FEATURE ARTICLE

Aspects of Protein Reaction Dynamics: Deviations from Simple Behavior

Martin Karplus

Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg France, and Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received: October 6, 1999

Proteins are complex systems so that their reactions may have characteristics not often observed in small molecules. Two common aspects of small molecule reactions are an exponential time dependence and Arrhenius-like temperature dependence. This review shows that corresponding behavior is not always observed in reactions involving proteins. Bond-making and bond-breaking reactions, as well as conformational changes, which do not involve chemical bonds, are discussed. As examples of bond making, bond breaking reactions, enzyme catalysis by triosephosphate isomerase and ligand binding by myoglobin are considered. Several reactions involving conformational changes are described. They include aromatic side chain rotations, loop/lid opening and closing motions, quaternary structural changes in hemoglobin, and protein folding. In some of these cases, deviations from simple small molecule behavior have been observed, but it is evident that more information is needed for a complete understanding. The protein folding reaction, which is most complex because it involves the entire polypeptide chain, is analyzed in some detail.

I. Introduction

Macromolecules of biological interest include proteins, nucleic acids, lipids and carbohydrates. They form complex mesoscopic systems with many degrees of freedom and multidimensional potential energy surfaces. Their motions cover a wide range of time scales (from femtoseconds to seconds or even longer) and length scales (from 0.01 to 10 Å or even larger). A fundamental question concerns the relationship between the microscopic interactions, which govern the potential energy surface ("energy landscape"), and the observed macroscopic behavior. Of particular interest is whether and, if so how, the microscopic complexity manifests itself in the simplicity or complexity of macroscopic observables.

This review focuses on proteins with the expectation that a corresponding analysis is applicable to other macromolecular biological systems. Proteins participate in a wide range of reactions, all of which involve internal motions. Some, as in the binding and release of oxygen in myoglobin and hemoglobin and the catalysis of reactions by enzymes, involve the making and breaking of chemical bonds. Many other reactions of proteins involve conformational changes, such as hinge-bending motions, subunit reorientations, and the folding and unfolding reactions of the polypeptide chain. They are governed by relatively weak nonbonded interactions arising from the van der Waals and electrostatic terms in the potential energy. Any or all of these reactions may have complex time and temperature dependences because of the wide range in the time and length scales of the underlying protein motions.

Proteins in their native state have been shown to have a large number of similar configurations (or substates). This conclusion is based on a variety of experiments; the earliest of these involved studies of the ligand rebinding reaction of myoglobin.[1] The nature of and the transitions between these substates have been analyzed by quenched molecular dynamics simulations.[2] Since these substates are likely to have functionally different properties, their role in protein reactions has to be considered. More generally, there is the question of whether the distribution of protein configurations plays a role in the kinetics of biologically important processes at room temperature (or at other temperatures). What is the effect on protein reactions of the rate at which substates interconvert? Are the effects of the configurations of the protein that govern the rate of a given phenomenon localized or do they involve the protein as a whole? Does nonexponential behavior arise from relaxation phenomena in a nonequilibrium situation? Questions of this type, which are of interest for localized events (e.g., ligand binding) become even more important for global reactions such as protein folding, for which much less is known about the nature and complexity of the energy surface.

To focus the present analysis, it is useful to define first what is meant by "simplicity" and "complexity" in a reaction. Given this framework, some examples are used to illustrate what has been found theoretically and experimentally in reactions involving proteins. The examples include the catalysis of the conversion of DHAP to GAP by the enzyme triosephosphate isomerase, ligand rebinding after photolysis of myoglobin, and certain cases of localized conformational changes in native proteins. Finally, the protein folding reaction, which involves the entire polypeptide chain, is discussed. The concluding section summarizes the present situation and the outlook for a more complete understanding of reactions involving proteins.

II. Kinetic Background and Definitions

Simple versus complex behavior for a reaction is defined here in terms of two not unrelated criteria. A reaction is considered to have simple behavior if a phenomenological expression involving rate constants describes the time dependence of the

10.1021/jp993555t CCC: $19.00
Published on Web 12/11/1999

reaction and if the temperature dependence of the rate is Arrhenius-like.

We consider the unimolecular reaction (A $\rightleftarrows$ B), which is appropriate for many protein reactions, including geminate rebinding in myoglobin, an enzymatic reaction after the Michaelis complex has formed, conformational change, and protein folding and unfolding. The differential expressions for the rate of the reaction can be written[3]

$$\frac{dB}{dt} = k_A A(t) - k_B B(t) \tag{1a}$$

$$\frac{dA}{dt} = -k_A A(t) + k_B B(t) \tag{1b}$$

where k_A and k_B are the rate coefficients for the forward and reverse reaction and $A(t)$ and $B(t)$ are the concentrations of species A and B, respectively, at time t; low concentrations (the limit of infinite dilution) are assumed to avoid the need to consider interactions among molecules. If the rate coefficients k_A and k_B are essentially independent of time (i.e., $k_A(t) = k_A$, $k_B(t) = k_B$), they are referred to as the rate constants for the reaction. The integrated rate expressions are then

$$\Delta A(t) = A(t) - A_{eq} = \Delta A(0)\, e^{-kt} \tag{2a}$$

$$\Delta B(t) = B(t) - B_{eq} = \Delta B(0)\, e^{-kt} \tag{2b}$$

where A_{eq} and B_{eq} are the concentrations of A and B at equilibrium and $\Delta A(t)$ and $\Delta B(t)$ and $\Delta A(0)$ and $\Delta B(0)$ are the deviations from equilibrium at time t and $t = 0$, respectively. The constant k in eqs 2 is the sum of the rate constants ($k = k_A + k_B$).

Equations 2 are the standard rate expressions for a unimolecular reaction, and their validity corresponds to the first criterion for simplicity. If k is a function of time, the reaction is not simple by our definition. For higher order reactions, corresponding expression involving a rate constant can be written for the differential rate law, though the integrated expression is more complex, except under certain limiting conditions. An insightful discussion of the treatment of cases where k depends on the time and the reaction rate is nonexponential has been given by Zwanzig.[124] He makes a useful separation between "static disorder" (different molecules in the ensemble have different activation barriers) and "dynamic disorder" (the value of k within a molecule varies with time). The term "strange kinetics" has also been introduced to describe such systems.[125] I return to this point in section IV.B.

The second criterion for a simplicity of a reaction is that the temperature dependence of the rate follows an Arrhenius-like equation. This requires that the rate constant can be written to a good approximation in the form

$$k = A(T)\, e^{-\Delta E^{\ddagger}/RT} \tag{3a}$$

where $\Delta E^{\ddagger}$, the activation energy, is approximately independent of temperature and $A(T)$ is the pre-exponential factor with only a weak temperature dependence. Given eq 3a, we have

$$\frac{d \ln k}{d(1/T)} \cong -\frac{\Delta E^{\ddagger}}{R} \tag{3b}$$

i.e., a plot of $\ln k$ versus $(1/T)$ approximates a straight line. More generally we can write

$$k = A'(T)\, e^{-\Delta G^{\ddagger}/RT} \tag{4a}$$

and, by the Gibbs−Helmholtz relation,

$$\frac{d \ln k}{d(1/T)} \cong -\frac{\Delta H^{\ddagger}}{R} \tag{4b}$$

where $\Delta G^{\ddagger}$ and $\Delta H^{\ddagger}$ are the free energy and enthalpy of activation and $A'(T)$ is the appropriate pre-exponential factor, which is significantly different from $A(T)$ in many cases.

For elementary gas-phase reactions involving small molecules, both conditions for simplicity are usually satisfied over a wide range of temperatures.[4] However, even in these systems, where the entropic contribution to the reaction is small, it is important for quantitative interpretations to consider the temperature dependence of the pre-exponential factor. An example is provided by the $H + H_2$ exchange reaction. The temperature dependence of the rate constant calculated by the quasiclassical trajectory method[5] can be fitted by the Arrhenius expression

$$k(T) = 4.3 \times 10^{13} \exp[-7.435/RT]\ \text{cm}^3\ \text{mol}^{-1}\ \text{s}^{-1} \tag{5a}$$

over the temperature range 300 to 1000 K. A slightly better fit is obtained by taking the temperature dependence of the pre-exponential factor into account. The result is

$$k(T) = 7.9 \times 10^{9} T^{1.18} \exp[-6.23/RT]\ \text{cm}^3\ \text{mol}^{-1}\ \text{s}^{-1} \tag{5b}$$

Equations 5 make clear that, even for this simple reaction with a high activation barrier, some care has to be used in interpreting the numerical values obtained by fitting Arrhenius-like equations.

By contrast for systems such as disordered crystals and glasses,[6−8] nonexponential relaxation after a perturbation and non-Arrhenius temperature dependence for motional phenomena are commonly observed. Although phenomenological models exist that ascribe such behavior to the multiplicity of time scales, the details of the origins of the complexity are generally not fully understood at the atomic level. Observing corresponding complexities in protein reactions would not be surprising, given the multiplicity of time scales for their internal motions.

III. Aspects of Simple Behavior

III.A. Criterion for Existence of a Rate Constant. There are three essential aspects of a reaction that are involved in the existence of a well-defined rate constant (see, for example, ref 3). We consider here a single elementary step in a reaction; if the overall reaction has several steps (as in many enzymes[9]), the discussion would apply separately to each one. The first requirement is that it is possible to define a coordinate, the reaction coordinate (or other progress variable), for the transition from reactants to products; the second requirement is that there exists a well-defined barrier with a free energy several times kT separating the reactant and product states along the reaction coordinate, and the third, intimately related to the other two, is that the relaxation time of all the degrees of freedom involved in the transformation, other than the reaction coordinate, is fast relative to motion along the reaction coordinate. Given that these three requirements are satisfied, a rate constant can be defined and simple exponential behavior is expected. It can be true that a rate constant (simple rate expression) exists at one temperature but not at another because the relaxation times of different modes of the system depend differently on the temperature.

To make clearer what is involved in the three requirements, it is useful to review the activated dynamics method that is widely used for calculating the rates of chemical reactions. The method was originally proposed for gas-phase reactions[10−12] and

has been developed and applied in its present form to reactions in solution[13,14] and in proteins.[15]

Following the formulation of Chandler,[3] we can write the rate coefficient in the form

$$k = \frac{1}{2}\kappa\langle|\dot{\zeta}|\rangle_{\ddagger}\frac{\rho(\zeta^{\ddagger})}{\int_i \rho(\zeta)\delta\zeta} \tag{6}$$

where ζ is the reaction coordinate and the double dagger ($\ddagger$) defines the value of the coordinate at the transition state. The symbol $\rho(\zeta)$ represents the probability density for the system as a function of ζ, which corresponds to the free energy (potential of mean force), $W(\zeta)$, defined by

$$W(\zeta) = -RT \ln \rho(\zeta)/C \tag{7}$$

with appropriate normalization constant, C. The activation free energy in eq 4a can now be written

$$\Delta G^{\ddagger} = -RT \ln\left[\frac{\rho(\zeta^{\ddagger})}{\int_i \rho(\zeta)\, \mathrm{d}\zeta}\right] \tag{8}$$

where the integral $\int_i$ is over the values of the reaction coordinate associated with the reactant valley. Such a potential of mean force is meaningful if and only if for each value of the reaction coordinate ζ, the system can be assumed to be in equilibrium with respect to all of the other degrees of freedom; i.e., the latter must relax rapidly relative to motion along the reaction coordinate. That the barrier be sufficiently high and well defined means that the essential step determining the rate of the reaction is the passage over the barrier, which is a rare event. If the barrier were very low or nonexistent, one might expect to observe a diffusive rate for a reaction in solution that would not generally satisfy the conditions for a simple reaction. The quantity $\langle|\dot{\zeta}|\rangle_{\ddagger}$ is the equilibrium average value of the time derivative of ζ ($\dot{\zeta} = \mathrm{d}\zeta/\mathrm{d}t$) evaluated with ζ restrained to the transition state $\ddagger$. This can be determined computationally by restricting the system to $\zeta = \zeta^{\ddagger}$ and doing simulations that are long enough to obtain the average velocity; again this procedure is meaningful only if there is a separation of time scales. Finally, κ is the transmission coefficient, which in gas phase reactions usually has a value close to unity unless surface crossing or very high collision energies are involved. However, in solution or in proteins, κ often is significantly less than unity. The time-dependent transmission coefficient $\kappa(t)$ can be calculated from the reaction flux expression[3,13]

$$\kappa(t) = N\langle(\dot{\zeta})(0)\theta[\zeta(t) - \zeta^{\ddagger}]\rangle_{\ddagger} \tag{9}$$

where $\theta[\zeta(t) - \zeta^{\ddagger}]$ is a step function that equals 1 when the reaction coordinate is in the product state and is zero otherwise; N is a normalization constant such that $\kappa(0+)$ is equal to unity. Equation 7 corresponds to the transition state theory (TST) expression if $\kappa(t)$ equals unity; that is, it is assumed in TST that there are no recrossings, which would reduce the rate constant, so that the TST value is an upper limit to the reaction rate. The function $\kappa(t)$ is, in fact, equal to unity at very short times, corresponding to the fact that no recrossings have occurred. For a well-defined rate constant, there is then a short relaxation period in which systems may recross the transition state one or more times until they settle into the reactant well (no contribution to reaction rate) or into the product well (a contribution to reaction rate). The requirement for a high activation barrier (relative to kT) corresponds to the fact that, particularly for a unimolecular reaction, collisions remove kinetic energy from the reaction coordinate as soon as the system has left the transition state region so the system is quenched in the reactant or product region. This is true in a sufficiently dense system like a macromolecule or a liquid.[16]

If a separation of time scales were not present, $\kappa(t)$ would not reach a plateau value and the rate coefficient would have a significant time dependence; i.e., the reaction rate would not show the simple exponential behavior corresponding to eqs 2. Concomitant with this is the possibility that the definition of a potential of mean force barrier, as in eqs 6−8, is not appropriate for the system; that is, the coupling of different degrees of freedom may be such that a reaction coordinate, which varies slowly relative to other degrees of freedom, does not exist. In proteins, where there are many different types of degrees of freedom with different time scales of relaxation, deviations from simple behavior is not unexpected.

There are two additional points. The first is that we have used the description of the activated dynamics method primarily as a way of indicating the factors involved in the existence of a rate constant. It is, of course, not necessary for the existence of simple exponential behavior that we can determine the reaction coordinate or appropriate progress variable. It is required only that there exists such a coordinate with the properties we have described. This is clearly relevant to protein folding reactions where, as we discuss below, the definition of a reaction coordinate is a major problem. More generally, a difficulty in applying the activated dynamics method to complicated many-body systems is the determination of an optimal reaction coordinate and the transition state as a function of that coordinate. This problem is rooted in the conceptual and computational complexity associated with finding the minimal number of atomic coordinates which adequately specify a dividing surface between a "reactant" configuration and a "product" configuration. The second point is that the use of eq 6 does not require knowing the optimal transition state, which is essential only for the application of transition state theory. With eq 6, a poor choice of reaction coordinate or transition state (dividing surface) does not invalidate the method, in principle. The result is that the transmission coefficient will have a reduced value. However, the efficiency of the transition state sampling by activated dynamics decreases rapidly as the choice of reaction coordinate and transition state become less than optimal.

To illustrate the possible complexities of determining the reaction coordinate we show two potential energy surfaces in Figure 1. The first (Figure 1a) is a calculated collinear surface for the $H + H_2$ reaction.[5] Its shape is such that the reaction can be followed by the coordinate R_{BC}, for example, with R_{AB} expressed as a function of R_{BC}. Alternatively, the reaction coordinate can be expressed as a uniquely defined linear combination of R_{BC} and R_{AB}.[4] Figure 1b represents the free energy surface calculated with a 125-mer bead model for the protein folding reaction;[102] a free energy surface is shown because the entropy plays an important role in the reaction. The surface is drawn as a function of two coordinates that are adequate for describing the complexity of the reaction.[102] The variable Q_c is the number of "core" contacts and Q_s is the number of "surface" contacts; details are given in the original paper. The figure shows two dominant average paths that make clear that a single reaction coordinate cannot be used. For the "slow" path on the left, Q_s is not a monotonic function of Q_c because an intermediate, I, is formed and the Q_s values decrease as the reaction proceeds from I; that is, some contacts must be broken for the reaction to go forward from the intermediate.

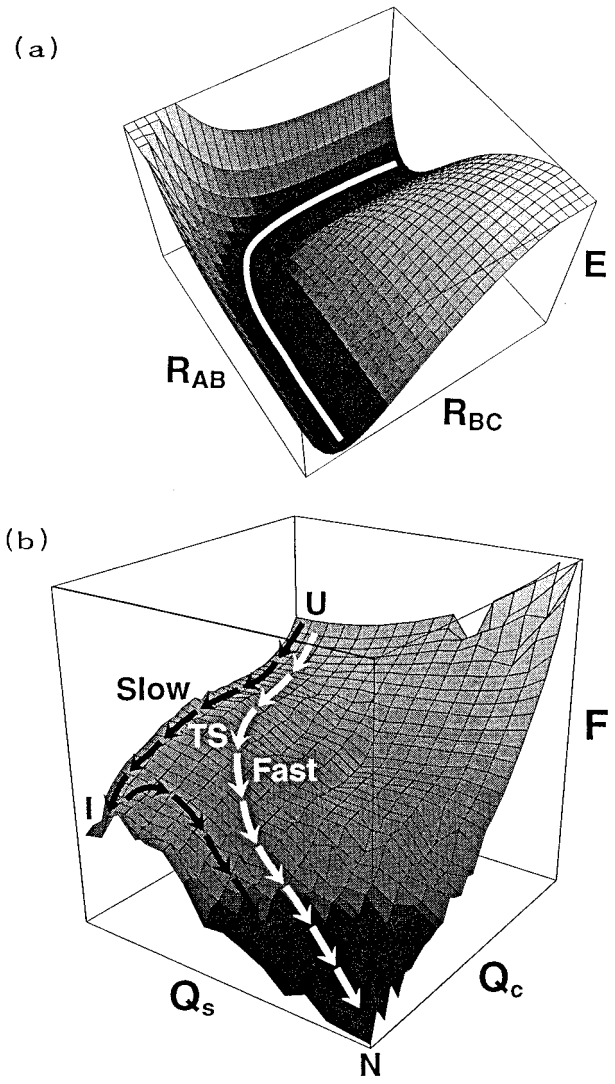

Figure 1. Simple and complex reaction surfaces. In both cases, the reaction proceeds from the upper right to the lower left. (a) Potential energy surface for the reaction $H_{AB} + H_C \rightarrow H_A + H_{BC}$. The surface is calculated for a linear collision as a function of R_{AB}, the H_A-H_B distance and R_{BC}, the H_B-H_C distance. The well-defined reaction coordinate is shown as the heavy white line in the figure. (adapted from refs 5 and 93). (b) Free energy surface for a 125-mer heteropolymer protein model. The two coordinates are a minimal set required to describe the free energy surface and the possibility of multiple pathways on that surface. Q_c represents the fraction of native "core" contacts and Q_s the fraction of native "surface" contacts. Two average paths are shown; one (white arrows) represents "fast track" folding without an intermediate and the other (black arrows) corresponds to the most common "slow track" with an intermediate (adapted from ref 102, which presents the details of the model and its analysis).

By contrast, for the "fast track", it appears that Q_c could be used as the reaction coordinate; that is, Q_s could be expressed a function of Q_c to describe the reaction in this region. However, this is not true because the path shown in the figure is an average over an ensemble of paths, each with different values of Q_s as a function of Q_c.

III.B. Validity of Arrhenius Equation. The second criterion for simplicity is concerned with the variation of the rate constant, if it exists, with temperature. If the reaction has an Arrhenius-like temperature, as defined in eqs 3 and 4, it is regarded as simple from this point of view. Gas-phase reactions of small molecules with well-defined barriers at least several times kT tend to have rate constants that obey the Arrhenius law; a plot of ln k versus $1/T$ approximates a straight line with an increase in ln k as T increases (eqs 4b with $\Delta H^{\ddagger}$ constant).

Deviations from such a simple Arrhenius temperature dependence in complex systems can be of several types. At high temperature, significant curvature in the Arrhenius plot may appear with the possibility of a decrease in the rate with increasing temperature. This has been observed in the protein folding reactions and is discussed below. At low temperatures, there are several types of deviation from Arrhenius behavior. One of these corresponds to a decrease in the temperature dependence (i.e., the rate constant falls off more slowly than predicted by the Arrhenius law) that can arise from the contribution of tunneling to the reaction rate. The other deviation corresponds to a higher order temperature dependence; that is, the reaction rate decreases more rapidly at low temperatures than expected from the Arrhenius law. Such behavior is often referred to as a "super-Arrhenius" temperature dependence. We consider both types of deviations from Arrhenius behavior in the relevant sections.

IV. Examples of Protein Reactions

IV.A. Dynamics of Enzymatic Reactions. The enzyme triosephosphate isomerase (TIM) catalyzes the interconversion of dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP) in a central step of the glycolytic cycle.[9] TIM is of particular interest because it has been characterized as a "perfect enzyme"; that is, the chemical steps in the reaction are accelerated such that the overall rate is limited by the diffusion-controlled binding of the substrate and release of the product.[17] The reaction (DHAP $\rightleftarrows$ GAP) has several steps. The first of these, which we consider in what follows, consists of the transfer of a proton from DHAP to the residue Glu 165 of TIM, which acts as a base to form an enediolate intermediate (see Figure 2). The rate enhancement of this step is about 5×10^5 relative to the reaction catalyzed by a general base in solution.[18] As has been shown,[19,20] the reduction of the activation barrier is due to interactions of the reactants and the transition state with specific residues of the enzyme. Displacements of atoms in the enzyme by 1 Å or so can have a large effect on the activation barrier.[21,22] Such a tight coupling of structure and reactivity raises the question of its possible effect on the dynamics of the reaction; that is, do the structural requirements for transition state stabilization result in a significant reduction in the effective rate of crossing the barrier? Is the reaction simple in terms of the present definition? To begin to answer these questions, the activated dynamics method summarized in section III.A was applied to the first step of the TIM reaction;[23] that is, a reaction coordinate was selected, the transition state was determined by calculating the potential of mean force along the reaction coordinate, and the transmission coefficient was evaluated by use of the reaction flux method. This is the first time the transmission coefficient has been determined for an enzymatic reaction. Warshel and co-workers have made important studies of the dynamics of reactions in solution and enzymes,[24,25] but their conclusions were based on more approximate treatments that did not involve calculations of the transmission coefficient.

Although the activation energy had been studied by a semiempirical molecular orbital QM/MM method,[19] a simpler representation of the potential energy surface was needed to calculate the large number of trajectories that are required for evaluating the transmission coefficient. The representation used describes the surface for the reaction in terms of the coupling of two states, one corresponding to the reactants and the other to the products.

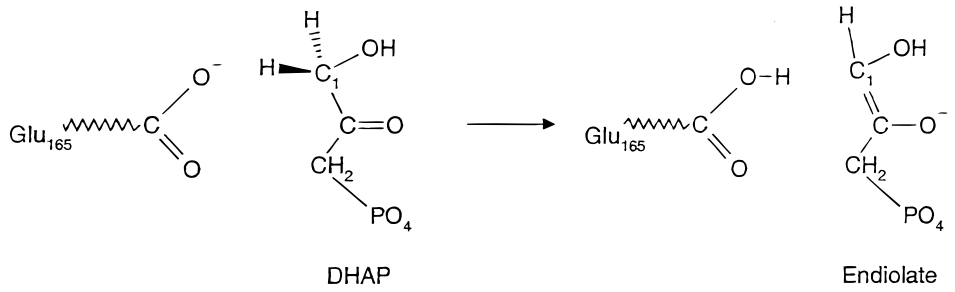

Figure 2. Initial proton transfer step in the catalytic mechanism of triosephosphate isomerase. A proton is transferred from the C1 carbon of DHAP to one of the carboxyl oxygens of GLU165 yielding an endiolate intermediate.

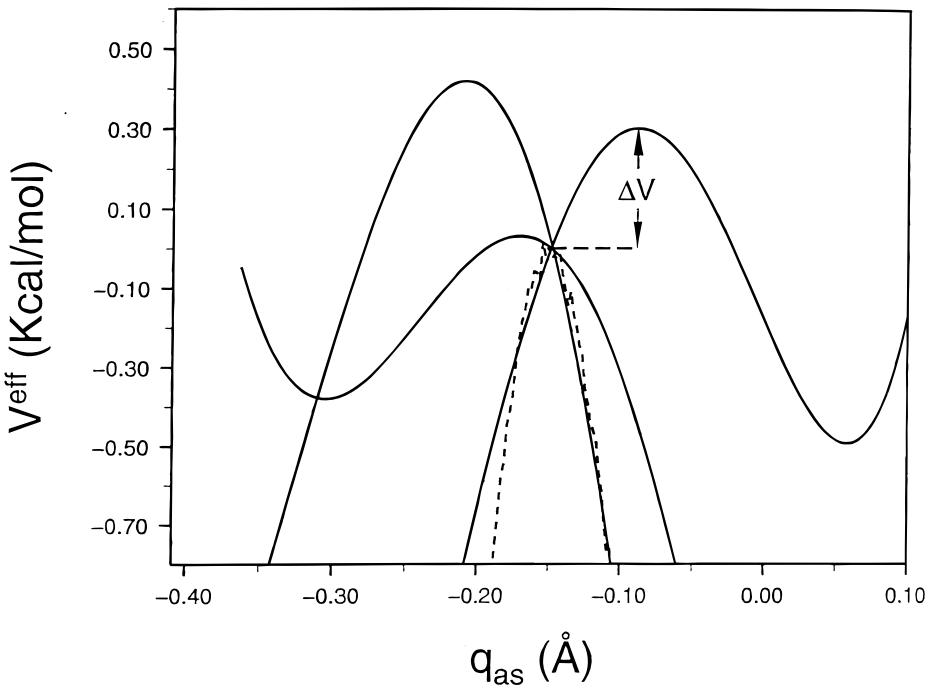

Figure 3. Potential of mean force near the transition state (- - -) and the potential energy for three initial configurations used in the calculation of the transmission coefficient. The potential energy was calculated with all degrees of freedom, other than the reaction coordinate, fixed. The zero of energy is at $q_{as}^{\#}$ and ΔV is the height of the instantaneous barrier that trajectories need to surmount.

In each state the potential surface is represented by a standard molecular mechanics (MM) force field. The possibility of reaction is introduced by coupling the two electronic states.[26,27] The potential function $V(\mathbf{R})$ is chosen to have the EVB form[20,27]

$$V(\mathbf{R}) = \frac{1}{2}(V_R(\mathbf{R}) + V_P(\mathbf{R})) - \frac{1}{2}\sqrt{(V_R(\mathbf{R}) - V_P(\mathbf{R}))^2 + 4V_{RP}^2} \quad (10)$$

where $\mathbf{R}$ is the set of all coordinates, $V_R(\mathbf{R})$ and $V_P(\mathbf{R})$ are the MM potential energies in the reactants and products states, respectively, and V_{RP} is approximated by a constant coupling, assumed to be independent of $\mathbf{R}$ for simplicity.

An appropriate reaction coordinate for the proton transfer from carbon C_1 of DHAP to the O of Glu 165 (see Figure 2) is the asymmetric stretch defined by

$$q_{as} = \frac{1}{m_C + m_O}(m_C r_{CH} - m_O r_{OH}) \quad (11)$$

where r_{CH} and r_{OH} are the distances of the proton from the donor carbon and the acceptor oxygen, respectively, and m_C and m_O are their masses; for a collinear arrangement q_{as} has a reduced mass of 0.97 amu. In the reaction, the proton moves from the vicinity of the donor carbon atom ($q_{as} = -1.1$ Å) to the vicinity of the accepting oxygen ($q_{as} = 0.3$ Å). To determine the location of the transition state, the potential of mean force (PMF)[28] as a function of q_{as} was determined by use of umbrella sampling.[29] As already described, the PMF is an effective potential for the reaction coordinate that is based on an equilibrium average over all other degrees of freedom. The resulting free energy profile for the reaction is shown in Figure 3.

As a starting point for the transition state trajectory calculations, equilibrium configurations in the transition state were sampled by running a 400 ps trajectory; $q_{as}(t)$ was constrained to $q_{as}^{\#}$ with a SHAKE-like algorithm,[30] and 40 configurations at 10 ps intervals were saved. For each of these configurations, 100 trajectories were initiated with different values of $\dot{q}_{as}(0)$ sampled from a Maxwell distribution. The function $\kappa(t)$ calculated from these trajectories is shown in Figure 4. It decays rapidly and has almost reached a plateau value by 10 fs. The plateau value is $\kappa = 0.43 \pm 0.08$, a significant reduction from the TST value of unity. However, the reduction by a factor of 2 is a small effect relative to the specific interactions in the protein that lower the potential of mean force barrier for the reaction so as to increase the rate by 5 orders of magnitude.

The rapid decay of $\kappa(t)$ to the plateau value makes it likely that most degrees of freedom coupled to the reaction coordinate do not change significantly during the time required to trap the trajectory in the reactant or product wells. Defining the substrate

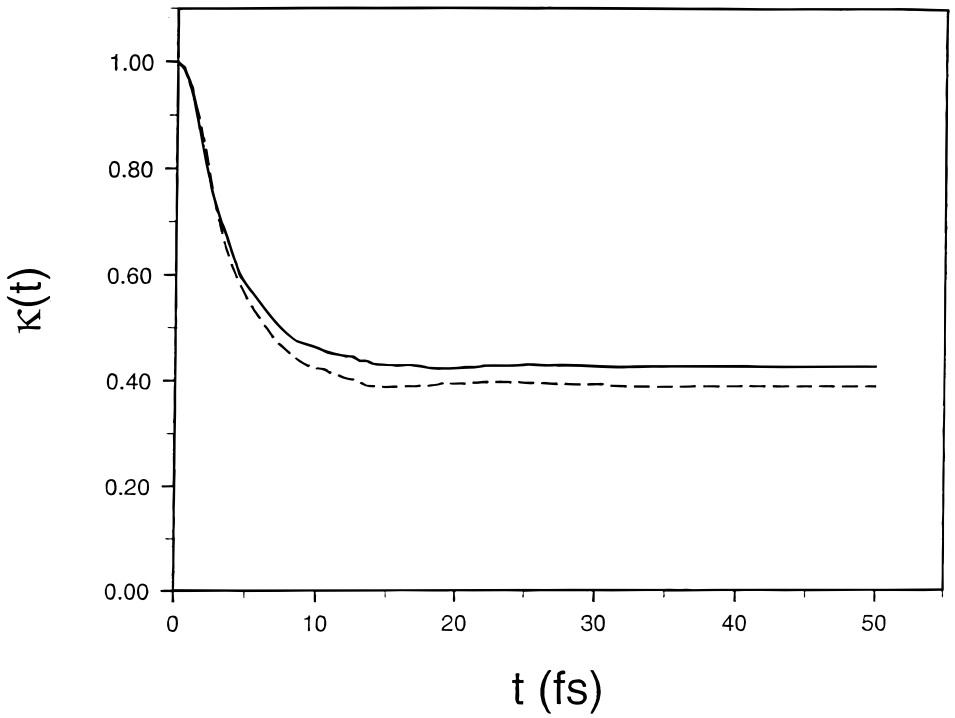

Figure 4. The time-dependent transmission coefficient, $\kappa(t)$, when all atoms are allowed to move (—) and in a rigid environment (- - -) (see text).

and the GLU165 atoms as the "intramolecular" part and the rest of the system as the "environment", the calculation of the transmission coefficient was repeated from the same initial configurations, but with only the intramolecular atoms allowed to move and the environment atoms fixed in their initial ($t = 0$) positions. The behavior of $\kappa(t)$ is very similar to that obtained when all atoms are allowed to move (Figure 4). This demonstrates that the dynamics of the environment atoms is unimportant for the barrier crossing per se and suggests that the structure at $t = 0$ determines whether a reaction takes place.

If all degrees of freedom other then the reaction coordinate are frozen ("frozen bath assumption") and there is a barrier ΔV, in addition to the potential of mean force barrier, for a particular set of bath coordinates (a configuration chosen from the 400 ps trajectory), the transition state trajectory has to surmount that barrier to complete the reaction; the range in ΔV observed in the calculations is about 0.5 kcal/mol (see Figure 3). This reduces the rate from the TST value, which is based on the potential of mean force barrier. With such a "frozen bath" assumption, the transmission coefficient is given by[31]

$$\kappa^{\text{froz}} = \langle e^{-\Delta V/k_B T} \rangle \tag{12}$$

where k_B is Boltzmann constant, T the is temperature, and the average is over all initial transition state configurations (Figure 3). The average over the 40 transition-state configurations gives $\kappa^{\text{froz}} \cong 0.40$, in good agreement with the activated dynamics simulations. This confirms that the frozen bath approximation is valid; that is, the surrounding system is fixed within the short time scale of the barrier crossing.

Two mechanisms have been proposed for the influence of the environment on the instantaneous barrier. The first, found in a study of a symmetric SN_2 reaction in water,[31] corresponds to a nonequilibrium solvation mechanism. In that reaction, the instantaneous barrier arises from different solvation states of the intramolecular subsystem; for example, a fluctuation in the environment that results in better solvation of the reactant relative to the products introduces an instantaneous barrier located between the transition state and the products state, and vice versa. In an alternative mechanism for the environmental influence, the instantaneous barrier is determined by the configuration of the intramolecular subsystem, but the intramolecular dynamics is coupled to low-frequency fluctuations of the environment; for example, intramolecular motions that are required for the reaction are hindered by the environment, which appears rigid on the time scale of the barrier crossing (~10 fs) but fluctuates on a picosecond time scale. The two mechanisms were tested for the proton transfer in the TIM reaction, and it was demonstrated that the second mechanism is dominant in the barrier modulation and in the reduction of the rate from the transition state limit. Which mechanism operates for the proton transfer in solution from DHAP to acetate, for example, and the magnitude of the effect (i.e., the value of $\kappa(t)$) have not been determined (Cui and Karplus, work in progress).

It is possible also that, in enzymes or in solution, a single reaction coordinate is not sufficient to describe the reaction, as discussed by reference to Figure 1. This appears not to be the case for TIM, unlike ligand rebinding to myoglobin after photodissociation (see section IV.B).

From the above analysis, there is a well-defined rate constant for the reactions; i.e., the reaction is simple according to the first criterion and eqs 2 are applicable. As the calculations have shown, there are three time scales for the reaction. There is the very fast time scale of relaxation of the high-energy transition state trajectories to a thermal distribution (~10 fs); there is the time scale of the motions associated with the enzyme environment that determines the height of the transition state barrier (~1 ps). The dominant modulations of the barrier are of a low-frequency character ($\omega < 300$ cm^{-1}) and originate from the interaction between the intramolecular subsystem and the rest of the enzyme. Finally there is the time scale associated with the overall reaction rate (~1 ms). Since the third is much longer than the other two, simple behavior, as exemplified by the well-defined plateau value for the transmission coefficient, is expected and was obtained in the calculations. However, it

should be noted that the trajectories covered a time scale (400 ps) many orders of magnitude shorter than that of the overall reaction rate (ms). Consequently, it is not excluded that there exist very slow structural fluctuations (on the same time scale or longer than the reaction) that alter the barrier. This would lead to nonexponential behavior, but it appears very unlikely at room temperature for such a "slow" reaction; at low temperatures, nonexponential kinetics might occur (see section IV.b, below). Interestingly, single molecule studies of alkaline phosphatase have shown significantly different activities for individual molecules.[32] However, it is likely that in this case there was a difference in composition (e.g., glycosylation) that was the source of the variation in reaction rates. In the future, single molecule techniques are likely to provide important new information in this area.[121]

Concerning the other aspect of simplicity, the Arrhenius-like temperature dependence, data for the reaction catalyzed by TIM are not available.

IV.B. Rebinding of Ligands after Photodissociation of Myoglobin. A well-documented case of complex behavior in a protein reaction is the geminate rebinding of ligands to myoglobin after flash photolysis.[1,33,34] This reaction has been studied experimentally over a wide range of temperatures and time scales and numerous theoretical analyses of the results have been made. Moreover, the high-resolution X-ray structures of liganded and unliganded myoglobin do not reveal any path by which ligands can move between the heme binding site and the outside of the protein.[35,36] Since fluctuations must therefore be involved in the overall ligand binding reaction,[1,2,37−39] myoglobin has become a model system for studying the relation of motion to function in proteins.

The phenomenological description of the kinetics of photodissociation and rebinding of a ligand X to myoglobin (Mb) can be written as

$$\mathrm{MbX} \rightleftarrows (\mathrm{Mb} + \mathrm{X})_{\mathrm{pocket}} \rightleftarrows (\mathrm{MB} + \mathrm{X})_{\mathrm{protein}} \rightleftarrows (\mathrm{Mb} + \mathrm{X})_{\mathrm{solution}} \quad (13)$$

where the subscripts refers to the location of the ligand X. Each of the designated species may involve several different states or substates on a microscopic level. At low temperatures (below 160 K in mixed glycerol/water), it is found that the geminate rebinding of the CO ligand [$(\mathrm{Mb} + \mathrm{X})_{\mathrm{pocket}} \rightarrow \mathrm{MbX}$] is nonexponential. As the temperature is raised, the geminate recombination of CO becomes essentially exponential. Thus, the first criterion for a simple reaction (eqs 1 and 2) is satisfied at room temperature but violated at low temperature. For NO, in contrast to CO, the geminate rebinding at room temperature is nonexponential;[40] the temperature dependence of NO rebinding has not been studied. Thus, the difference between the CO and NO behavior can aid in elucidating the source of complexity in this reaction.

The very long times over which the CO rebinding is nonexponential and time-delay flash experiments at low temperatures[1] suggest that an inhomogeneous model is involved. Thus, a description corresponding to the frozen bath model (i.e., protein molecules are in different substate with different activation barriers; see section IV.A), with the equilibration between bath configurations slow compared to the overall reaction, appears to be appropriate. This is in accord with equilibrium molecular dynamics simulations at 80 K[41] and incoherent neutron scattering experiments.[42,43] They indicate that protein fluctuations at such temperatures are relatively harmonic and correspond to oscillations in a single well or substate. This means that a distribution of barriers for a population of photolyzed myoglobin molecules is required for describing the observed reaction, rather than a single potential of mean force surface. The number of unliganded Mb hemes, $N(t)$, as a function of time after photolysis can then be written as

$$N(t) = \int_0^\infty \mathrm{d}E\, g(E) \exp[-k(E)t] \quad (14)$$

where $g(E)$ is the distribution of barrier heights and $k(E)$ is the rate constant for a barrier of energy E. Equation 14 results in nonexponential relaxation; that is, the rate coefficient is time-dependent and its form is determined by the choice of $g(E)$ and $k(E)$. Both power law or stretched exponential expressions for $N(t)$ can be fitted to the experimental data. The temperature dependence of the rebinding was obtained by assuming that $k(E)$ in eq 14 can be expressed in Arrhenius form (eqs 3) with a temperature-independent expression for $g(E)$; that is, the distribution of substates with different barrier heights was taken to be frozen in below 160 K within the precision of the experiments. Independent measurements of other markers of internal motion have demonstrated that certain relaxation phenomena in myoglobin[44,45] are nonexponential in time and have a super-Arrhenius temperature dependence that can be expressed in the form

$$k = A \exp[-(E/RT)^2] \quad (15)$$

over a wide temperature range. These results suggest that most of the rebinding below 160 K is a relatively local phenomenon in which protein relaxation does not play an important role. This is in accord with a dissociation simulation at 10 K (K. Kuczera and M. Karplus, unpublished), which demonstrated that the initial motion of the iron out of the heme plane still occurs in 40 fs (as it does at room temperature) and that at least some of the undoming required for rebinding can take place without protein relaxation.

For CO at room temperature, because of the large barrier to ligand binding, one does not even begin to observe recombination until about 100 ns after photodissociation.[33] Thus, the transitions between the substates are fast relative to the overall reaction and the protein is expected to have relaxed nearly completely to the unliganded configuration by the time significant CO rebinding occurs. A description of the type outlined in section III.A is expected to be applicable with $\kappa(t)$ having a well-defined plateau value. This implies that a rate constant should exist and exponential rebinding should occur, in agreement with experiment. At some intermediate temperatures (near 160 K), the rate of protein relaxation and rebinding will be on the same time scale, and complex behavior is expected. It is in fact observed that the rate decreases with increasing temperature in this range.[45,46] This is a clear manifestation of non-Arrhenius temperature dependence.

Because of the difference in the electronic structure of the two ligands, there is a much lower barrier for rebinding of NO than CO. This is exemplified by the fact that in myoglobin at room temperature 80% of the NO molecules have rebound after 300 ps.[40] An analysis of the Mb potential energy surface for the native state[2] based on a room temperature 300 ps molecular dynamics trajectory[47] demonstrated that the interconversion rates between substates range from 0.1 ps for those that are very similar to at least the length of the simulations (300 ps) for those with larger structural differences; experiments[48,49] indicate that there are relaxation processes that extend from femtoseconds into the microsecond range. These results show that the substate transitions are on the same time scale as protein relaxation after photolysis in myoglobin, as would be expected since they cover

the same conformational space. In fact, the rms difference between MbCO and Mb in crystal structures is 0.3 Å for the main chain and 0.7 Å for all non-hydrogen atoms,[36] somewhat smaller than the conformational space (about 2 Å) explored by the single Mb simulation at room temperature.[2]

From these results, it is clear that the rigid bath model used to explain nonexponential rebinding of CO at low temperatures is not applicable to NO at room temperature. Instead, the similarity of the time scales of the protein motions and the rebinding rate suggests that there is a coupling between the two. This is likely to arise from the difference of the position of the heme iron relative to the heme plane in the liganded and unliganded structures; that is, the heme is essentially planar with the six-coordinate iron in the heme plane in the former and the heme is "domed" with the iron out of the heme plane in the latter. The barrier to rebinding is expected to be lowest when the five-coordinate iron is in the heme plane and to increase as the heme moves out of the plane and the protein relaxes to the unliganded equilibrium structure. To examine the possibility that the nonexponential rebinding of NO can be produced by a time-dependent evolution of the barrier, we describe a simple model; simulations that support this model have been made (Becker and Karplus, to be published). The disappearance of unliganded protein, $N(t)$, is assumed to follow the modified first-order decay law,

$$-\mathrm{d}N/\mathrm{d}t = k(t)N \tag{16}$$

where $k(t)$ is a time-dependent rate coefficient which is assumed to obey the modified Arrhenius equation

$$k(t) = A \exp[-E(t)/RT] \tag{17}$$

with

$$E(t) = (E_0 - E_{eq}) \exp(-k_{bar}t) + E_{eq} \tag{18}$$

Here k_{bar} is the rate constant that determines the decay of the initial barrier height, E_0, to its equilibrium value, E_{eq}. Reasonable parameters (i.e., $E_0 = 0.027$ kcal/mol, $E_{eq} = 1.2$ kcal/mol, $k_{bar} = 1.5 \times 10^{10}$ s^{-1}) give a good fit to NO rebinding for myoglobin at room temperature; the limiting rate constants are 42 ps at $t = 0$ and 300 ps at $t = \infty$.[40]

The model just described requires that there be a slow component to the iron motion out-of-the heme plane. The existence of such a component has been demonstrated in simulations by Kuczera et al.;[50] interestingly, simulations with somewhat different protein models[51,52] did not find a nonexponential component. The simulations of Kuczera et al.[50] showed that after photodissociation there is an ultrafast relaxation process with an average time constant of about 40 fs; the range of relaxation times varies from 30 to 70 fs in different trajectories. It consists of an iron out-of-plane displacement coupled to local adjustments of the heme. The ultrafast behavior is very similar to that obtained following the dissociation of a heme−histidine−CO complex in the absence of the protein. The remaining (slower) processes correspond to structural and energetic relaxation involving the protein, as well as the heme. This is due to the fact that the doming of the heme group exerts a force on the surrounding protein atoms. The iron out-of-plane displacement averaged over several trajectories, $\langle z(t)\rangle$, which was used as the reaction coordinate for the structural transition leading from the liganded to the unliganded state, shows nonexponential behavior. Over the time scale of the simulations (100 ps) the results can be fitted to a power law or a stretched exponential. Figure 5 shows the calculated relaxation behavior;

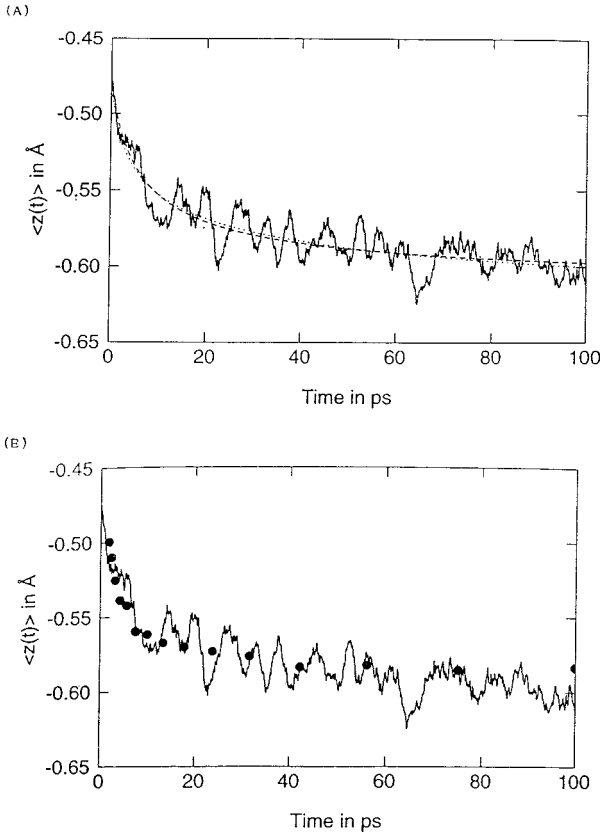

Figure 5. Myoglobin photodissociation showing the out-of-plane motion of iron, $\langle z(t)\rangle$ in angstroms, as a function of time in picoseconds. The solid line is the average of four simulations smoothed with 2 ps windows: (a) fit with a power law, two slightly different fits are shown; (b) superposition of the simulated $\langle z(t)\rangle$ for the out-of-plane motion of the iron and experimental $\Delta\gamma(t)$ for the spectral shift of band III (measurements by Lim et al.).[48] The experimental data (filled circles) were subjected to an overall linear scaling (see text). Adapted from Kuczera et al.[50]

the fluctuations in $z(t)$ correspond to high-frequency local oscillations superposed on the global relaxation. Also shown is the measured frequency shift, $\Delta\nu(t)$, of band III,[48] scaled in magnitude to correspond to the amplitude of $\langle z(t)\rangle$. Band III is a heme charge-transfer band whose frequency is related to the iron out-of-plane distance. The nonexponential time dependence of $\Delta\nu(t)$ and $\langle z(t)\rangle$ are in excellent agreement over the range of the simulation. This supports the simulation results and indicates as well that the band III shift is a linear function of the iron out-of-plane displacement. Steinbach et al.[45] have suggested that there is also a linear relationship between the barrier height for rebinding and the position of band III. Further, the temporal evolution of $\langle z(t)\rangle$ from the simulations is in agreement with time scales found in the experimental studies of NO ligand rebinding.[40]

The position of band III is a marker for conformational relaxation of myoglobin after photodissociation. It has been shown, as already mentioned, that the shift in band III is nonexponential and takes place over a time scale from femtoseconds to microseconds.[48,49] Hagen and Eaton[53] have developed a model for the nonexponential decay to equilibrium of a protein after a perturbation, such as the photodissociation reaction. The model is based on the assumption of two relaxation time scales: one (faster) being the transition from the liganded to a certain portion of the unliganded manifold of states and the other (slower) the redistribution of states to the equilibrium population in the unliganded manifold. To obtain a description that can be

used for simple calculations, Hagen and Eaton assumed a Gaussian (random energy) model for the distribution of energies in the unliganded and liganded manifold of states and introduced a single transition state[54] that governs the rates of all transitions (within each manifold and between manifolds). With reasonable parameters, they are able to fit the model to the stretched exponential type behavior of the measured shift of band III. The details of the relaxation behavior depend on the assumption of the connectivity between states; i.e., whether any state can make a transition to any other state, as in the protein folding model of Bryngelson and Wolynes[55] (see also ref 56) or whether the allowed transitions are restricted to states with similar energies. A major question in this type of model concerns the relation between energy and structure; that is, what is the structure of the postulated single transition state, if one exists, and is it possible to assume that states of similar energy have similar structures since any physical transition will most likely involve two similar structures. Hints concerning the complex nature of the behavior that can be expected are given by the topological and kinetic analysis[57] of the potential surface calculated for an alanine tetrapeptide by Czerminsky and Elber,[58,59] as described in section V.

The present analysis suggests that there are two limiting possibilities for protein reactions, in general. In one limit, many protein substates with different barrier heights exist at equilibrium and it is the substate interconversion rate versus the reaction rate that determines whether exponential behavior is observed. In the other limit the barrier and binding rate are altered by protein relaxation from the liganded to the unliganded structure after dissociation so that the relative rate of relaxation and rebinding determines whether the latter shows nonexponential kinetics. The two possible sources of nonexponential behavior are not mutually exclusive and are likely to play a greater or lesser role, depending on the nature of the ligand, the protein and the temperature, as well as other conditions specifying the system. Single molecule measurements would be useful to distinguish between these possibilities.[53,60]

Another protein reaction for which nonexponential kinetics has been observed is the fast primary electron-transfer step in the bacterial photosynthetic reaction center.[126] The origin of this nonsimple behavior has been analyzed by use of molecular dynamics simulations.[127]

I have gone into considerable detail in reviewing the myoglobin rebinding after photolysis because it is the protein reaction for which the most detailed experimental and theoretical studies exist. Nevertheless, many questions concerning this well-studied reaction remain unanswered.

IV.C. Specific Studies of Conformational Change. There are many reactions in proteins involving conformational change. It is likely that one of the reasons for having molecules as large as proteins carrying out important functions in living systems is that this makes it possible to have the well-defined conformational transitions needed for control and signaling. A conformational change in the context of this section is a reaction in which the initial and final states have well-defined structures, unlike the protein folding or unfolding reactions, where one of the "states" is a superposition of a very large number of different conformations; this very interesting case is considered in section V. The most common examples of conformational changes, of which there are many, involve an alteration in the relative stability of two structures due to the binding of a ligand or a substrate. Such conformational changes include lid opening and closing motions (e.g., triosephosphate isomerase,[61] loop reorientation (e.g., as in rasp21),[62] hinge-bending motions (e.g., as in lysozyme[63] and in GroEL subunits),[64] and the relative motion of subunits (e.g., as in the quaternary transition of hemoglobin).[65] For many of these cases, a reaction coordinate would be relatively easy to define. This contrasts such transitions with those concerned with globally distributed conformational changes that involve motions of all or most of the atoms of the protein, as in the transition of myoglobin from the liganded to the unliganded state discussed above (section IV.B) and the protein folding transition considered below (section V).

IV.C.i. Ring Flips in Bovine Pancreatic Trypsin Inhibitor (BPTI). This conformational change whose dynamics has been simulated in detail is a "degenerate" one, in that the initial and final states are the same. It consists of aromatic ring isomerization (ring "flips") in the BPTI.[15,66−68] Room temperature activated dynamics simulations very similar to those described for triosephosphate isomerase (section IV.A) were performed for the tyrosine ring (Tyr 35), which has the slowest flipping rate of the four tyrosines in BPTI. Even for this apparently simple case, the obvious reaction coordinate, corresponding to the ring dihedral angle, is not satisfactory; instead, a more complex reaction coordinate that takes account of essential coupling between the ring and the local protein backbone was introduced. On the basis of this reaction coordinate, the transmission coefficient was calculated (eq 9) and a well-defined plateau value equal to 0.22 was reached in 0.1 ps. Thus, the calculations indicate that the conformational change shows simple behavior, at least at room temperature. This is as expected because the calculated potential of mean force barrier for the ring flip is on the order of 10 kcal/mol and the barrier crossing rate is much slower than the collisional relaxation of the activated ring kinetic energy.[15] It is not unlikely that, at a lower temperature, nonexponential behavior would be observed, in analogy to that found for ligand rebinding in myoglobin. However, such low-temperature behavior has not been studied theoretically or experimentally. The experimental data on the ring flip from nuclear magnetic resonance spectroscopy have been used with the assumption of simple kinetic behavior of the first and second type to obtain the reaction rate.[67] The measured values of the activation enthalpy and entropy above 300 K are 37 kcal/mol and 68 eu, respectively. This contrasts with calculated values of 10 kcal/mol for the enthalpy and approximately zero for the entropy. Such a large disagreement suggested that the experimental separation into enthalpy and entropy of activation, based on the assumption that both are temperature independent, may be in error[69] (see also section V.B). In fact, the measured rates as a function of temperature showed significant curvature, suggesting that complexity of the second type is present. Otting et al.[70] have remeasured the Tyr 35 ring flips as part of a study of the equilibration of two conformers of the Cys 14−Cys 38 disulfide bond of BPTI. In the range 299−323 K, similar to that used by Wagner et al.,[67] the measurements (33 kcal/mol and 53 eu) confirmed the activation parameters found in the earlier work; as before, a kT/h pre-exponential factor was used. For the same temperature range, the disulfide conformational exchange had "similar" activation parameters (values are not given in the paper). What is interesting is that, at lower temperatures (277−293 K), the disulfide activation parameters were found to be very different; the activation energy was on the order of 10 kcal/mol and the activation entropy was small and negative (on the order of −15 eu). Thus, there is a striking difference between the low- and high-temperature disulfide conformational transition with the former having activation parameters much closer to those calculated for the ring flip; unfortunately, the Tyr 35 ring flip

rate could not be measured at the lower temperature. Otting et al.[70] suggested that at high temperatures the disulfide transition and ring flip are concerted to explain the increased activation parameters. This could indicate that the discrepancy between the earlier experiments and the calculations for the ring flip arises from the fact that the measurements represent the high-temperature process, while the calculations correspond to the low-temperature process; indeed no disulfide transition was observed in the ring flip calculation. It is likely, therefore, that the ring flip transition and the disulfide reorientation are not simple by the second criterion; i.e., in this case, the reaction appears to follow a different path at low and high temperatures and the activation parameters have a complex temperature dependence, even though the native structure of the protein is essentially independent of temperature.

IV.C.ii. Triosephosphate Isomerase (TIM) Lid/Loop Transition. For most other cases, where the occurrence of a conformational change is demonstrated by the observation of two different crystal structures (see, for example, Gerstein et al.[71] for a list and classification of many conformational changes in proteins), there are no measurements of the dynamics. One case, which has been studied experimentally and theoretically, is the conformational change of a loop region that protects the active site of TIM from solvent during the reaction. We briefly describe the results because they demonstrate what is known and what is not known about a rather localized conformational change in a protein. An analysis based on the "open" and "closed" crystal structures of the enzyme suggested that the peptide loop motion in going from one structure (unliganded, open) to the other (liganded, closed) is best described as a rigid "lid" motion; i.e., the internal conformation of the loop/lid does not change in the transition and certain residues at the two ends of the lid can be identified as hinges by use of pseudodihedral angles based on the Cα atom of each residue.[61,72] High-temperature molecular dynamics simulations were required to go from a closed to an open structure (analogous to protein unfolding; see section V) in a reasonable simulation time. It was found that the loop opens and closes in a "jump-like fashion" within approximately 20 ps at 1000 K. This suggests that there is a barrier to opening even in the absence of substrate or product. Wade et al.[73] simulated the loop motion by Brownian dynamics in the presence of the electrostatic field due to the rest of the (rigid) protein. They used a simplified representation of the loop residues, analogous to that employed by McCammon et al.[74] in Brownian dynamics simulations of α-helix unwinding. Use of such a model permitted simulations as long as 100 ns, which would have been impossible with a more detailed description. The simulations showed that the opening and closing motion occurred on a 1 ns time scale. Moreover, many dihedral transitions within the loop lid were observed, in contrast to the analysis of Joseph et al.[61] A recent simulation of the lid motion[75] with an all-atom model using stochastic dynamics supports the rigid-lid description and suggests a sizable activation barrier for the transition.

There have been two experimental studies aimed at the dynamics of the loop/lid motion. One of these uses line shape analysis of the quadrupole powder pattern determined by solid-state deuterium NMR[76] to probe the motion of the perdeuterated indole side chain of Trp 168. This residue is near the end of the loop (166−176); in fact, one of the hinges involves motion about the pseudodihedral angles 166−167 and 167−168, though the displacement of the former is larger than that of the latter. The experimental data were interpreted by assuming displacements of the Trp 168 side chain that correspond to the difference observed in the crystal structures; with this assumption, a transition rate of about 10^4 s^{-1} was obtained. This is much slower than that observed in the Brownian dynamics simulation. Using a pre-exponential factor of 10^{12} s^{-1}, which may not be appropriate for what is likely to be a diffusive transition, an activation energy of 12 kcal/mol was estimated; no temperature dependence measurements were made to confirm this. Somewhat surprisingly, the same rate was observed for the free enzyme and that with the bound inhibitors. Also, the same ratio of 10 to 1 between two forms (assumed open to closed for the free enzyme and closed to open for the inhibited enzyme) was obtained by fitting the spectra. An obvious complication in the interpretation is that there is no direct evidence that the two conformations used to fit the data are actually the open and closed form of the loop, rather than two unrelated slowly interconverting conformations of Trp 168, such as have been found to exist in studies of tryptophane fluorescence in other proteins. With a barrier of 12 kcal/mol, the 20 ps time for the transition found in the molecular dynamics simulation at 1000 K yields a rate of about 10^4 s^{-1} at room temperature, close to that estimated from the NMR experiment. One other study examined the temperature dependence of a TIM transition by using ^{31}P NMR of Glu 165 modified covalently with the substrate analogue 3-chloroacetol phosphate.[77] Signals corresponding to two conformations were observed (again there is no direct evidence that they represent the open and closed lid), and the temperature dependence of the transition rate between them was fitted to an Arrhenius plot; it yielded a barrier of 34 kcal/mol. The actual data indicate significant curvature, though it is difficult to determine whether this is meaningful since only three temperature points were measured.

In summary, it is evident from the differences in the results obtained from the various experiments and simulations that the present understanding of this transition is incomplete and that more work is needed to determine the details of its behavior, including whether the reaction can be described as simple or complex.

IV.C.iii. Quaternary Transition in Hemoglobin. A conformational transition that has been studied at various levels of detail over many years is the quaternary transition in hemoglobin. Many measurements of the kinetics of the transition have been made because of its important role in hemoglobin function, as exemplified by the Monod−Wyman−Changeaux model.[78] From the structural data of Perutz and co-workers,[65] it is known that there are two quaternary structures, often referred to as T and R, with the former more stable in the unliganded state (no O_2 bound) and the latter more stable in the fully liganded state (one O_2 bound to each heme). Since both tertiary and quaternary transitions play a role in the different states of hemoglobin, a full analysis of the kinetics is rather complicated.[79,80] However, in the present report we are not concerned with these aspects. To a good approximation, the quaternary transition of the hemoglobin tetramer can be described in terms of two $\alpha\beta$ dimers ($\alpha_1\beta_1$ and $\alpha_2\beta_2$) that undergo a relative rotation of 15° with respect to each other; there is also a small relative translation coupled to the rotation.[81] The most complete kinetic data are available for component I of trout hemoglobin, whose quaternary transition is similar to that of human hemoglobin.[79] The experiments were done by photolysis of hemoglobin fully liganded with CO, which provides a trigger analogous to that used in the myoglobin experiments described in section IV.B. It was possible to isolate spectrally the R to T transition of the fully photodissociated tetramers (i.e., those with no CO bound). A simple rate expression was assumed in the analysis of the

data; there is not enough information to determine a nonexponential time dependence for the reaction, even if it were to occur (W. Eaton, private communication). Measurements of the temperature dependence over the range 275−338 K showed simple Arrhenius behavior. Thus, within the limits of the available experimental data, the quaternary transition appears simple. The measured activation energy was 8 kcal/mol. Use of this value, a gas-phase pre-exponential factor of (kT/h), and a transmission coefficient of unity yielded an activation entropy equal to -11.7 cal mol^{-1} K^{-1}. However, a simple diffusive model is probably better for obtaining the pre-exponential factor. The Kramers equation in the high friction limit[82,83] and a simplified sphere representation of the $\alpha\beta$ dimer with a 20 Å radius yield a time of $\sim 10^9$ s^{-1} for the 15° motion involved in the R to T transition, in the absence of an activation barrier. This suggests that the factor $\kappa(kT/h)$ is significantly smaller than the value used by Hofrichter et al.,[79] in agreement with a suggestion in their paper. Use of eq 4a with this pre-exponential factor yields a positive entropy of activation. This seems more reasonable since simulations suggest that there is a loosening of the tetramer in the transition (Fischer et al., unpublished).

V. Protein Folding

Protein folding is the protein reaction of greatest complexity. The entire molecule is directly involved in the reaction, unlike some of the cases considered above. Moreover, the nonbonded van der Waals and electrostatic interactions that determine the potential energy surface for protein folding and lead to the stability of the native state are individually weak (between 0.1 and 2−3 kcal/mol). If one is considering effective energy surfaces or potentials of mean force surfaces, these energy terms correspond to the effective interaction energy between pairs of atoms in the presence of a canonically averaged solvent and so depend on temperature.[84] There are a very large number of these contribution; even for a small protein like the well-studied C terminal fragment of CI2 with 64 residues, there are on the order of 10^5 potential energy terms if they are represented as atom−atom pair interactions. Nevertheless, the free energy difference between the native and denatured state is only on the order of tens of kcal/mol at ambient temperatures (e.g., 7 kcal/mol for a rather stable protein like CI2 at 298 K),[87] which corresponds to about 0.4 kcal per residue and 0.04 kcal/atom (the enthalpy of unfolding is significantly larger; it is about 30 kcal/mol for CI2 at 298 K).[87] Clearly, a delicate balance of interactions is involved in the effective energy and free energy surface. Another element in the complexity of the protein folding reaction is the large number of conformations accessible to the polypeptide chain. This has two aspects. One, of course, is the search problem, embodied in the so-called Levinthal paradox,[85,86] and the other, closely related, is the possibility that the configurational entropy has an important role in determining the activation barrier.

Despite the microscopic complexity, most measurements of folding kinetics[87−89] indicate that the rate of formation of the native state obeys the simple unimolecular rate law given in eqs 1 and 2. More specifically, the observed rates of folding (and unfolding) can be represented either by a simple rate law or by a kinetic scheme consisting of a small number of intermediates, in addition to the denatured and native state, and each step of the overall reaction can be described by a simple rate law.[9,90] Thus, the kinetics of the protein folding reaction appears to be simple in terms of the first criterion; i.e., an exponential time dependence for the folding rate is observed within the accuracy of available experiments. However, the second criterion, that the temperature dependence obeys an Arrhenius law, is not satisfied by the protein folding reactions whose temperature dependence has been studied.[91] It appears to be Arrhenius-like at physiological temperatures, but very clearly deviates from Arrhenius behavior at higher temperatures. Thus, essential questions concerning protein folding are why the reaction is simple in terms of the first criterion, despite the complexity of the energy surface and the many degrees of freedom involved, and why the protein folding reaction is complex in terms of the second criterion.

V.A. Exponential Time Dependence of Protein Folding. From the discussion of reaction rate theory (see section III.A) and its applications to reactions involving proteins, a number of points are clear. A simple kinetic rate law is applicable if and only if there is a separation of time scales; that is, the rate of the reaction (measured as flux passing over a barrier) is slow compared to the elementary collisional events that lead to equilibration of the other degrees of freedom of the system. This requirement is expected to be satisfied if there exists a single significant free energy barrier (several times kT) for the reaction. Experiments show that, at room temperature, proteins have activation free energy barriers for the folding reaction that are large by this criterion (for CI2, the value of $\Delta G^{\ddagger}$ based on the measured rate and an assumed pre-exponential factor of kT/h is about 15 kcal/mol at 298 K, but see below)[91] so that the observed simple exponential behavior is in accord with this qualitative criterion. However, it raises the question of why there is *a* dominant free energy barrier, if there is one, a question which does not have a simple answer. There is a biologically relevant reason for unfolding to be endothermic so as to satisfy the thermodynamic criterion for a unique native structure that is stable under physiological conditions. However, this does not provide a reason for having a large (several kT) activation barrier for folding.

In previous sections (sections I and IV), we described the multiminimum surface of the native state and its effects on reactions in native proteins. It is very likely that the polypeptide energy surface for non-native states is also a complex multiminimum surface. However, no detailed calculations of effective potential energy surfaces for proteins, including a satisfactory sampling of the non-native conformations, are available. One of the few studies of a complete potential surface deals with the alanine "tetrapeptide".[58,59] Even this ultra small system with only seven soft dihedral angles has a number of points of interest for the protein folding problem. The surface has 139 minima and 502 transition states. A diagram of a part of the surface is shown in Figure 6.[57,92] The minima can be organized into local basins by an energy criterion. Each basin contains several minima but the barriers between them are such that there is generally rapid equilibration within a basin, relative to the passage from one basin to another. A kinetic analysis of the approach of the system to the native state from an initial distribution over the denatured states shows approximately exponential behavior at 300 K. If such results can be extrapolated to the effective energy surface for protein folding, they suggest that, despite the existence of many minima, there is a dominant barrier ("the" barrier to folding) and that transition rates connecting different basins are fast relative to the passage along the reaction coordinate (progress variable) over the "rate determining" barrier. The overall folding kinetics should then obey a simple rate law, possibly complicated by intermediates, and the global description in terms of simple kinetic schemes is a useful one, even if the species involved may represent populations with a wide range of structures.[93]

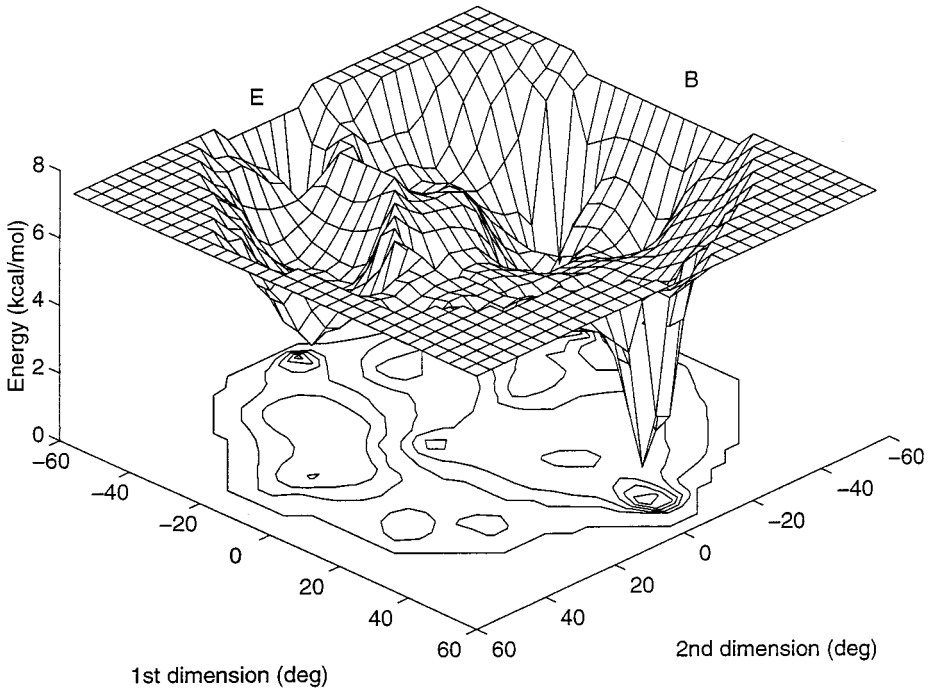

Figure 6. The alanine "tetrapeptide": a simplified diagram of the energy surface of the main basin, which includes the minimum energy conformer, projected on to two principal components (a reduced representation relative to the seven soft dihedral angles specifying the conformation). It is interesting to note that the accessible region is rather broad down to an energy of about 4 kcal mol^{-1}, when the different basins separate. Adapted from Becker.[92]

The most direct support for this view of protein folding were originally provided by simulations of a random heteropolymer 27-bead model on a cubic lattice;[94,95] closely related studies have been made by Socci et al.,[96] and Chan and Dill[97] have used a similar model on two-dimensional square lattices. Although oversimplified relative to actual proteins, such models have important protein-like properties; that is, they have a large number of conformations (e.g., 10^{16} for a 27-mer) and a unique ground state; the "unique" ground state in this model encompasses the large number of substates with very similar structures (within 2 Å rms) that in reality contribute to the native state of proteins and have a role in the complex kinetic behavior exhibited in the ligand rebinding to myoglobin, for example (see section IV.B). Unlike more detailed protein models, calculations of the folding kinetics and thermodynamics are sufficiently fast that they can be explored in detail for a large number of sequences at different temperatures. Figure 7 shows the calculated kinetics for a 27-mer sequence.[86,93,98] For all the temperatures studied, it is found that the rate of folding is exponential; that is, the heteropolymer folding reaction is simple by the first criterion. This is true even though the time dependence of the decay of the correlation function for the structural overlap between coordinate sets is best fitted by a stretched exponential with the exponent of the time near 0.5, corresponding to an approximately diffusive time dependence (Šali and Karplus, unpublished). For some designed sequences (i.e., sequences that are optimized by selecting the interactions to speed up folding) more complex folding behavior has been found;[99] at high temperatures, folding can be described by a single exponential, but at lower temperature there is both a fast and slow exponential phase. This was explained, in analogy to interpretations of experiments on protein folding, by the existence of an (off-pathway) intermediate.

The effective free energy surfaces at two different temperatures drawn as a function of the progress variable Q, which is the fractional overlap (number of native contacts divided by the total number of native contacts) with the native state are in

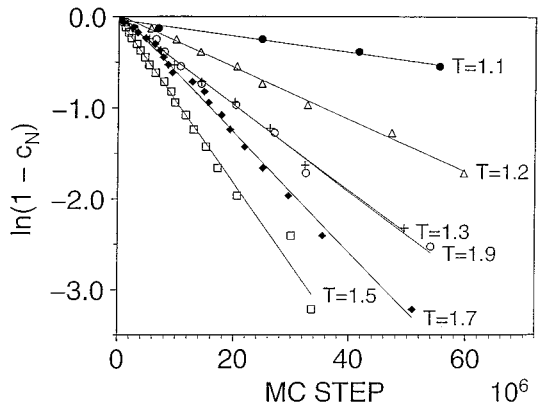

Figure 7. Distribution of the mean first passage folding times for a 27-mer sequence; c_N represents the native fraction. The distribution of the mean first passage times is shown at the temperatures indicated on the plot. The lines are linear least-squares fits to the points. To obtain each of the points, 100 independent folding trials were done. Adapted from Karplus et al.[98]

accord with the calculated simple rate behavior (see parts b and d of Figure 8). For both cases, the free energy surface has a significant barrier near the native state that dominates the folding kinetics. In the (high- and low-temperature) limits shown (for the more general temperature dependence of the surfaces, see Šali et al.),[95] the free energy surface appears relatively smooth and has a single important barrier, as required for the validity of a phenomenological rate equation. This is true even at the lower temperature for which the effective energy surface is quite "rough" (see Figure 8a) due to a partial cancellation between the "roughness" in the energy and entropy.[93,94]

In Figure 8, we have used Q, the fractional overlap with the native state, as the progress variable to describe the free energy of the reaction. Because of the complexity of the protein folding reaction and the many degrees of freedom involved, it is much less simple to define a reaction coordinate than for small molecule reactions. In fact, it is likely that a reaction coordinate

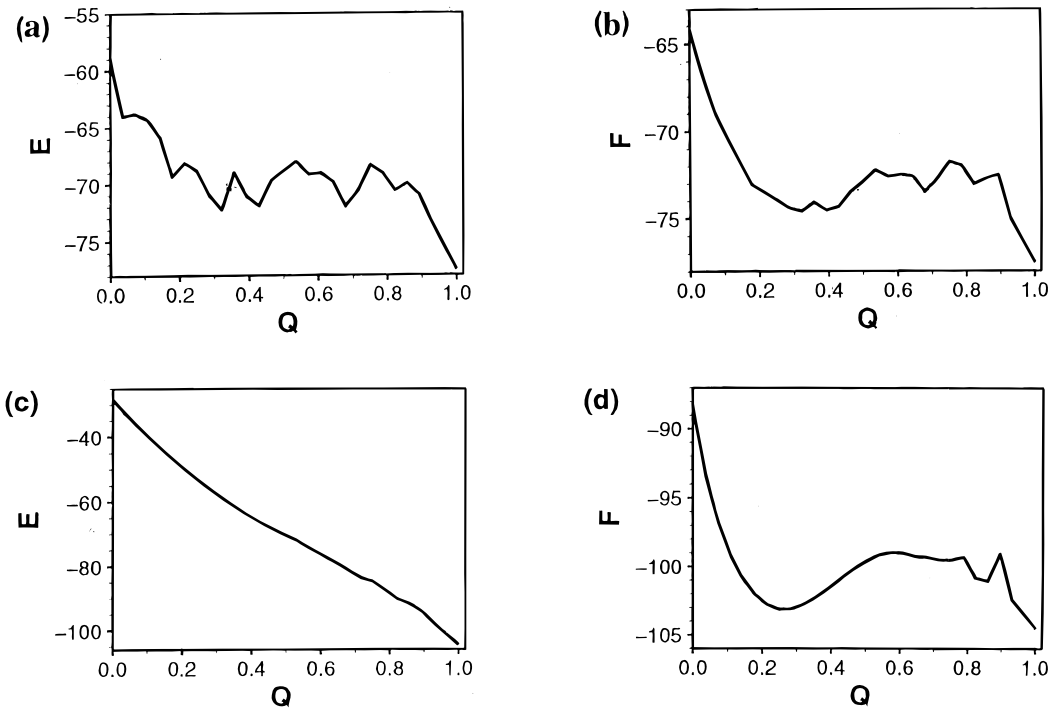

Figure 8. Folding effective energy and free energy surfaces of a 27-bead heteropolymer on a cubic lattice. (a) The average effective energy, E, as a function of Q, the fraction of native contacts at a low temperature ($T = 0.7$); there are 28 contacts in the native state, which is a $3 \times 3 \times 3$ cube. (b) The average effective free energy, F, as a function of Q for $T = 0.7$. (c) and (d) are the same as (a) and (b) but at a high temperature ($T = 2$). Results adapted from Figure 4 of Šali et al.[95] Details of the definition and methods are given in that paper.

corresponding to that for small molecule reactions does not exist; that is, there is no well-defined linear combination of coordinates that describes the motion from the reactant to the product state. This follows if the folding trajectories of individual polypeptide chains are very different and cover a wide range of possibilities. Clearly, if the fractional number of native contacts, Q, is used to specify the progress of the reaction, many different sets of contacts correspond to the same value of the progress variable, except for the unique native state. (For the use of Q as the progress variable in a protein unfolding simulation with an atomic model, see Lazaridis & Karplus.[100]) It is not self-evident that motion along Q is always an appropriate description of the protein folding reaction. There is the question of whether motion along Q corresponds to the slow step so that the effective free energy as a function of its value can be used to analyze the reaction; e.g., to show that there is a single dominant free energy barrier. For the random 27-mers studied by Šali et al.,[94,95] consistent results are obtained with Q as the progress variable. Moreover, the transition states as defined by the free energy barriers (see parts b and d of Figure 8) correspond to the values of Q, where the probability of reaching the native state increases rapidly from essentially 0 to about 0.5 (Šali and Karplus, unpublished). The choice of Q for the reaction coordinate of short chains, as in the 27-mers, has been supported by the work of Socci et al.[96] and Pande et al.[101] For longer chains, more than one progress variable may be required to describe the folding process (see section III and Figure 1b).[93,102] Alternative progress variables for protein folding have been used by Du et al.[103] and by Chan and Dill.[104] For the 27-mers, Q satisfies the conditions for a useful progress variable or "transition coordinate" suggested by Du et al.[103] Chan and Dill[104] found cases in lattice folding simulations where the condition of a separation of time scale for the reaction coordinate and the other coordinates of the system appears not to be satisfied; it would be interesting to know whether the folding rate is nonexponential in that case. In this regard, it is important to point out again that for the validity of the simple phenomenological rate expression, it is not necessary to "know" the appropriate progress variable; it is necessary only that such a coordinate exists and that motion along it is slow relative to the relaxation of other degrees of freedom of the system.[3] However, to do meaningful calculations or interpret the results at the atomic level, a knowledge of the appropriate slow variable that describes the reaction is necessary.

Given the simulations, which demonstrate simple kinetic behavior of the first type (eqs 1 and 2), a number of models have been introduced to justify the experimental and theoretical results. The essential idea[97,101,105,106] (Neria and Karplus, unpublished) is that the native state is coupled to the very large number of configurations making up the denatured state by a much smaller number of "doorway" states. The analysis of the 27-mer lattice simulations at low temperature suggests on the order of 1000 such doorway states (the "transition state ensemble"). If the doorway states have similar free energies and the equilibration among the denatured states is fast relative to the motion that leads to the doorway states, a transition-state-like expression for the flux through the doorway states can be derived and an expression for the rate constant can be obtained. Several of the models[97,105] (Neria and Karplus, unpublished) have used a master equation approach to calculate the kinetics. A somewhat different model[106] has elements in common with that of Hagen and Eaton[53] for the conformational relaxation in myoglobin. The two essential assumptions of the model of Gutin et al.[105] are that the mean energy of the system relaxes rapidly (fast time scale) to the equilibrium value given by the random energy model; below a critical temperature the equilibrium energy is essentially that of the native state. On a slower time scale, the system finds one of the small number of transition states from which it folds rapidly to the native state. It is assumed that the transition states all have essentially the same energy, in line with the kinetic models of Kopfer and Hilhorst,[54] Hagen and Eaton,[53] and Savens et al.[56] The model differs from

that of Hagen and Eaton[53] in that the equilibration of degrees of freedom other than the reaction coordinate is rapid so that exponential kinetics is obtained.

It is important to consider briefly the pre-exponential factor in eq 4a since, given the activation enthalpy of the reaction from a measurement of the rate as a function of temperature (eq 4b), the activation entropy and activation free energy depend on its value. Although the gas-phase (Eyring) value, kT/h, which equals $\sim 6 \times 10^{12}$ at 300 K, is often used, it is not generally valid for solution reactions, as has been mentioned already in section IV. For the motions involved in the folding reaction, the system is likely to be in the diffusive regime because of size of the particles (residues or larger groups). However, the microscopic complexity is such that a range of different structural entities may be involved during various portions of the folding reaction. This is clearly one of the problems that still needs to be resolved. In fact, it may be very difficult to determine the appropriate "average" value of the pre-exponential factor for use in eq 4a. Values such as 10^{10} s^{-1} [89] and 6×10^{12} s^{-1} [91] have been used, but there is little justification for these or any other values. Since the elementary step in residue-based treatments of protein folding (e.g., Monte Carlo lattice simulations) is likely to be a dihedral or pseudodihedral angle transition, a value in the neighborhood of 10^9 s^{-1} appears to be suitable for that case.[74] However, if larger entities are involved (e.g., as in the diffusion−collision model[107]), values closer to 10^7-10^8 s^{-1} may be appropriate. Hagen et al.[108] have suggested a value in the neighborhood of 10^6 s^{-1} for the fastest forming loops in proteins based on measurements of cytochrome *c* and a theoretical model for the formation of a contact by collapse of a random chain. This value seems somewhat too small, not for the process specifically considered but for use generally in the protein folding reaction. Some proteins fold on the microsecond time scale and interestingly an Arc repressor mutant has a bimolecular rate constant for folding of 3×10^8 M^{-1} s^{-1}.[109] There is also the question of whether the diffusive rate slows down significantly as the transition state for the folding reaction is approached (see Socci et al.[96] for a diffusive, Kramers-like,[82] protein folding model). It is not clear whether this is a major factor since both the effective diffusion constant for the entities involved and the range over which diffusion is required are expected to decrease simultaneously.

V.B. Temperature Dependence of Folding Rate. One of the striking differences between small molecule kinetics and protein folding is the temperature dependence of the rate constant. Although the unfolding rate is often approximately Arrhenius-like (i.e., a plot of ln k vs $1/T$ is essentially a straight line),[110] the folding rate is not. Parts a and b of Figure 9 show the results obtained, respectively, for folding with the 27-mer cube lattice model that we have already discussed[94,95,98] and for folding of CI2 from the experiments of Oliveberg et al.[91] The interpretation of the lattice model results is straightforward and provides an increased understanding of the folding dynamics. At low temperatures ($1/T$ large in Figure 9a), there is Arrhenius-like behavior associated with a positive activation energy (or enthalpy), as in eqs 3 and 4. This means the barrier to folding is dominated by the apparent activation energy in correspondence with the effective energy surface in Figure 8a. Actually, the barrier in Figure 8a is smaller than that estimated from Figure 9a, suggesting that activated diffusion may also contribute to the barrier. As the temperature increases ($1/T$ small in Figure 9a), the Arrhenius plot turns over and the activation energy becomes negative. In the present model, this arises from temperature-independent interactions and an effective (funnel-like) energy surface in which there is a monotonic decrease in the energy. The corresponding surface is shown in Figure 8c, which yields a negative energy of activation, as confirmed by analyzing the results in Figure 9a with eq 4b. However, the free energy of activation, $\Delta G^{\ddagger}$, is still positive due to the activation entropy ($-T\Delta S^{\ddagger}$), whose effect increases with increasing temperature. The activation entropy is associated with the decrease in the number of configurations accessible with increasing Q; i.e., there is an entropic "bottleneck" associated with the transition state and this bottleneck becomes more important at higher temperatures because of the larger number of states accessible in the denatured region. At very low temperatures, there is no indication in the lattice calculations (Figure 9a and Gutin et al.[106]) or in the available experiments (Figure 9b), that there is a deviation from the Arrhenius law; that is, no "super-Arrhenius" behavior, analogous to that described by eq 15, for example, has been observed in protein folding.

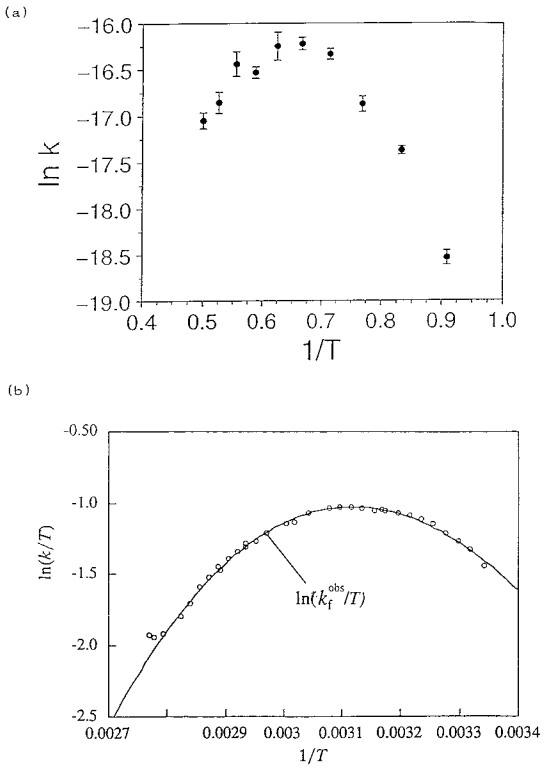

Figure 9. (a) Calculated Arrhenius plot of the folding reaction rate constant as a function of temperature for the sequence used in Figure 7; results adapted from Karplus et al.[98] (b) Arrhenius temperature dependence of the measured refolding rate constant for CI2, adapted from Oliveberg et al.[91]

The experimental Arrhenius plot for the folding of CI2, shown in Figure 9b, is very similar in form to that found in the lattice simulations (Figure 9a). Consequently, the formal interpretation in terms of activation enthalpy, $\Delta H^{\ddagger}$, and activation entropy $\Delta S^{\ddagger}$, is exactly the same. As the temperature increases, the activation enthalpy becomes negative, while the activation free energy remains positive. One explanation of these results is that the effective enthalpy surface of protein CI2 (and other proteins) corresponds, in fact, to those shown in parts a and c of Figure 8 at low and high temperatures, respectively, and that the positive activation free energy at high temperatures arises from the configurational entropy term, as in the lattice models. If $T = 1.6$, the maximum in the rate of the lattice simulations (Figure 9a) is set equal to 325 K, a factor of 10 decrease in the rate at higher temperatures corresponds to about 400 K, somewhat

larger than the experimental value of 360 K for CI2; this is in accord with the fact that the entropic contribution in CI2, which has 64 residues, is expected to be larger than for the 27-mer. The effective energy as a function of the number of native contacts obtained in unfolding simulations of CI2 at high temperature[100] does show a monotomically decreasing energy with increasing Q, in accord with this model. This is a very interesting result[100] (Lazaridis and Karplus, to be published) because it suggests that the experimental temperature dependence of the protein folding reaction is in accord with funnel-like effective energy surfaces at high temperatures but not at room temperature. Since for many proteins[89,91,110] the temperature dependence of the folding rate is similar in form to that of CI2, these results suggest that, under physiological conditions, the effective energy surface is not funnel-like; that is, it is more like that shown in Figure 8a than Figure 8c.

It has been suggested[89,104,110] that the temperature dependence of the interactions themselves, due to the hydrophobic effect, is involved in the non-Arrhenius behavior. The hydrophobic effect increases with increasing temperature over the range of interest (275−330 K), as based on small molecule studies and their application to interpret protein thermodynamics.[111] The dominant contribution of the hydrophobic effect is on the denatured state, which is enthalpically less stabilized as the temperature increases. If the transition state is less unfolded than in the denatured state (for an analysis of CI2, see ref 100), it would have a proportionally smaller contribution from the hydrophobic effect and the magnitude of $\Delta H^{\ddagger}$ would be expected to decrease with increasing temperature. This contribution to the curvature in the Arrhenius plots contrasts with the lattice results in which the interactions are independent of temperature and the change in activation enthalpy is due to the difference in the energies of the configurations sampled as a function of temperature. Both types of contributions may play a role in protein folding.

VI. Summarizing Discussion

This review provides a brief background in chemical kinetics to serve as a basis for comparing the "simple" behavior of small molecule reactions with the more complex reactions of proteins. Simple behavior was defined by two (related) aspects of reactions. The first is that a simple phenomenological rate law with an exponential time dependence for the rate applies and the second is that the temperature dependence of the rate follows the Arrhenius equation. We have seen that although simple behavior is found in some protein reactions, significant deviations from both types of simplicity have been documented and interpreted theoretically. Ligand rebinding after photolysis in myoglobin was discussed in considerable detail because this reaction has been studied more than any other and has been shown to deviate from both requirements for simplicity under certain conditions.

For protein folding, the situation is somewhat different. Although deviations from simplicity might have been expected, there is considerable evidence that a simple rate law with an exponential overall folding rate applies to most proteins that have been studied so far; there are exceptions (intermediates), but they are "trivial" from the present viewpoint. By contrast, the second criterion for simplicity is not satisfied. The temperature dependence of the folding rate of all proteins that have been studied show significant deviations from Arrhenius behavior. We have given a determination of the results based on lattice simulations, but the isolation of the actual mechanism for the non-Arrhenius type behavior in proteins will require additional studies.

From the history of the experimental and theoretical investigations of ligand rebinding after photolysis in myoglobin, it seems likely that the applicability of a simple phenomenological rate law to the overall protein folding reaction hides an underlying complexity. One set of experiments that will help to reveal this complexity involves studies with a short time resolution for systems under nonequilibrium conditions. A range of methods for triggering the folding (and unfolding) reaction have been developed, and these processes can be studied, in principle, on time scales from nanoseconds to seconds or longer.[93,112] However, techniques for monitoring the structural details as a function of time still need to be improved. Techniques that are likely to make important contributions are time-resolved X-ray crystallography,[113] real time NMR,[93] as well as infrared spectroscopy[114] and vibrational Raman optical activity.[115] More generally for protein dynamics, infrared photon echo measurements,[116,117] and optical line broadening studies[118,119] are expected to yield significant new results.

Another type of approach that is likely to be of great importance for many reactions involving proteins, including enzyme catalysis, ligand binding, and protein folding, is the monitoring of single molecules.[32,120−122] This provides a direct way of determining the diversity that exists for the various reactions and will make it possible to distinguish, in principle, whether the macroscopic complexity is evident in individual molecules or whether it is a consequence of averaging over a population of molecules. It is interesting to mention in this regard that single molecule behavior has been studied for a long time for the case of membrane channel conductance, in which the molecules involved undergo a significant structural change between open and closed states. For most channel systems, the structural data required for atom-based simulations are not available. However, the high-resolution structure of a potassium channel determined very recently[123] opens the way for detailed theoretical studies on this very interesting system; many simulations will certainly be made. Finally, the more global variation of experimental conditions (e.g., folding at low temperatures) will aid in determining the difference in the temperature (or other) dependence of different relaxation processes.

The future holds great promise for this exciting field. It is likely that, as measurements and simulations become more precise and cover a wider range of conditions, complex behavior of the type discussed in this manuscript will be found for an increasing number of reactions in proteins.

Note Added in Proof. As stated in the Summarizing Discussion, it was to be expected that the "underlying complexity" of the protein folding reaction would be revealed by "studies with a short time resolution under nonequilibrium conditions". In a recent study of the folding of phosphoglycerate kinase and an ubiquitin mutant, Sabelko et al.[128] have observed nonexponential folding induced by a temperature jump. The origin of the nonexponential behavior and its relation to that discussed in section IV is not clear. Further experiments will be needed to determine the validity of the interesting interpretation proposed by Sabelko et al.

Acknowledgment. I thank William Eaton and Hans Frauenfelder for helpful comments and Aaron Dinner and Andrej Šali for preparing Figure 1. I thank Stefan Fischer for calling my attention to reference 70. Partial support for the research reported here was provided by the National Science Foundation and the National Institutes of Health. A preliminary version of the material reported here is being published in the "Dahlem Workshop Report on Simplicity and Complexity in Protein and

Nucleic Acids," H. Frauenfelder, J. Deisenhofer, and P. G. Wolynes, Editors (Freie Universität Berlin, 1999).

References and Notes

(1) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355−5373.

(2) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318−321.

(3) Chandler, D. *J. Chem. Phys*. **1978**, *68*, 2959−2970.

(4) Weston, R. E.; Schwarz, H. A. *Chemical Kinetics*; Prentice Hall: New York, 1972.

(5) Karplus, M.; Porter, R. N.; Sharma, R. D. *J. Chem. Phys.* **1965**, *43*, 3259−3287.

(6) Bai, Y. S.; Fayer, M. D. *Phys. Rev.* **1989**, *B39*, 11066−11084.

(7) Greer, J.-L.; Fan, J.; Angell, C. A. *J. Phys. Chem.* **1994**, *98*, 13780−13790.

(8) Angell, C. A. *Science* **1995**, *267*, 1924−1927.

(9) Fersht, A. *A Guide to Catalysis and Protein Folding*; W. H. Freeman & Co.: New York, 1999.

(10) Keck, J. C. *Adv. Chem. Phys.* **1967**, *13*, 85−121.

(11) Anderson, J. B. *J. Chem. Phys.* **1973**, *58*, 4684−4692.

(12) Wigner, E. P. *Trans. Faraday Soc.* **1938**, *34*, 29−41.

(13) Grote, R. F.; Hynes, J. T. *J. Chem. Phys*. **1980**, *73*, 2715−2732.

(14) Montgomery, J. A.; Chandler, D.; Berne, B. J. *J. Chem. Phys.* **1979**, *70*, 4056−4066.

(15) Northrup, S. H.; Pear, M. R.; Lee, C.-Y.; McCammon, J. A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4035−4039.

(16) Skinner, J. L.; Wolynes, P. G. *J. Chem. Phys.* **1980**, *72*, 4913−4927.

(17) Knowles, J. R. *Philos. Trans. R. Soc. London B* **1991**, *332*, 115−121.

(18) Richard, J. P. *J. Am. Chem. Soc*. **1984**, *106*, 4926−4936.

(19) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826−5832.

(20) Åqvist, J.; Fothergill, M. *Biol. Chem*. **1996**, *271*, 10010−10016.

(21) Knowles, J. R. *Nature* **1991**, *350*, 121−124.

(22) Joseph-McCarthy, D.; Petsko, G. A.; Karplus, M. *Protein Eng.* **1995**, *8*, 1103−1115.

(23) Neria, E.; Karplus, M. *Chem. Phys. Lett.* **1997**, *267*, 23−30.

(24) Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1978**, *110*, 5297−5311.

(25) Warshel, A.; Sussman, F.; Hwang, J. K. *J. Mol. Biol.* **1988**, *201*, 139−159.

(26) Brumer, P.; Karplus, M. *Faraday Soc. Discuss.* **1973**, *55*, 80−91.

(27) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523−2544.

(28) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.

(29) Valleau, J. P.; Torrie, G. M. In *Statistical Mechanics*; Berne, B. J., Ed.; Plenum Press: New York, 1977; Part A, p 169−194.

(30) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J.-C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(31) Bergsma, J. P.; Gartner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1356−1376.

(32) Craig, D. B.; Arriaga, E. A.; Wong, J. C. Y.; Lu, H.; Douichi, N. J. *J. Am. Chem. Soc.* **1996**, *118*, 5245−5283.

(33) Henry, E. R.; Sommer, J. H.; Hofrichter, J.; Eaton, W. A. *J. Mol. Biol.* **1983**, *166*, 443−451.

(34) Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E. T.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5000−5004.

(35) Perutz, M.; Matthews, B. T. *J. Mol. Biol.* **1965**, *21*, 199−202.

(36) Takano, T. *J. Mol. Biol.* **1977**, *110*, 537−568.

(37) Case, D. A.; Karplus, M. *J. Mol. Biol.* **1979**, *132*, 343−368.

(38) Frauenfelder, H.; Petsko, G. A.; Tsernoglu, D. *Nature* **1979**, *280*, 558−563.

(39) Debrunner, P. G.; Frauenfelder, H. *Annu. Rev. Phys. Chem.* **1982**, *33*, 283−299.

(40) Petrich, J. W.; Lambry, J.-C.; Kuczera, K.; Karplus, M.; Poyart, C.; Martin, J.-L. *Biochemistry* **1991**, *30*, 3975−3987.

(41) Kuczera, K.; Kuriyan, J.; Karplus, M. *J. Mol. Biol.* **1990**, *213*, 351−373.

(42) Doster, W.; Cusack, S.; Petry, W. *Nature* **1989**, *337*, 754−756.

(43) Smith, J.; Kuczera, K.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1601−1605.

(44) Iben, I.; Braunstein, D.; Doster, W.; Frauenfelder, H.; Hong, M. K.; Johnson, J. B.; Luck, S.; Ormos, P.; Schulte, A.; Steinbach, P. J.; Xie, A. H.; Young, R. D. *Phys. Rev. Lett.* **1989**, *62*, 1916−1919.

(45) Steinbach, P.; Amsari, A.; Berenndzen, J.; Braunstein, D.; Chu, K.; Cowen, B.; Ehrenstein, D.; Frauenfelder, H.; Johnson, J. B.; Lamb, D.; Luck, S.; Mourant, J.; Nienhaus, G.; Ormos, P.; Philipp, R.; Xie, A.; Young, R. D. *Biochemistry* **1991**, *30*, 3988−4001.

(46) Agmon, M.; Hopfield, J. J. *J. Chem. Phys.* **1983**, *79*, 2042−2053.

(47) Levy, R. M.; Sheridan, R. P.; Keepers, J. W.; Dubey, G. S.; Swaminathan, S.; Karplus, M. *Biophys. J.* **1985**, *48*, 509−518.

(48) Lim, M.; Jackson, T. A.; Anfinrud, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5801−5804.

(49) Jackson, T. A.; Lim, M.; Anfinrud, P. A. *Chem. Phys.* **1994**, *180*, 131−140.

(50) Kuczera, K.; Lambry, J.-C.; Martin, J.-L.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5805−5807.

(51) Li, H.; Elber, R.; Straub, J. E. *J. Biol. Chem.* **1993**, *268*, 17908−17916.

(52) Schaad, O.; Zhan, H. X.; Szabo, A.; Eaton, W. A.; Handy, E. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9547−9551.

(53) Hagen, S. J.; Eaton, W. A. *J. Chem. Phys.* **1996**, *104*, 3395−3398.

(54) Kopfer, G.; Hilhorst, H. *Europhys. Lett.* **1987**, *3*, 1213−1217.

(55) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902−6915.

(56) Saven, J. G.; Wang, J.; Wolynes, P. G. *J. Chem. Phys.* **1994**, *101*, 11037−11043.

(57) Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495−1517.

(58) Czerminski, R.; Elber, R. *Proc. Nat. Acad. Sci.* **1989**, *86*, 6963−6967.

(59) Czerminsky, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580−5601.

(60) Wang, J.; Wolynes, P. G. *Phys. Rev. Lett.* **1994**, *74*, 4317−4320.

(61) Joseph, D.; Petsko, G. A.; Karplus, M. *Science* **1990**, *249*, 1425−1428.

(62) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11905−11910.

(63) McCammon, J. A.; Gelin, B. R.; Karplus, M.; Wolynes, P. G. *Nature* **1976**, *262*, 325−326.

(64) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8502−8507.

(65) Perutz, M. *Nature* **1971**, *232*, 408−413.

(66) Snyder, G. H.; Rowan, R.; Karplus, S.; Sykes, B. D. *Biochemistry* **1975**, *14*, 3765.

(67) Wagner, G.; DeMarco, A.; Wüthrich, K. *Biophys. Struct. Mech.* **1976**, *2*, 139−158.

(68) McCammon, J. A.; Karplus, M. *Biopolymers* **1980**, *19*, 1375−1405.

(69) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, & Thermodynamics*; Adv. Chem. Phys. LXXI; John Wiley & Sons: New York, 1988.

(70) Otting, G.; Liepinsh, E.; Wüthrich, K. *Biochemistry* **1993**, *32*, 3571−3582.

(71) Gerstein, M.; Lesk, A. M.; Chothia, C. *Biochemistry* **1994**, *33*, 6739−6749.

(72) Gerstein, M.; Chothia, C. *J. Mol. Biol.* **1991**, *220*, 133−149.

(73) Wade, R. C.; Davis, M. E.; Luty, B. A.; Madura, J. D.; McCammon, J. A. *Biophys. Soc. 64*, 9−15.

(74) McCammon, J. A.; Northrup, S. H.; Karplus, M.; Levy, R. M. *Biopolymers* **1980**, *19*, 2033−2045.

(75) Derreumaux, P.; Schlick, T. *Biophys. J.* **1998**, *74*, 72−81.

(76) Williams, J. C.; McDermott, A. E. *Biochemistry* **1995**, *34*, 8309−8319.

(77) Yüksel, K. Ü.; Sun, A.-Q.; Gracy, R. W.; Schnackerz, K. D. *J. Biol. Chem.* **1994**, *269*, 5005−5008.

(78) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88−118.

(79) Hofrichter, J.; Henry, E. R.; Szabo, A.; Murray, L. P.; Ansari, A.; Jones, C. M.; Coletta, M.; Falcioni, G.; Brunori, M.; Eaton, W. A. *Biochemistry* **1991**, *30*, 6583−6598.

(80) Henry, E. R.; Jones, C. M.; Hofrichter, J.; Eaton, W. A. *Biochemistry* **1997**, *36*, 6511−6528.

(81) Baldwin, J.; Chothia, C. *J. Mol. Biol.* **1979**, *129*, 175−200.

(82) Kramers, H. A. *Physica* **1940**, *7*, 284−293.

(83) McCammon, J. A.; Karplus, M. *Nature* **1977**, *268*, 765−766.

(84) Karplus, M.; Shakhnovich, E. Protein Folding: Theoretical Studies of Thermodynamics and Dynamics. In *Protein Folding*; Creighton, T., Ed.; W. H. Freeman & Sons: New York, 1992; pp 127−195.

(85) Levinthal, C. How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems*, Proceedings of a Meeting held at Allerton House, Monticello, IL; Debrunner, P., Tsibris, J. C. M., Münck, E., Eds.; University of Illinois Press: Urbana, 1969; p 22.

(86) Karplus, M. *Folding Des.* **1997**, *2*, 569−576.

(87) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428−10435.

(88) Schindler, T.; Schmid, F. X. *Biochemistry* **1996**, *35*, 16833−16842.

(89) Scalley, M. L.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10636−10640.

(90) Fersht, A. R.; Itzhaki, L. S.; ElMasry, N. F.; Matthews, J. M.; Otzen, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10426−10429.

(91) Oliveberg, M.; Tan, Y.-J.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8926−8929.

(92) Becker, O. M. *J. Mol. Struct*. (*THEOCHEM*) **1997**, *398*−*399*, 507−516.

(93) Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 869−893.
(94) Šali, A.; Shakhnovich, E.; Karplus, M. *J. Mol. Biol.* **1994**, *235*, 1614−1636.
(95) Šali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248−251.
(96) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860−5868.
(97) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1994**, *100*, 9238−9257.
(98) Karplus, M.; Caflisch, A.; Šali, A.; Shakhnovich, E. In *Modelling of Biomolecular Structures and Mechanisms*; Pullman, A., et al., Eds.; Kluwer Academic Publishers: Norwell, MA, 1995; pp 69−84.
(99) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Chem. Phys.* **1994**, *101*, 6052−6062.
(100) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928−1931.
(101) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Folding Des.* **1997**, *2*, 109−114.
(102) Dinner, A.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 7976−7994.
(103) Du, R.; Pande, V. J.; Grasberg, A. Y.; Tanaka, T.; Shakhnovich, E. *J. Chem. Phys.* **1998**, *108*, 334−350.
(104) Chan, H. S.; Dill, K. A. *Proteins* **1998**, *30*, 2−33.
(105) Zwanzig, R. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 148−150.
(106) Gutin, A.; Šali, A.; Abkevich, V.; Karplus, M.; Shakhnovich, E. *J. Chem. Phys.* **1998**, *108*, 6466−6483.
(107) Karplus, M.; Weaver, D. L. *Nature* **1976**, *260*, 404−406.
(108) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615−11617.
(109) Waldburger, C. D.; Jonsson, T.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2629−2634.
(110) Segawa, J. J.; Sugihara, M. *Biopolymers* **1984**, *23*, 2473−2488.
(111) Makhatadze, G. I., Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307−425.
(112) Eaton, W. A.; Thomson, D. A.; Chan, C.-K.; Hagen, S. J.; Hofrichter, J. *Structure* **1996**, *4*, 113−119.
(113) Srajer, V.; Teng, T.; Ursby, T.; Pradervand, C.; Ren, Z.; Adachi, S.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; Moffat, K. *Science* **1996**, *274*, 1726−1729.
(114) Phillips, C. M.; Mizutami, Y.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7292−7296.
(115) Wilson, G.; Hecht, L.; Barron, L. D. *J. Mol. Biol.* **1996**, *261*, 341−347.
(116) Tokmakoff, A.; Fayer, M. D. *Acc. Chem. Res.* **1995**, *28*, 437−445.
(117) Rector, K. D.; Rella, C. W.; Hill, J. R.; Kwok, A. S.; Sligar, S. G.; Chien, E. Y. T.; Dlott, D. D.; Fayer, M. D. *J. Phys. Chem.* **1997**, *101*, 1468−1475.
(118) Thorn Leeson, D.; Wiersma, D. A. *Nature Struct. Biol.* **1995**, *2*, 848−851.
(119) Thorn Leeson, D.; Wiersma, D. A.; Fritch, K.; Friedrich, J. *J. Phys. Chem. B* **1997**, *101*, 6331−6340.
(120) Askin, A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4853−4860.
(121) Lu, H. P.; Xie, S. *Nature* **1997**, *385*, 143−146.
(122) Zocchi, G. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10647−10651. Weiss, S. *Science* **1999**, *283*, 1676−1683.
(123) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science* **1998**, *280*, 69−77.
(124) Zwanzig, R. *Acc. Chem. Res.* **1990**, *23*, 148−152.
(125) Shlesinger, M. F.; Zaslavsky, G. M.; Klafter, J. *Nature* **1993**, *363*, 31−37.
(126) Wang, Z.; Pearlstein, J. M.; Jia, Y.; Fleming, G. R.; Norris, J. R. *Chem. Phys.* **1993**, *176*, 421−425.
(127) Gehlen, J. N.; Marchi, M.; Chandler, D. *Science* **1994**, *263*, 499−502.
(128) Sabelko, J.; Ervin, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6031−6036.