell of the exposed atom. The opening

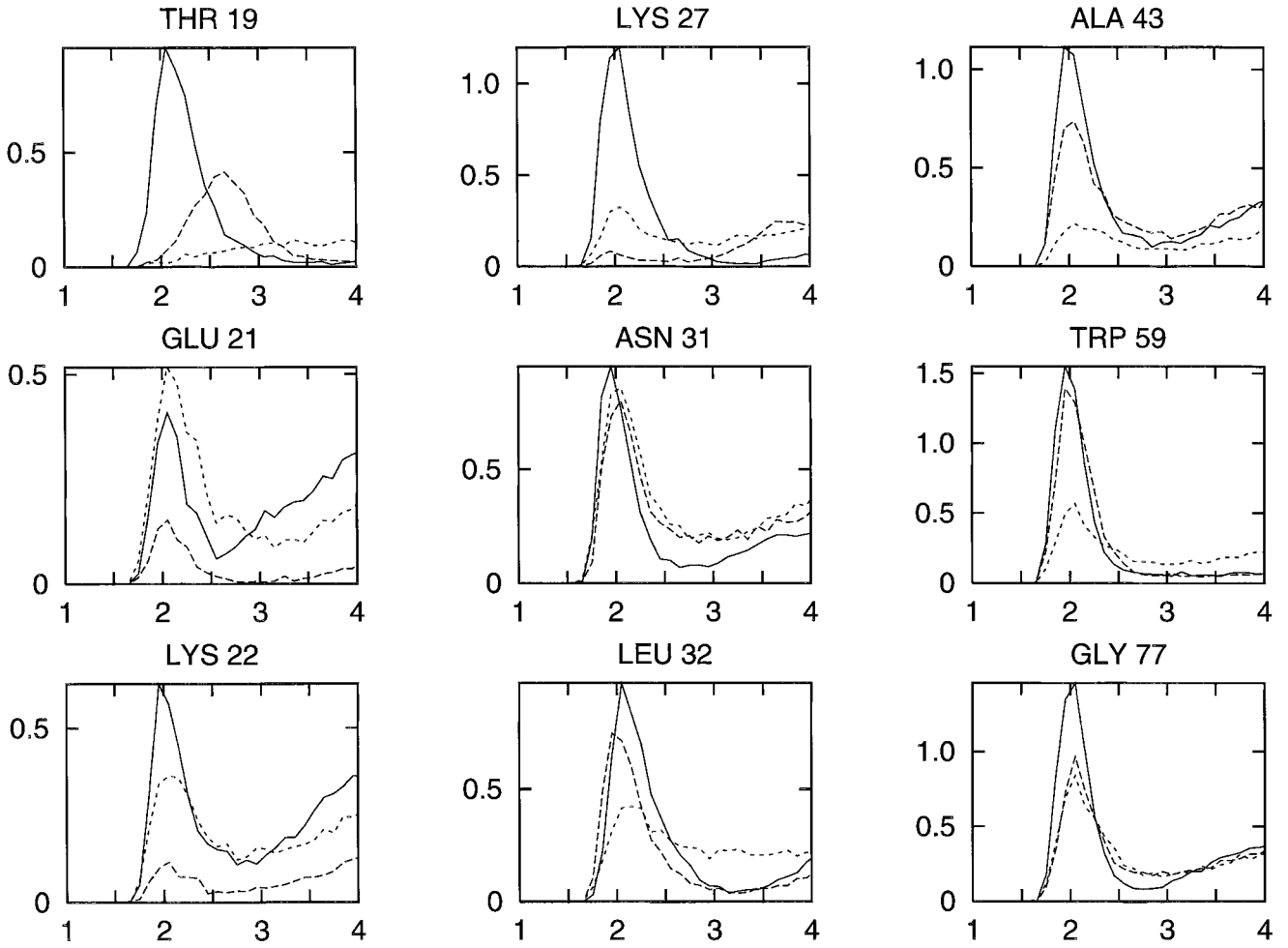

Fig. 10. Radial distribution functions as a function of radial distance r (Å), from the amino group hydrogen, and calculated over the last 1.5 ns of each MD trajectory at T = 300 K (solid line), 360 K (long dashed line), and 430 K (short dashed line) for selected amino groups. Water molecules in the first hydration shell of these amino groups exhibit mean residence times longer than 100 ps at T = 300 K. That is, these amino groups show large exposure to structural water molecules and slow exchange with other water molecules in the solvent.

and closing of hydrogen bonds involved in secondary structure formation can be better understood by monitoring the amino group interactions with water through the HN-OW pair correlation function and the number of waters that occupy the first hydration shell of these atoms.

Figure 13 shows the time block averages (over 10 ps) of the α helical, (i, i+ 4) pattern, and 3–10, (i, i + 3) pattern backbone hydrogen bond formation during the 1.5 ns production of the MD simulations at 300 K and 550 K. Darker shades indicate the larger probability of the hydrogen bond formation over non-overlapping 10 ps time periods. For T = 300 K, we see clearly defined N-terminal, 60's, and C-terminal alpha and 3–10 helices. We also see a persistent hydrogen bonding in the region (α_2) containing amino acids 49–54. Notice, however, that large variations do occur in the fraction of the time that the hydrogen bonds in these regions are formed. We will show that water does not make hydrogen bonds with these amino hydrogens, even when the hydrogen bonds break. For T = 550 K, we see that hydrogen bonds with (i, i + 4) and (i, i + 3) patterns form and break rapidly. These hydrogen bonds occur all along the chain, although they are more predominant in the N-terminal, 60's, and C-terminal helical regions, where some helical shape is maintained even at 550 K. Water molecules coordinate significantly ($N(R) \geq 0.03$; i.e., one water molecule makes a hydrogen bond in every 100 stored configurations) to the amino hydrogens in the N-terminal helix, with the exception of F10, V11, N12, C17, and H18. This is consistent with experimental evidence for the presence of residual structure in the vicinity of H18 and C17, even in the fully denatured state.[2,53] T67 and L68 in the 60's helix, M80, I81, and I82 in Ω_3, and L94, I95, and T102 in the C-terminal helix are also protected from water at 550 K. With the exception of M80 and I81, these regions have been observed to exchange slowly (green and blue regions in Bai et al. notation.[16])

Figure 14 shows plots of the water coordination number to individual backbone amide hydrogens, the percentage of time that a hydrogen bond involving each amino hydrogen breaks, and the correlation (scatter plot) between the

TABLE I. Amino Hydrogen Coordination Number at Various Temperatures†

2ary str.	Amino acid	Res. num.	300 K N(R)	360 K N(R)	430 K N(R)	550 K N(R)
	GLY	1	1.27	1.19	1.06	0.77
	GLY	1	1.23	1.19	1.09	0.74
	GLY	1	1.29	1.17	1.05	0.76
α_N	ASP	2	0.04	0.041	0.27	0.39
α_N	VAL	3	0.51	0.48	0.32	0.22
α_N	GLU	4	0.38	0.28	0.21	0.09
α_N	LYS	5	0.0069	0.022	0.022	0.10
α_N	GLY	6	0	0	0.13	0.13
α_N	LYS	7	0	0	0.11	0.10
α_N	LYS	8	0	0	0.0017	0.06
α_N	ILE	9	0.00033	0	0.012	0.03
α_N	PHE	10	0	0	0.0047	0.01
α_N	VAL	11	0	0.00034	0	0.004
α_N	GLN	12	0	0	0	0.007
α_N	LYS	13	0	0.22	0.037	0.04
α_N	CYS	14	0	0.26	0.058	0.04
α_N	ALA	15	0	0.00069	0.0070	0.03
	GLN	16	0.0071	0.011	0.062	0.09
	CYS	17	0.0095	0	0.034	0.02
	HIS	18	0.00065	0	0	0.001
	THR	19	0.76	0.20	0.044	0.11
Ω_1	VAL	20	0.0045	0.54	0.42	0.18
Ω_1	GLU	21	0.29	0.099	0.44	0.22
Ω_1	LYS	22	0.43	0.086	0.35	0.11
Ω_1	GLY	23	0.86	0.087	0.32	0.09
Ω_1	GLY	24	0.70	0.15	0.48	0.14
Ω_1	LYS	25	0.49	0.83	0.51	0.13
Ω_1	HIS	26	0.74	0.052	0.52	0.08
Ω_1	LYS	27	0.85	0.058	0.28	0.06
Ω_1	THR	28	0.048	0.55	0.14	0.04
Ω_1	GLY	29	0.13	0.18	0.17	0.13
Ω_1	ASN	31	0.59	0.63	0.73	0.25
Ω_1	LEU	32	0.73	0.58	0.42	0.25
Ω_1	HIS	33	0.53	0.39	0.29	0.16
Ω_1	GLY	34	0.10	0.22	0.082	0.15
Ω_1	LEU	35	0	0.0031	0.00033	0.08
Ω_2	PHE	36	0.00033	0.012	0.075	0.10
Ω_2	GLY	37	0.47	0.32	0.075	0.21
Ω_2	ARG	38	0.11	0.0038	0.020	0.15
Ω_2	LYS	39	0.87	0.82	0.57	0.27
Ω_2	THR	40	0.0020	0	0.31	0.20
Ω_2	GLY	41	0.41	0.18	0.39	0.24
Ω_2	GLN	42	0.93	0.13	0.36	0.22
Ω_2	ALA	43	0.79	0.60	0.21	0.16
Ω_2	GLY	45	0.43	0.71	0.49	0.32
Ω_2	PHE	46	0.28	0.52	0.12	0.18
Ω_2	THR	47	0.83	0.64	0.43	0.07
Ω_2	TYR	48	0.16	0.35	0.42	0.07
Ω_2/α_2	THR	49	0.00065	0.0051	0.14	0.09
Ω_2/α_2	ASP	50	0.86	0.78	0.48	0.22
Ω_2/α_2	ALA	51	0.59	0.53	0.25	0.22
Ω_2/α_2	ASN	52	0	0	0.0083	0.10
Ω_2/α_2	LYS	53	0.55	0.031	0.038	0.03
Ω_2/α_2	ASN	54	0.23	0.0021	0.065	0.03
Ω_2	LYS	55	0	0.0010	0.083	0.04
Ω_2	GLY	56	0	0.0027	0.15	0.13
Ω_2	ILE	57	0	0	0.26	0.08
Ω_2	THR	58	0.86	0.74	0.45	0.12
Ω_2	TRP	59	0.91	0.93	0.45	0.08
α_{60}	LYS	60	0.14	0.017	0.011	0.07
α_{60}	GLU	61	0.60	0.82	0.55	0.17
α_{60}	GLU	62	0.37	0.45	0.20	0.09
α_{60}	THR	63	0.00033	0	0.032	0.04
α_{60}	LEU	64	0	0	0	0.03
α_{60}	MET	65	0	0	0.0023	0.03
α_{60}	GLU	66	0.0065	0.0027	0.016	0.04
α_{60}	TYR	67	0.00033	0	0.00033	0.01
α_{60}	LEU	68	0	0	0	0.0076
α_{60}	GLU	69	0	0.0010	0.00067	0.04
α_{60}/Ω_3	ASN	70	0.0091	0.0014	0.015	0.03
Ω_3	LYS	72	0.35	0.59	0.17	0.09
Ω_3	LYS	73	0.034	0.0021	0.0020	0.06
Ω_3	TYR	74	0	0.0010	0.085	0.13
Ω_3	ILE	75	0	0	0.049	0.12
Ω_3	GLY	77	0.93	0.68	0.68	0.27
Ω_3	THR	78	0.0029	0.079	0.0017	0.09
Ω_3	LYS	79	0.92	0.13	0.037	0.12
Ω_3	MET	80	0	0	0.00033	0.02
Ω_3	ILE	81	0.35	0.028	0.25	0.01
Ω_3	PHE	82	0.091	0.090	0.12	0.01
Ω_3	ALA	83	0.79	0.63	0.65	0.04
Ω_3	GLY	84	0.67	0.61	0.071	0.21
Ω_3	ILE	85	0.068	0.35	0.066	0.14
	LYS	86	0.77	0.63	0.23	0.24
α_C	LYS	87	0.83	0.65	0.51	0.20
α_C	LYS	88	0.34	0.20	0.57	0.26
α_C	THR	89	0.0026	0.0041	0.31	0.17
α_C	GLU	90	0.0016	0.0058	0.34	0.21
α_C	ARG	91	0.51	0.43	0.11	0.11
α_C	GLU	92	0.61	0.45	0.27	0.15
α_C	ASP	93	0.43	0.31	0.21	0.08
α_C	LEU	94	0.0059	0.018	0.040	0.02
α_C	ILE	95	0.77	0.49	0.18	0.02
α_C	ALA	96	0.00033	0.0031	0.0070	0.03
α_C	TYR	97	0	0	0.039	0.05
α_C	LEU	98	0	0	0	0.07
α_C	LYS	99	0	0	0	0.11
α_C	LYS	100	0.00065	0.012	0.0067	0.07
α_C	ALA	101	0.0013	0.0017	0.0067	0.03
α_C	THR	102	0.00033	0.00034	0	0.02
α_C	ASN	103	0.049	0.0014	0.0053	0.03
α_C	GLU	104	0.19	0.0014	0.044	0.07

†Secondary structure assignment follows the classifications used by Bai et al.[16] The following amino acids are included in each secondary structure assignment: 2–15 α_N; 20–35 Ω_1; 36–61 Ω_2; 61–70 α_{60}; 70–85 Ω_3 and 87–104 α_C. We also label the region 49–54 as α_2.

hydrogen bond breaking and water coordination. These curves are calculated from the last 1.5 ns of production at each of the four temperatures simulated. The coordination number and hydrogen bond breaking probability curves seem to be highly correlated, as patterns of high hydrogen breaking probability also show a large water coordination number. However, amino acid regions corresponding to the N-terminal, 60's, and C-terminal helices show a small coordination number, but with hydrogen bond breaking probabilities of about 25–30%, at T = 300 K. As T is increased, the percentage of hydrogen bond breaking increases, but the water coordination number remains

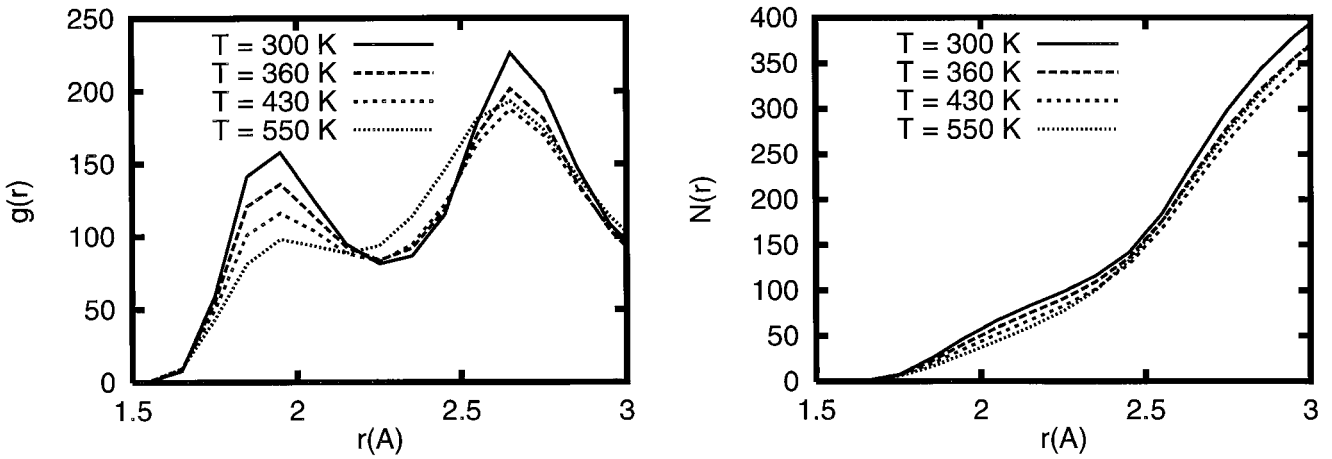

Fig. 11. The right-hand-side plot shows the cyt c–water proximity correlation function calculated during the 1.5 ns production period of MD simulations at T = 300, 360, 430, and 550 K. The proximity correlation function is plotted as a function of the distance between water and the closest protein atom. The peak at 1.75 Å represents the correlation between water molecules and protein hydrogen bond donors. The second peak at 2.75 Å represents the correlation between water molecules and hydrogen bond acceptors in the protein. The right-hand-side plot shows the proximity water coordination number around cyt c as a function of the distance between a layer around cyt c and the distance to the closest atom.

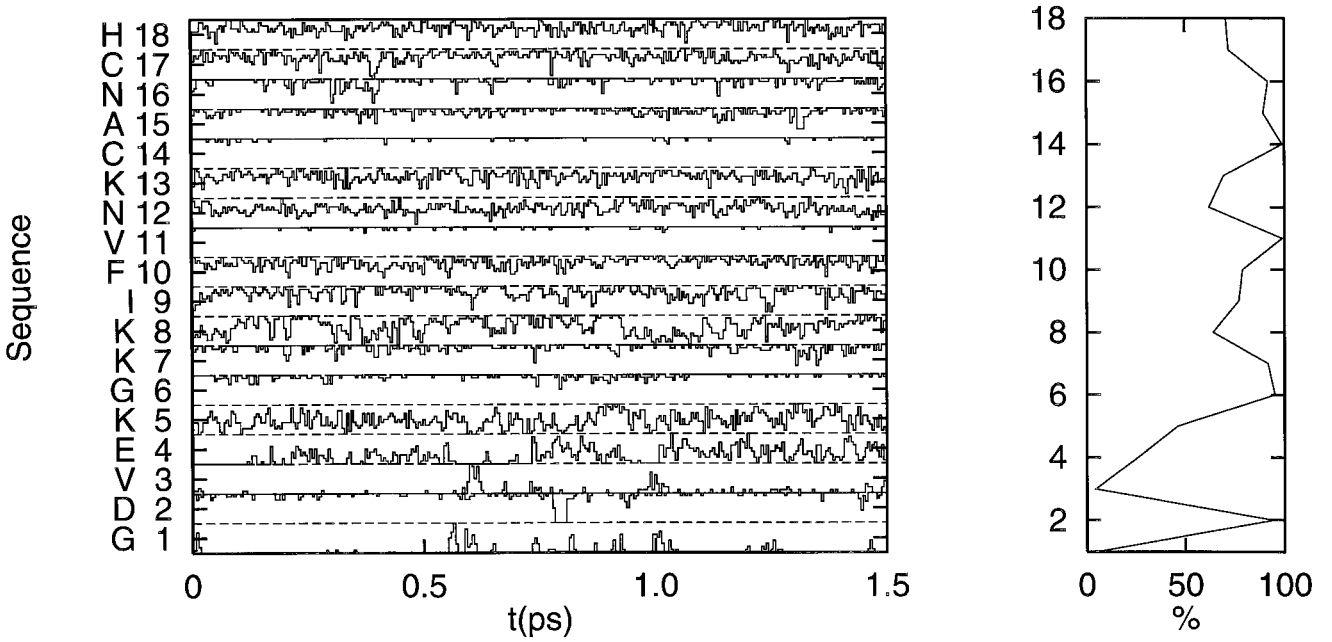

Fig. 12. **Left:** time series of the formation probabilities of intramolecular hydrogen bonds by amino hydrogens on the amino terminal helix of cyt c during the production stage of the MD simulation at 300 K. Hydrogen bond formation probabilities are obtained by averaging the number of times a hydrogen bond is formed over 5 ps blocks (10 configurations). A hydrogen bond is formed when the distance between the oxygen acceptor and hydrogen bond donor is $\leq$ 3.2 Å and the angle between the donor, the hydrogen and the acceptor is $\geq$ 120 degrees. **Right:** Fraction of time that the corresponding amino hydrogen forms intramolecular hydrogen bonds. Amino hydrogen 6 to 14 involved in the N-helix are formed 95, 92, 64, 77, 79, 99, 61, 69, and 99 percent, while the longest time that the hydrogen bonds are consecutively formed are 266, 259, 74, 62, 133, 792, 19, 30, and 792 ps for the amino acids 6 to 14, respectively. In contrast, the coordination number *N(R)*, for the amino hydrogens 6 to 8 and 9 to 15 are zero, and 0.0003 for I9 (i.e., water formed a hydrogen bond twice during the 1.5 ns MD simulation).

close to zero for the helical regions, except at T = 550 K, where helices are disrupted. The scatter plots of the water coordination number as a function of the hydrogen bond breaking percentage show that, despite strong correlations of some backbone amino hydrogens, many hydrogens either have water coordination number close to zero and large hydrogen bond breaking percentage, or hydrogen bond breaking fraction of 1.0, and water coordination number varying from 0 to 1. Notice that the overall water coordination number decreases with temperature, while the hydrogen bond breaking percentage increases. Figures 11 and 13 illustrate the following scenario: as T increases,

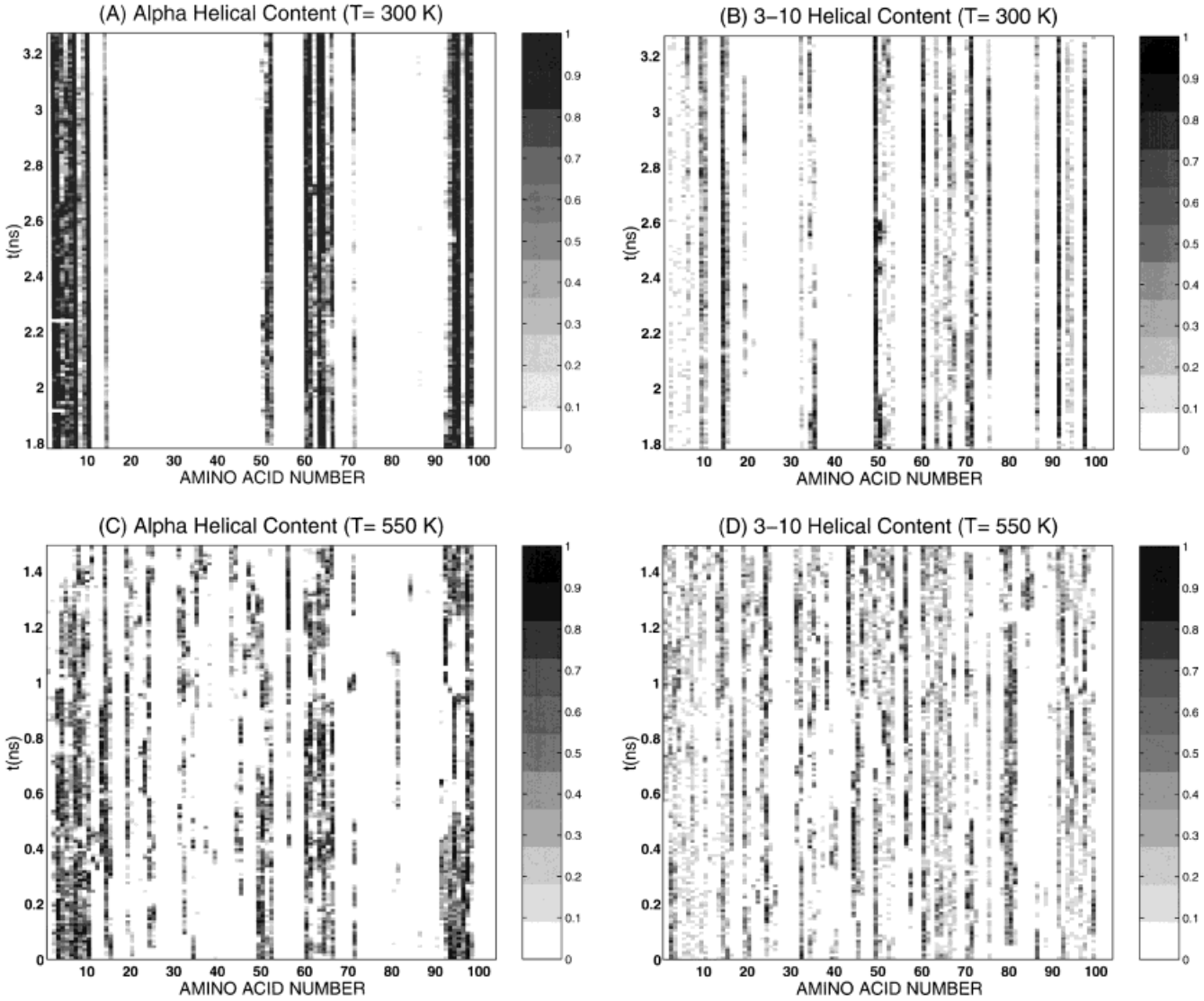

Fig. 13. Time block averages (over 10 ps) of the alpha helical (**A,C**; $i, i+4$) pattern, and 3–10 (**B,D**; $i, i+3$) pattern backbone hydrogen bond formation during the 1.5 ns production of the MD simulations at 300 K (A,B) and 550 K (C,D). The darkness of the shading indicates the proportion of the hydrogen bond formation over non-overlapping 10 ps time periods.

the number of water molecules coordinated to the protein decreases, thus increasing the solvent entropy. Also, the number of protein intramolecular hydrogen bonds increases, but these hydrogen bonds are short-lived and exchange partners rapidly. This broad ensemble of hydrogen bond partners will also have higher entropy than an ensemble of persistent, long-lived hydrogen bonding networks associated with a rigid secondary structure.

CONCLUSIONS

We have investigated the dynamics of oxidized horse heart cyt c by molecular dynamics simulations in aqueous solution at four temperatures—300 K, 360 K, 430 K, and 550 K. Each simulation covers a production time of 1.5 nanoseconds. The resulting trajectories are mapped to a reduced set of collective coordinates that best describes the system fluctuations. This analysis produces a set of collective coordinates that describe most of the fluctuations in the protein. As we see here and with other proteins, these modes are nonlinear in the sense that they describe a multi-basin dynamics. Quasi-harmonic dynamics will misrepresent the character and thermodynamics of these motions, due to the apparent non-linearities. This multi-basin character is present at all temperatures simulated.

At the lower temperature (T = 300, 360 K), we observe that the trajectories sample multiple minima with each basin being sampled for a few hundred picoseconds. Normal modes analysis of proteins of this size show the longest oscillations to be in the 10 ps time scale,[54] much faster than the nonlinear dynamics found here. Structural changes associated with inter-basin displacements are small changes in the Ω loops and coiled regions and

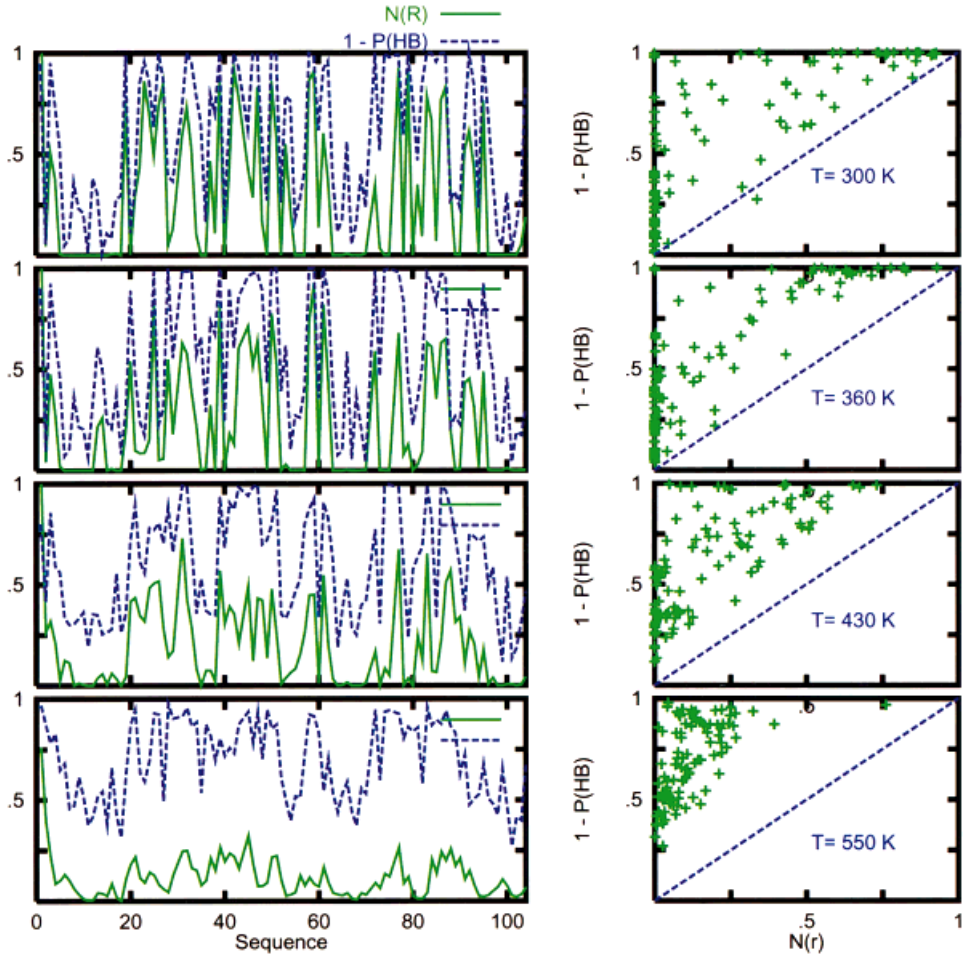

Fig. 14. Corelation between hydrogen bond breaking and amino group water coordination. The left-hand-side plots show water coordination number (solid line) and probability of breaking of hydrogen bonds (dashed line; 1-*P*(HB), where *P*(HB) is the fraction of time that a hydrogen bond is formed), for each backbone amino hydrogen, obtained during the 1.5 ns production period of the MD simulations at T = 300, 360, 430, and 550 K, top to bottom, respectively. The right-hand-side plots show scatter plots of the probability of breaking hydrogen bonds as a function of the water coordination number for all backbone amino hydrogens. The dashed line would indicate perfect correlation between hydrogen bond breaking and water coordination.

relative motions of the α-helices as rigid bodies. The most interesting phenomenon is observed at T = 430 K. At this temperature, a basin near the folded state is sampled for over 400 ps around what first appears to be a stable minimum. During the next 500 ps, a transition out of the folded state basin occurs, resulting in a partially unfolded protein. This conformational change occurs via an intermediate state where the 36 to 61 loop opens and later folds into a different conformation. This structural change, being limited to a 25 amino acid loop, shows a 9 Å *rmsd* movement of the C_α atoms in this region. This change involves an overall *rmsd* fluctuation of 0.6 Å. After this structural change, the protein remains near this basin for the last 300 ps of the simulation. Longer simulations may show other changes of equal or larger magnitude. All α helices in the protein remain intact during the 1.5 ns simulation. We should expect the protein to unfold at this temperature, but on a longer time scale.

The dynamics of cyt c at the four temperatures described here share the following features. The trajectories exhibit multi-basin dynamics at 100 ps time scales, with small overdamped oscillations at shorter time scales. Most fluctuations are described by the interbasin (multimodal) dynamics, not by the intrabasin (unimodal) dynamics. Short time-scale motions are characteristic of a non-Brownian random walk that suppresses diffusion in the protein configurational space (i.e., the system is trapped) for time scales of 100 ps. At longer time scales, the concerted motions of atoms enhances sampling of configurational space, resulting in super-diffusion.

At higher temperatures, the characteristics of the dynamics are similar to those seen at lower temperatures, but the amplitude of the interbasin motions are larger and these motions occur at shorter time scales. This suggests a hierarchical structure of the energy landscape. We put forward the hypothesis that thermal fluctuations under folding and unfolding conditions occur via many interbasin transitions, similar in character to those shown in Figure 7, although covering larger regions of the protein conformational space. In all simulations, the protein only samples a small region of the allowed configurational space due to time scale limitations.

This trapping at faster time scales, combined with large displacements in the longer time scales has been previously observed in 5.0 ns MD simulations of crambin in the crystal environment.[41] The character of these motions can be statistically described by a distribution of waiting times (configurational trapping), coupled to a Levy walk describ-

ing large (in *MSD* distance) interbasin jumps. Suppressed diffusion (sub-diffusion) and enhanced diffusion (super-diffusion) result from the hierarchical nature of the trajectories of proteins in aqueous solution, and are typical of nonlinear dynamical systems in general.[55] This kind of dynamics has been labeled as *strange kinetics* by Shlesinger et al.[56] The large interbasin jumps observed in our calculations dominate the total *MSD* of the system at all simulated temperatures. These large interbasin jumps involve collective motions of protein secondary structural domains, which are the same domains identified by Bai et al. as partially unfolding units in their hydrogen exchange experiments.

The characteristics of the dynamics described by our simulations is in agreement with time-resolved spectroscopic studies of the rebinding kinetics of carbonmonoxyde to myoglobin.[21,57] These experiments observe a power law dependence of the rebinding kinetics, which can be interpreted in terms of a hierarchical, rugged energy landscape.[57]

The dynamics of a system with such an energy landscape is not harmonic. The role of multi-basin, nonlinear motions seem to contradict the suggestion that a purely harmonic model describes the hydrogen exchange.[58] We have shown that *MSD* and B-factors are only directly related for harmonic systems.[59] Even more, B-factors contain only limited information about the multi-basin dynamics. The fluctuations of cyt c studied by three MD simulations show qualitative agreement with the hydrogen exchange experiments of Bai et al.[13–17] That is, fluctuations in the 70–85 and the 36–61 loops are seen at all temperatures, and are therefore identified with low energy activation barriers, while the α helices in the N and C termini, and the 60's helix, involved in slow exchange, exhibit small fluctuations. If we accept this criterion, our results agree with the experimental data and with the results of Bahar et al.,[58] which are based on a much simpler model. However, we have shown here that large *MSD* fluctuations are not necessarily correlated to hydrogen exchange. For example, regions of the proteins forming α helices and turns show significant *MSD*, but as rigid bodies, i.e., the hydrogen bonds involved in the formation of these structures do not break in proportion to these fluctuations. Geometrical definitions of hydrogen bond opening/closing are not a good measure of hydrogen exchange probability. The opening and closing of backbone hydrogen bonds involved in secondary structure stabilization can be better understood by monitoring the amino group interactions with water through the HN-OW pair correlation function and the number of waters that occupy the first hydration shell of these atoms. Although no theory of hydrogen exchange in proteins is provided, the particle nature (granularity) and the kinetics of water molecules are found to be relevant. In addition, evidence is provided showing that correlated hierarchical motions observed in the dynamics of proteins are related to motions of secondary structural domains.

ACKNOWLEDGMENTS

We thank Prof. Tobin Sosnick for enlightening discussions, and Prof. Neville R. Kallenbach for suggesting this work and for fruitful discussions.

REFERENCES

1. Sosnick TR, Mayne L, Hiller R, Englander SW. The barriers in protein folding Nat Struct Biol 1994;1:149–156.
2. Elove GA, Bhuyan AK, Roder H. Kinetic mechanisms of cytochrome c folding: involvement of the heme and its ligands. Biochemistry 1994;33:6925–6935.
3. Yeh S-R, Takahashi S, Fan B, Rousseau DL. Ligand exchange during cytochrome c folding Nat Struct Biol 1997;4:51–56.
4. Pascher T, Chesick JP, Winkler JR, Gray HB. Protein folding triggered by electron transfer. Science 1996;271:1558–1560.
5. Mines GA, Pascher T, Lee SC, Winkler JR, Gray HB. Cytochrome c folding triggered by electron transfer. Chem Biol 1996;3:491–497.
6. Shastry MCR, Roder H. Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nat Struct Biol 1998;5:385–392.
7. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
8. Fersht AR. Nucleation mechanisms in protein folding Curr Opin Struct Biol 1997;7:3–9.
9. Northrup SH, Pear MP, McCammon JA, Karplus M. Molecular dynamics of ferrocytochrome c. Nature 1980;286:304–305.
10. Simonson T, Perahia D. Microscopic dielectric properties of cytochrome c from molecular dynamics simulations in acqueous solution. J Am Chem Soc 1995;117:7987–8000.
11. Muegge I, Qi PX, Wand AJ, Chu ZT, Warshel A. The reorganization energy of cytochrome c revisited. J Phys Chem B 1997;101:825–836.
12. Onuchic JN, Beratan DN, Winkler JR, gray HB. Pathway analysis of protein electron-transfer reactions. Annu Rev Biophys Biomol Struct 1992;21:349–377.
13. Bai YW, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. Proteins 1993;17:75–86.
14. Bai YW, Milne JS, Mayne L, Englander SW. Protein stability parameters measured by hydrogen exchange. Proteins 1994;20:4–14.
15. Bai YW, Englander SW. Hydrogen bond strength and beta-sheet propensities—"the role of a side chanin blocking effect. Proteins 1994;18:262–266.
16. Bai YW, Sosnick TR, Mayne L, Englander SW. Protein folding intermediates: native-state hydrogen exchange. Science 1995;269:192–197.
17. Bai YW, Englander SW. Future directions in folding: the multi-state nature of protein structure. Proteins 1996;24:145–151.
18. Kallenbach NR. Breathing life into the folding pathway of cytochrome C. Nat Struct Biol 1995;2:813–816.
19. Xu Y, Mayne L, Englander SW. Evidence for an unfolding and refolding pathway in cytochrome c. Nat Struct Biol 1998;5:774–778.
20. Englander SW, Kallenbach NR. Hydrogen exchange and structural dynamics of proteins and nucleic acids. Q Rev Biophys 1984;16:521–655.
21. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–1603.
22. García AE. Large-amplitude nonlinear motions in proteins. Phys Rev Lett 1992;68:2696–2699.
23. García AE, Soumpasis DM, Jovin TM. Dynamics and relative stabilities of parallel- and antiparallel-stranded DNA duplexes. Biophys J 1994;66:1742–1755.
24. García AE, Harman JG. Simulations of CRP-$(cAMP)_2$ in non-crystalline environments show a subunit transition from the open to the closed conformation. Protein Sci 1996;5:62–71.
25. Gilmanshin R, Williams S, Callender RH, Woodruff WH, Dyer RB. Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin Proc Natl Acad Sci USA 1997;94:3709–3713.

26. Hagen SJ, Hofrichter J, Eaton WA. Rate of interchain diffusion of unfolded cytochrome c. J Phys Chem 1997;101:2352–2365.
27. Ballew RM, Sabelko J, Gruebele M. Direct observation of fast protein folding: the initial collapse of apomyoglobin Proc Natl Acad Sci USA 1996;93:5759–5764.
28. Privalov PL. Cold denaturation of proteins. Crit Rev Biochem Mol Biol 1990;25:281–305.
29. Privalov PL, Makhatadze GI. Contribution of hydration to protein folding thermodynamics. 2. The entropy and gibbs energy of hydration. J Mol Biol 1993;232:660–679.
30. Garde S, Hummer G, García AE, Paulaitis ME, Pratt LR. Origin of entropy convergence in hydrophobic hydration and protein folding. Phys Rev Lett 1996;77:4966–4968.
31. Qi PX, DiStefano DL, Wand AJ. Solution structure of horse heart ferrocyctochrome c determined by high-resolution NMR and restrained simulated annealing. Biochemistry 1994;33:6408–6417.
32. Qi PXR, beckman RA, Wand AJ. Solution structure of horse heart ferricytochrome-c and detection of redox-related structural changes by high-resolution H-1-NMR. Biochemistry 1996;35:12275–12286.
33. Bushnell GW, Louie GV, Brayer GD. High-resolution 3-dimensional structure of horse heart cytochrome c. J Mol Biol 1990;214: 585–595.
34. Trewhella J, Carlson VAP, Curtis EH, Heidorn DB. Difference in the solution sturctures of oxydized and reduced cytochrome c measured by small-angle x-ray scattering. Biochemistry 1988;27: 1121–1125.
35. Banci L, Bertini I, Gray HB, et al. Solution structure of oxydized horse heart cytochrome c. Biochemistry 1997;1997;36:9867–9877.
36. Jones CM, Henry ER, Hu Y, et al. Fast events in protein folding initiated by nanosecond laser photolysis. Proc Natl Acad Sci USA 1993;90:11860–11864.
37. Hartshorn RT, Moore GR. A denaturation-induced proton-uptake study of horse ferricytochrome-c. Biochem J 1989;258:595–598.
38. Cornell WD, Cieplak P, Bayley CI, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
39. Darden T, York D, Pedersen L. Particle mesh Ewald: an *N*log(*N*) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
40. Pearlman DA, Case DA, Caldwell JW, et al. *AMBER, V. 4.1*, 1995.
41. García AE, Hummer G, blumfield R, Krumhansl JA. Multi-basin dynamics of a protein in a crystal environment. Physica D 1997;107:225–239.
42. García AE. Multi-basin dynamics of a protein in acqueous solution. In Peyrard M, editor. Nonlinear excitations in biomolecules, les editions des physique. Berlin: Springer; 1995. p 191–208.
43. Amadei A, Linssen ABM, Berendsen HC. Essential dynamics of proteins. Proteins 1993;17:412–425.
44. Hayward S, Go N. Collective variable description of native protein dynamics. Annu Rev Phys Chem 1995;46:223.
45. Wang L, Chen RX, Kallenbach NR. Proteolysis as a probe of thermal unfolding of cytochrome c. Proteins 1998;30:435–441.
46. Fontana A, Zambonin M, De Filippis V, Bosco M, Polverino de Lareto P. Limited protoeolysis of cytochrome c in trifluoroethanol. FEBS Lett 1995;362:266–270.
47. Lazaridis T, Lee I, Karplus M. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci 1997;6:2589–2605.
48. Daggett V, Levitt M. Realistic simulations of native-protein dynamics in solution and beyond. Annu Rev Biophys Biomol Struct 1993;22:353–380.
49. Brooks CL. Simulations of protein-folding and unfolding. Curr Opin Struct Biol 1998;8:222–226.
50. Milne JS, Roder H, Wand AJ, Englander SW. Determinants of protein hydrogen exchange studied in equine cytochrome c. Protein Sci 1998;7:739–745.
51. García AE, Stiller L. Computation of the mean residence time of water in the hydration shells of biomolecules. J Comput Chem 1993;14:1396–1406.
52. Mehrotra PK, Beveridge DL. Structural analysis of molecular solutions based on quasi-component distribution-functions: application to [H2C0]aq at 25-degrees-c. J Amer Chem Soc 1980;102: 4287–4294.
53. Howry WA, Sauder JM, Roder H, Scheraga HA. Definition of amide protection factors for early kinetic intermediates in protein folding. Proc Natl Acad Sci USA 1998;95:4299–4302.
54. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsinoinhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 1985;181:423–447.
55. Klafter J, Shlesinger MF, Zumofen G. Beyond Brownian motion. Physics Today 1996;49:33–39.
56. Shlesinger MF, Zaslavsky GM, Klafter J. Strange kinetics. Nature 1993;363:31–37.
57. Ansari A, Berendzen J, Bowne SF, et al. Protein states and proteinquakes. Proc Natl Acad Sci USA 1985;82:5000–5004.
58. Bahar I, Wallquist A, Covel DG, Jernigan RL. Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. Biochemistry 1998;37:1067–1075.
59. García AE, Krumhansl JA, Frauenfelder H. Variations on a theme by Debye and Waller: from simple crystals to proteins. Proteins 1997;29:153–160.