Principal Components of the Protein Dynamical Transition

Alexander L. Tournier and Jeremy C. Smith*

Computational Molecular Biophysics, Interdisciplinary Center for Scientific Computing (IWR), Im Neuenheimer Feld 368, Universität Heidelberg, 69120 Heidelberg, Germany

(Received 10 June 2003; published 14 November 2003)

Proteins exhibit a solvent-driven dynamical transition at 180–220 K, manifested by nonlinearity in the temperature dependence of the average mean-square displacement. Here, molecular dynamics simulations of hydrated myoglobin show that the onset of the transition at ~180 K is characterized by the appearance of a single double-well principal component mode involving a global motion of two groups of helices. As the temperature is raised a few more quasiharmonic and multiminimum components successively appear. The results indicate an underlying simplicity in the protein dynamical transition.

DOI: 10.1103/PhysRevLett.91.208106 PACS numbers: 87.15.–v, 87.68.+z

Internal dynamics are essential to protein function. Experimental and computational work have shown that proteins exhibit a dynamical transition at ~180–220 K, manifested by an increase in the average atomic mean-square displacement for internal motion, $\langle u^2 \rangle$, above a linear, low-temperature regime [1–3]. This transition has been correlated with the onset of protein function [4–6] and remains the focus of much attention [7–17]. Recent work has emphasized that the dynamical transition is solvent controlled [7,8,11,16,18,19].

A consensus has yet to be reached concerning the forms and time scales of the protein motions activated at the dynamical transition in these structurally complex systems. The protein transition resembles that observed in glass-forming liquids, involving a critical slowing down of diffusive motion approaching from the high-temperature side [20,21]. A related interpretation uses the "conformational substates" model, in which proteins are frozen in multiple substates at low temperatures, leading to structural inhomogeneity, whereas above the transition substate interconversion is possible [17,22]. Neutron scattering data on the dynamical transition has been interpreted using an asymmetric double-well model [3], mode-coupling theory [23], and in terms of a "resilience" change with temperature-dependent effective force constants [24].

Here we use molecular dynamics (MD) simulation to determine the subnanosecond dynamical modes of hydrated myoglobin as a function of temperature. To do this, principal component analysis (PCA) is employed to characterize the MD trajectories. It is found that the dynamics activated at the transition can be described with a very small number of components, pointing to an underlying simplicity in the processes involved. The modes activated are found to be collective, and distributed over the protein molecule.

The MD simulations were performed using the CHARMM program [25], version 27b2. Details of the simulation setup have been reported previously [19]. The model system consists of one myoglobin molecule surrounded by a shell of water molecules, and has the same composition as the sample used in neutron scattering experiments performed on myoglobin in which the dynamical transition on the subnanosecond time scale was first demonstrated [3]. The system was simulated at a range of temperatures between 80 and 300 K and thermostatted using the Nosé-Hoover-Chains algorithm [26]. The production phase of each simulation was 1 ns long.

PCA is a convenient method for representing the conformational space explored in an MD trajectory [27,28]. PCA determines the essential motions present in the simulation: the principal component modes. The set of principal components is the solution to the eigenvalue problem of the second-moment matrix, $\mathbf{A}$, of the mass-weighted internal atomic displacements. The elements of the second-moment matrix, A_{ij}, are given by $A_{ij} = \sqrt{m_i m_j} \times \langle (\mathbf{r}_i(t) - \mathbf{r}_i^m)(\mathbf{r}_j(t) - \mathbf{r}_j^m) \rangle$ where $\mathbf{r}_i$ is the position, $\mathbf{r}_i^m$ the mean position, and m_i the mass of atom i, and the average is taken over the time frames of the trajectory. The diagonalization of $\mathbf{A}$ yields the eigenvectors, $\mathbf{w}_k$, i.e., the principal components, and their associated eigenvalues, ζ_k. ζ_k is the mass-weighted mean-square fluctuation along $\mathbf{w}_k$, and can be used to calculate the atom-averaged mass-weighted atomic fluctuation in the protein. For comparison with the principal component results, normal modes were calculated from the energy-minimized conformation of the model system using the harmonic approximation to the potential function. In a normal mode analysis, the Hessian matrix is diagonalized, yielding the normal modes, $\mathbf{w}_k^{\mathrm{NM}}$, with their corresponding angular frequencies, ω_k^{NM}. In what follows, the principal component mode numbering ascends with the effective frequency, ω_k (where $\omega_k^2 = k_B T/\zeta_k$), i.e., with the lowest-frequency mode labeled number 1.

The anharmonicity factor, ρ_k, is a measure of the deviation of mode k from harmonicity [28]. To determine this, a coefficient c_{kl}, projecting the normal mode vector space onto that of the principal components, is calculated:

$c_{kl} = (\mathbf{w}_k \cdot \mathbf{w}_l^{\mathrm{NM}})^2$. The harmonic variance along the kth principal component mode ζ_k^H, is then given by $\zeta_k^H = \sum_l c_{kl}\zeta_l^{\mathrm{NM}}$, and the anharmonicity factor is defined as $\rho_k = \sqrt{\zeta_k/\zeta_k^H}$. ρ is a number greater than 1. Modes with ρ close to 1 are considered harmonic.

Figure 1 shows ρ calculated for the first 30 principal component modes. There is no abrupt, qualitative change in the anharmonicity of the principal components at the onset of the transition (~180–210 K). Rather, the temperature dependence of the anharmonicity is highly mode dependent. As a guide to the eye, for each mode the temperature at which $\rho = 2$ is shown with a bar. Using this criterion, the lowest two modes are anharmonic at temperatures as low as 80 K but no other principal components reach $\rho = 2$ below 190 K. At higher temperatures a few other low modes successively become anharmonic. However, at 300 K still only 20 of the 6536 components have reached $\rho = 2$.

The trajectory, $q_k(t)$, of a principal mode $\mathbf{w}_k$ is given by the projection of the MD trajectory onto the kth principal component: $q_k(t) = (\mathbf{r}(t) - \mathbf{r}^m) \cdot \mathbf{w}_k$, where $\mathbf{r}(t)$ is the protein position vector in conformational space at time t and $\mathbf{r}^m$ the mean position vector. Useful complementary information on the anharmonicity is given by $P_k(q_k)$, the probability distribution along mode k [28]. For harmonic motion $P_k(q_k)$ is Gaussian with variance ζ_k. The standard error, σ_k, of a Gaussian fit to $P_k(q_k)$ is therefore zero for harmonic modes. σ is low for modes with temperature-dependent effective force constants but with close-to-quadratic effective potentials ("quasiharmonic" modes). σ is higher for multiminimum free-energy profiles.

Figure 2 shows σ for the lowest 15 modes. The temperature at which $\sigma = 1$ is indicated with a black bar. Here again very few principal components become non-Gaussian in the temperature range examined. Significant deviation is seen in modes 1 and 2 at low temperatures; these two modes reach $\sigma = 1$ at 140 and 170 K, respectively. Only five further modes reach $\sigma = 1$ by $T =$ 300 K, all other modes being close to Gaussian at all temperatures examined.

The above analysis allows the classification of the principal components into three types:

(i) *Harmonic* modes are those which exhibit approximately Gaussian probability distributions (defined here as $\sigma < 1$) and for which the anharmonicity is low ($\rho \leq 2$).

(ii) *Quasiharmonic* modes exhibit approximately Gaussian probability distributions ($\sigma < 1$) but possess considerable anharmonicity ($\rho > 2$).

(iii) *Multiminimum* modes cannot be well fitted by a Gaussian ($\sigma > 1$) and are also anharmonic ($\rho > 2$).

The contributions made to the mean-square displacement, $\langle u^2 \rangle$, by these three types of principal components are shown in Fig. 3. The increase in $\langle u^2 \rangle$ signaling the incipient phase of the transition (~180–210 K) is seen to arise from a very small number of principal components. By 210 K, only four PCA modes are not harmonic: three of these are multiminimum and one quasiharmonic. At 210 K, 75% of the increase over the linear $\langle u^2 \rangle$ (the "excess" $\langle u^2 \rangle$) is due to multiminimum dynamics. As the temperature increases more multiminimum and quasiharmonic components appear. However, at 300 K still only 20 modes deviate from harmonic behavior, i.e., only 0.3% of the total number of modes in the protein. 70% of the excess $\langle u^2 \rangle$ at 300 K arises from seven multiminimum modes, with the remaining 30% originating from 13 quasiharmonic principal components.

In contrast to harmonic normal modes, the forms of the individual principal components vary with temperature.

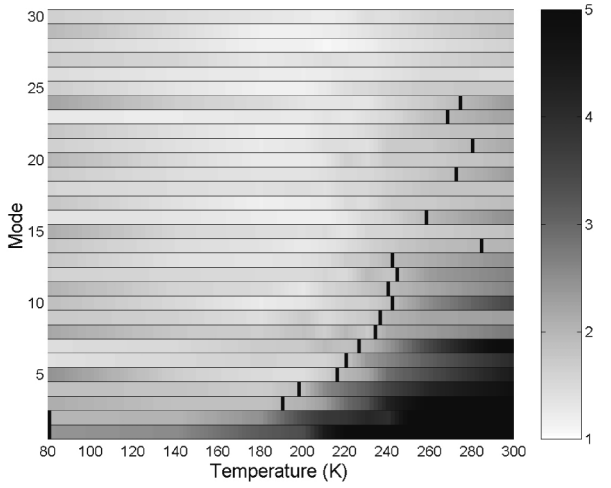

FIG. 1. Anharmonicity factor, ρ, for the first 20 principal component modes. For each principal component the temperature at which $\rho = 2$ is reached is indicated with a black bar.

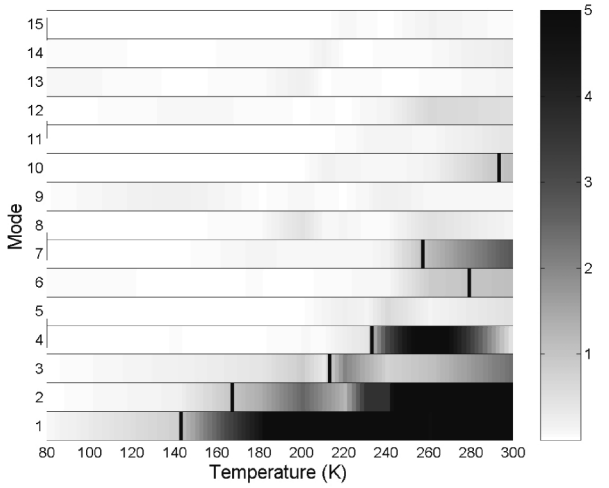

FIG. 2. Standard deviation, σ, of the Gaussian fit to the probability distribution of the first 20 principal component modes. For each principal component the temperature is indicated at which $\sigma = 1$ is reached. The error was defined as $\sigma^2 = 1000\,\overline{[P(q) - G(q)]^2}$, taken over 3 standard deviations of the probability distribution, where $G(q)$ is the Gaussian fit.

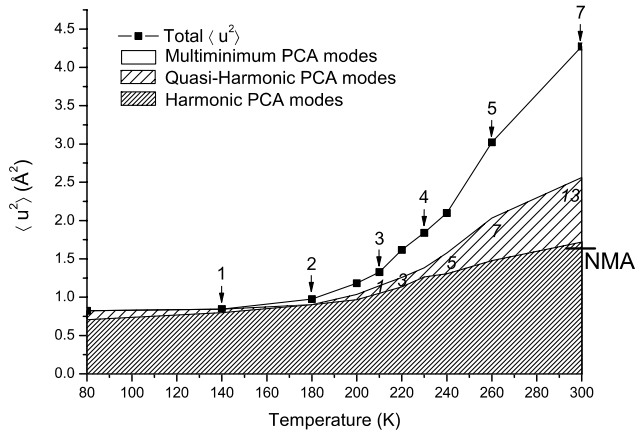

FIG. 3. Decomposition of the mass-weighted protein, $\langle u^2 \rangle$, into contributions arising from the harmonic, quasiharmonic, or multiminimum classes of principal component. Vertical arrows indicate the number of multiminimum modes, and italics the number of quasiharmonic modes at certain temperatures. The $\langle u^2 \rangle$ calculated at 300 K from the normal mode analysis using the same model system and potential function is indicated, the value of this concurs with that calculated from the "harmonic" principal component modes.

Therefore, to further characterize the dynamics associated with the onset of the transition at ~180–210 K it is necessary to examine the principal components concerned at the temperatures involved. The effective free energy along the kth mode is $\mu_k(q_k) = -k_B T \ln P_k(q_k)$. In Fig. 4 are plotted $\mu_k(q_k)$ of modes 1, 2, and 5. For mode 1 the profile is approximately quadratic below 180 K. The onset of the dynamical transition is characterized by the appearance at 180 K in mode 1 of double-well behavior with a free-energy barrier of $\sim k_B T$. This behavior is consistent with the idea that the glass transition originates in strong anharmonicity in low-frequency modes ("bosons") which permits the system to access multiple minima of its configuration space [20]. Above ~240 K the mode 1 profile is highly rugged. Mode 2 is approximately harmonic for $T \sim 200$ K, above which multiple minima again appear. Mode 5 is an example of a mode that is harmonic at low temperatures (here $T < 200$ K) and becomes quasiharmonic above.

Graphical examination of modes 1–5 over the temperature range 180–300 K shows that these principal components all involve collective dynamics distributed over most of the myoglobin molecule. In mode 1 the transition between the free-energy minima at 180 K comprises the relative motion of two large rigid-body blocks of ~900 atoms. One of these blocks contains helices A, H, and F and the other B, C, D, and G. The relative motion involves helices G and H rotating in opposite directions around the axis of the stationary helix E.

Figures 5(a) and 5(b) illustrate the myoglobin structures in the two free-energy minima of mode 1. Rigid-body helix translations and rotations are apparent. Because of rotations around the helix axes the side-chain displacements are generally larger than those of the main-chain atoms, as illustrated by ASP126 on helix H in Fig. 5(a).

Previous work has shown that solvent is the driving force for important internal protein motions [8,29]. The protein dynamical transition is accompanied by a qualitative change in the hydration water shell dynamics [15] involving activation of protein motions by translational water diffusion [7,19]. Hydration water interacts

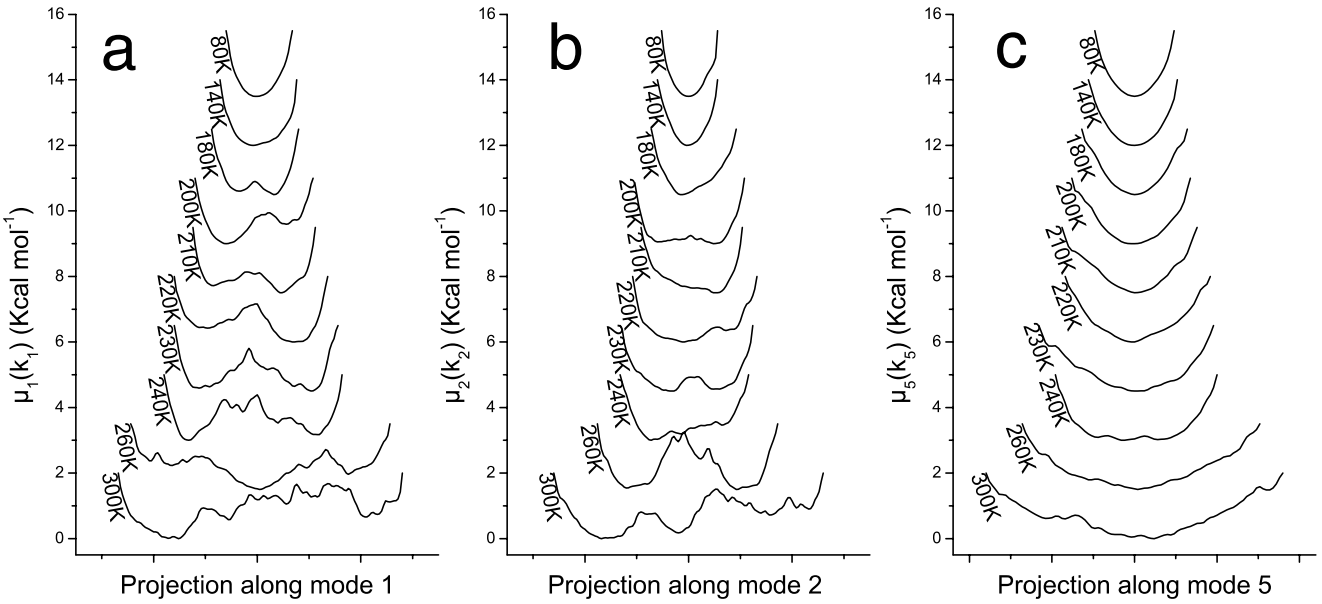

FIG. 4. Free energy profiles of (a) the 1st-, (b) the 2nd-, and (c) the 5th-lowest principal component modes.

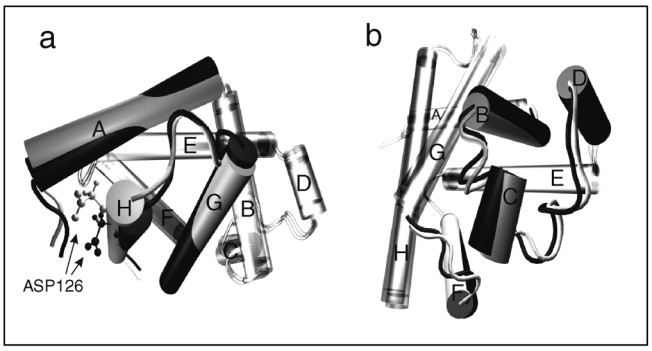

FIG. 5. (a),(b) Motion of myoglobin along the first principal mode calculated at 200 K. The structures of myoglobin are shown in the two minima (black and white) of the mode. The helices are represented as cylinders. For ease of viewing the amplitude of the motion is multiplied by a factor of 6. (a) also shows the sidechain of ASP126.

principally with the surface atoms, which correspondingly have been shown to exhibit the largest change in $\langle u^2 \rangle$ at the dynamical transition [9,19]. The present simulation analysis has provided a description of the subnanosecond protein dynamics activated at the transition. The results indicate that the solvent:surface interaction propagates to the interior of the protein via collective dynamics that can be described by a very small number of principal components. By 210 K only two quasiharmonic modes and two multiminimum modes deviate substantially from harmonic behavior.

The modes initiating the transition are global and distributed over the protein. The quasiharmonic components correspond to softened effective harmonic motion. The largest displacements, however, arise from the activation of barrier-crossing, multiminimum components. In the mode with the largest contribution to the onset of the transition in the present system the barrier crossing involves blocks of supersecondary structural elements moving relative to each other. This rigid-body motion qualitatively resembles structural changes seen in proteins in different functional states [30]. However, it is unclear to what extent motions activated in the dynamical transition describe functional protein dynamics.

As computer power improves it will be of interest to perform a similar analysis on longer time scales, so as to examine the time scale dependence [31] and the hierarchical substate structure [22,28] of the low-frequency principal coordinates. It is hoped that the underlying simplicity revealed here in the subnanosecond motions associated with the dynamical transition will facilitate future studies of the relationship between protein energy landscapes and function.

*Corresponding author.

[1] F. Parak and H. Formanek, Acta Crystallogr. Sect. A **A27**, 573 (1971).
[2] I. E. Iben *et al.*, Phys. Rev. Lett. **62**, 1916 (1989).
[3] W. Doster, S. Cusack, and W. Petry, Nature (London) **337**, 754 (1989).
[4] F. Parak *et al.*, FEBS Lett. **117**, 368 (1980).
[5] M. Ferrand, A. J. Dianoux, W. Petry, and G. Zaccai, Proc. Natl. Acad. Sci. U.S.A. **90**, 9668 (1993).
[6] B. F. Rasmussen, A. M. Stock, D. Ringe, and G. A. Petsko, Nature (London) **357**, 423 (1992).
[7] M. Tarek and D. J. Tobias, Phys. Rev. Lett. **88**, 138101 (2002).
[8] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak, Proc. Natl. Acad. Sci. U.S.A. **99**, 16 047 (2002).
[9] M. M. Teeter, A. Yamano, B. Stec, and U. Mohanty, Proc. Natl. Acad. Sci. U.S.A. **98**, 11 242 (2001).
[10] A. L. Lee and A. J. Wand, Nature (London) **411**, 501 (2001).
[11] V. Reat, *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 9961 (2000).
[12] J. Smith, K. Kuczera, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **87**, 1601 (1990).
[13] J. A. Hayward and J. C. Smith, Biophys. J. **82**, 1216 (2002).
[14] M. Tarek and D. J. Tobias, Biophys. J. **79**, 3244 (2000).
[15] A. R. Bizzarri, A. Paciaroni, and S. Cannistraro, Phys. Rev. E. Stat. Phys. **62**, 3991 (2000).
[16] D. Vitkup, D. Ringe, G. A. Petsko, and M. Karplus, Nat. Struct. Biol. **7**, 34 (2000).
[17] A. Ostermann, R. Waschipky, F. G. Parak, and G. U. Nienhaus, Nature (London) **404**, 205 (2000).
[18] T. Kleinert *et al.*, Biochemistry **37**, 717 (1998).
[19] A. L. Tournier and J. C. Smith, Biophys. J. **85**, 1871 (2003).
[20] C. A. Angell, Science **267**, 1924 (1995).
[21] P. G. Debenedetti and F. H. Stillinger, Nature (London) **410**, 259 (2001).
[22] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[23] W. Doster, S. Cusack, and W. Petry, Phys. Rev. Lett. **65**, 1080 (1990).
[24] G. Zaccai, Science **288**, 1604 (2000).
[25] B. R. Brooks, *et al.*, J. Comput. Biol. **4**, 187 (1983).
[26] M. E. Tuckerman and G. J. Martyna, J. Phys. Chem. **194**, 159 (2000).
[27] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981).
[28] A. Kitao, S. Hayward, and N. Go, Proteins **33**, 496 (1998).
[29] M. Diehl, W. Doster, W. Petry, and H. Schober, Biophys. J. **73**, 2726 (1997).
[30] M. Gerstein *et al.*, J. Mol. Biol. **234**, 357 (1993).
[31] B. Hess, Phys. Rev. E **65**, 031910 (2002).