The Nature of Folded States of Globular Proteins

J. D. HONEYCUTT[1] and D. THIRUMALAI[2]

[1]Biosym Technologies, Inc., 10065 Barnes Canyon Road, San Diego, California 92121; and [2]Department of Chemistry and Biochemistry, and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742

SYNOPSIS

We suggest, using dynamical simulations of a simple heteropolymer modelling the α-carbon sequence in a protein, that generically the folded states of globular proteins correspond to statistically well-defined metastable states. This hypothesis, called the metastability hypothesis, states that there are several free energy minima separated by barriers of various heights such that the folded conformations of a polypeptide chain in each of the minima have similar structural characteristics but have different energies from one another. The calculated structural characteristics, such as bond angle and dihedral angle distribution functions, are assumed to arise from only those configurations belonging to a given minimum. The validity of this hypothesis is illustrated by simulations of a continuum model of a heteropolymer whose low temperature state is a well-defined β-barrel structure. The simulations were done using a molecular dynamics algorithm (referred to as the "noisy" molecular dynamics method) containing both friction and noise terms. It is shown that for this model there are several distinct metastable minima in which the structural features are similar.

Several new methods of analyzing fluctuations in structures belonging to two distinct minima are introduced. The most notable one is a dynamic measure of compactness that can in principle provide the time required for maximal compactness to be achieved. The analysis shows that for a given metastable state in which the protein has a well-defined folded structure the transition to a state of higher compactness occurs very slowly, lending credence to the notion that the system encounters a late barrier in the process of folding to the most compact structure. The examination of the fluctuations in the structures near the unfolding $\rightarrow$ folding transition temperature indicates that the transition state for the unfolding to folding process occurs closer to the folded state.

INTRODUCTION

It is well known that in order for a protein molecule to carry out a specific biological function, it has to adopt a well-defined three-dimensional structure.[1,2] The mechanism of formation of the three-dimensional structure of a protein from a given sequence of amino acids remains one of the most important unsolved problems in molecular biology, and is sometimes referred to as the "second half" of the genetic code.[3] In the process of synthesis of a protein using the genetic machinery, it is believed that only a linear sequence of connected amino acids emerges. The formation of a well-defined three-dimensional structure, i.e., the cotranslational events that lead to the formation of a biologically active globular protein, is known as the "protein folding problem."[4] It has been shown that in some cases the folding process in the living cell may involve catalysts and other special enzymes.[5,6] Two specific examples are the formation of a triple helix in collagen and the folding of insulin. However, moderately sized proteins can spontaneously form folded structures in vitro,[7] and hence the understanding of the folding process in these systems can in principle be obtained using the laws of physical chemistry (Refs. 7–12; see also, for example, Ref. 13). The studies of Anfinsen and co-workers laid the foundation for the thermodynamic hypothesis of protein folding. Although

Biopolymers, Vol. 32, 695–709 (1992)
 CCC 0006-3525/92/060695-15$04.00

the absence of chaperon molecules and other cellular components makes in vitro studies of the folding process a better defined problem than the in vivo process of protein folding, the major difficulty in understanding the protein folding problem arises because of the diverse number of conformational states that are likely to occur. The characterization of these states and their associated stability makes this a formidable problem.

Although the thermodynamic hypothesis was accepted as the basis of interpreting the folding of globular proteins, its validity was seriously questioned by Levinthal.[14] The Levinthal argument became the basis of the kinetic control hypothesis for protein folding.[14,15] This hypothesis suggests that energy barriers or bottlenecks in phase space hinder the protein from forming the state with the lowest free energy, thus trapping it in a metastable state. The kinetic control hypothesis has led to several mechanistic proposals for the formation of the folded state of globular proteins.[16,17]

In this article we elaborate on a hypothesis suggested recently by us,[18] which is a generalization of the kinetic control process, and is referred to as the metastability hypothesis. This hypothesis appears to be in accord with a few experiments. However, we wish to caution the reader that the conclusions we have reached are based solely on computer simulations on a highly simplified model of a protein molecule. There is strong evidence that these simplified models are very useful even in portraying the characteristics seen in real protein molecules.[19-23] Nevertheless, the true test of the conclusions reached here should come from in vitro experiments and ultimately from in vivo studies.

The rest of this paper is organized as follows. In the next section we present and further clarify the metastability hypothesis. The third section is devoted to the detailed methodology used to obtain the results presented in the sections following it. The experimental consequences of our hypothesis are surveyed in the fourth section, and the penultimate section is concerned with the possible experimental evidence supporting our theoretical findings. This paper is concluded in the sixth section with additional remarks.

METASTABILITY HYPOTHESIS

The metastability hypothesis,[18] is based on the notion that the folding of a polypeptide chain is kinetically controlled. Our hypothesis, which can be thought of as a generalization of the kinetic control hypothesis, can be stated as follows: Several forms of folded conformations of a polypeptide chain exist, all of which exhibit very similar structural characteristics but differ in energies from one another. It is posited that almost all of these folded states are metastable, and the particular state into which the protein folds depends on the initial conditions. The content of this hypothesis is illustrated in Figure 1, in which a schematic sketch of the rough free energy landscape for a polypeptide chain is presented. Notice that for the purpose of illustrating the metastability hypothesis, roughness that typically is exhibited in the free energy landscape on small length scales has been suppressed in Figure 1. According to our hypothesis the various (folded) minima, labeled by numbers in Figure 1, differ in energy and are separated by barriers that are hard to overcome in typical time scales involved in the folding process. This time is usually on the order of a few seconds for the biological synthesis of a protein molecule.[2] The metastability hypothesis states that the structural characteristics measured (or calculated) correspond to statistical properties obtained by restricting the configurations to a given minimum. Furthermore, these structural properties are expected to be fairly similar among the different minima.

The second important aspect of the metastability hypothesis is the role of the initial folding conditions in directing the denatured chain into one of the minima corresponding to the folded states. These

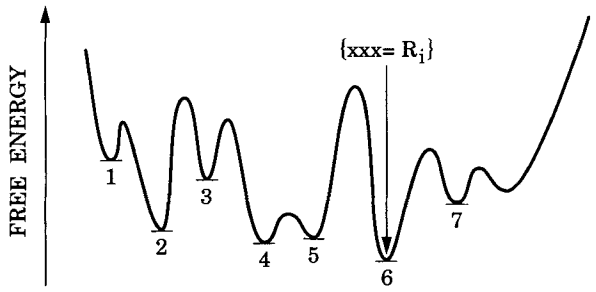

Figure 1. Schematic sketch of the free energy (or energy) profile corresponding to the folded region in the conformational space of a protein molecule. The sketch is used to illustrate the content of the metastability hypothesis. The various minima are separated by bottlenecks of differing amplitude such that the transition from one to another is unlikely to occur on the time scale of the folding process. The typical time scale for the folding process ranges from milliseconds to seconds. The particular minimum into which the protein falls depends on the initial condition. For example, in this figure the initial configurations of the protein marked by { xxx } map onto the minimum labeled 6.

initial configurations may be determined by the thermodynamic conditions, pH, and ionic strength, a group of amino acid residues directing the folding process, as well as the presence of other coenzymes or chaperon molecules if they are needed. It follows that the set of configurations that map onto a given minimum (for example, the configurations given by xxx map onto the folded minimum labeled 6 in Figure 1) is far less than the total set of allowed configurations. Thus the Levinthal argument is circumvented in our scenario.

A rough free energy landscape with certain common characteristics is thought to exist in a variety of physical systems (for a recent review, see Ref. 24; see also Refs. 25 and 26).* Most notably, in the context of protein folding, Wolynes and co-workers have suggested that concepts from spin glass theory may be used to understand certain global aspects of protein folding.[28-30] In various models of spin glasses the existence of many metastable states that can only interconvert on long time scales emerges from very simple Hamiltonians in which some sort of random interactions are assumed to exist.[26] Recently, several experimental studies have suggested that the dynamical aspects observed in proteins have much in common with what is seen in glassy systems.[28,30] We should emphasize, however, that the demonstration of the validity of the metastability hypothesis is a little more exacting. In order to validate our hypothesis we have to show that these multiple metastable folded states should emerge from Hamiltonians that are potentially capable of mimicking interactions in real proteins (without having to introduce randomness arbitrarily). Furthermore, the structural characteristics in each of the low-lying minima have to be similar. The notion of "glassy" behavior in our model emerges from the model naturally, and lends support to the conjecture that concepts developed to understand frustrated systems such as spin glasses and structural glasses may be useful in understanding certain kinetic aspects of protein folding.[31-33]

METHODOLOGY

The approach we have taken is to construct a model for the polypeptide chain using simple representations of the interaction potentials between the various residues. Then simulation techniques are used to investigate the dynamic nature of the unfolding to folding transition. These methods are based on a combination of molecular dynamics and quench techniques. In this section, a detailed presentation of the numerical methodology used in this paper is provided.

Construction of the Model

In order to illustrate the metastability hypothesis suggested in the preceding section, we have constructed a set of "minimal" models, whose low temperature states are of the β-barrel form. By "minimal" we mean the simplest model capable of capturing the essential features of the physical system. These minimal models contain some, but not all, of the features that play a role in imparting stability to real globular proteins. For example, our model for the heteropolymer does not contain side groups, which are known to be responsible for intramolecular hydrogen bonding.[1] The role of the solvent (usually water) is treated in a very simplified fashion.† The model is aimed at testing the metastability hypothesis, and therefore only certain salient features of a real polypeptide chain are retained in our simulations. Despite these limitations, it has become clear that models of the sort studied here and elsewhere[18-23] provide insight into crucial forces needed for the folding of a polypeptide chain. Furthermore, it is hoped that a detailed study of simplified models may prove useful in elucidating the pathways involved in the transition from the denatured state to the folded state. If these lessons are useful in interpreting in vitro experiments, then one can directly study the effects of mutations on protein stability using these simple dynamical models.[35]

The most significant criterion we have adopted in constructing a model Hamiltonian for the model polypeptide chain is that each individual term in the potential function should in principle be computable either using ab initio quantum mechanical methods or be calculable as a potential of mean force using statistical mechanics. We feel the main physical properties observed in proteins should be a consequence of such first-principle "minimal" (Ref. 36; for a review, see Ref. 37) models. Most notably, any "cooperativity" found in the protein folding process must result as a specific consequence of a given thermodynamic condition and the interplay among

* In the context of structural glasses, various ideas regarding the free energy landscape have been formulated. See, for example, Ref. 27.

† It has been recently argued by Levitt and Sharon[34] that molecular dynamics simulations that explicitly include the solvent molecules provide a more realistic description of dynamics in proteins.

the various (calculable) forces involved. In our models we do not include cooperativity effects in an ad hoc fashion. Since our objective is to study in detail the dynamics of folding of a model heteropolymer and to map out the various free energy minima, no attempt has been made to make the potential parameters compatible with those found in real proteins. The proposed model is partly inspired by a series of recent papers by Skolnick et al.[19-21]. These authors were attempting to obtain a model of a heteropolymer on a diamond lattice, whose low temperature phase is a unique β-barrel structure. One of our goals has been to study the kinetics of the folding (and unfolding) process. Accordingly, we have devised a continuum model satisfying the major criteria outlined above and used the noisy molecular dynamics technique to follow the kinetics of the folding–unfolding transition.

The studies by Skolnick et al. have already illustrated the basic ingredients needed to obtain a β-barrel structure.[19-21] The potential energy functions we have developed, which are very different from that used by Skolnick et al.[19-21] in their on-lattice simulation studies, also yield the β-barrel structure as the minimum energy (not necessarily unique) conformer. Our model incorporates the following types of forces: (a) Long-range forces such as the hydrophobic interaction (for a review of the importance of hydrophobic effects in protein folding, see Ref. 38). Here "long range" refers to the interactions between hydrophobic residues that may be separated by long distance in sequence space as measured by the length along the backbone of the chains. The actual range of the potential in coordinate space is relatively short and is typically of the order of a few molecular diameters. We have chosen a simple functional form to mimic these forces. (b) Simple bond angle potential between two successive bonds. (c) Dihedral angle potential due to the rotation about the peptide bond. The bonds are constrained to be of fixed length. Our model protein contains three different types of "residues" in the backbone of the chains, namely hydrophobic (B), hydrophilic (L), and neutral (N). The meaning of these terms in terms of the potential function is described shortly. For the moment, it is only necessary to note that B residues attract other B residues. The specific sequence we have used for our 46-member chains is the same as that given by Skolnick et al., namely

$$B_9N_3(LB)_4N_3B_9N_3(LB)_5L \qquad (1a)$$

In addition, we have also considered the dynamics of folding of a chain with $n = 58$ whose sequence is given by

$$B_9N_3(LB)_4N_3B_9N_3(LB)_5N_3B_{10} \qquad (1b)$$

In Eqs. (1), LB indicates that an L-type bead is bonded to a B-type bead. We have done most of our simulation work on the $n = 46$ model and have studied the $n = 58$ case less extensively. In both cases we have established that the low temperature form indeed has a well-defined β-barrel structure. The time required for the atoms in the folded β-barrel states to achieve registry is quite long. It is for this reason that the mapping of the various metastable minima for the $n = 58$ has not been thoroughly done. However, the robustness of our model in yielding β-barrels for a larger value of n has been established. The majority of the results presented in the fourth section are for the $n = 46$ case. The symmetry of the above sequence has been constructed so that if one has a folded structure, a three- or fourfold bend would emerge as long as n is greater than a certain minimum value. From simple energy considerations the minimum value of n turns out to be approximately 30. The neutral residues appear in regions where bends are desired for the native low temperature structure. These residues interact with each other only through a short-range repulsion. In real globular proteins these segments can be thought to be responsible for loops and turns. The dihedral angle forces are assumed to be weaker for the bonds involving the neutral residues so that the bend formation is enhanced. This constraint proves to be extremely important and is further discussed in the final section. The Hamiltonian describing the model heteropolymer is given by the sum of the three forces. The functional form for each is described below.

Long-Range Potential. The hydrophobic attraction, which occurs only between residues both of which are of type B, has the form

$$V_{BB}(r) = 4\epsilon_h[(\sigma/r)^{12} - (\sigma/r)^6] \qquad (2)$$

where r is the distance between the specified residues. For simplicity, σ is chosen equal to the bond length between two successive residues along the backbone and ϵ_h defines the energy scale. All energies are presented in terms of ϵ_h, and all lengths are given terms of σ. Thus $\epsilon_h = \sigma = 1$ in reduced units. The interactions between L-L pairs or L-B pairs are governed by the following potential

$$V_{L\alpha}(r) = 4\epsilon_L[(\sigma/r)^{12} + (\sigma/r)^6] \quad \alpha = \text{L or B} \quad (3)$$

where $\epsilon_L = \frac{2}{3}\epsilon_h$. Note that the above form is purely repulsive. The repulsive potential has a range that is somewhat longer than a pure soft sphere r^{-12} repulsion. The interaction between the N residues and L, B, or N residues takes the form

$$V_{N\alpha}(r) = 4\epsilon_h(\sigma/r)^{12} \quad \alpha = \text{N, L, or B} \quad (4)$$

Bond Angle Potential. The bond angles formed between three successive residues $(k, k+1, k+2)$ are constrained by a simple harmonic potential,

$$V(\theta) = \frac{k_\theta}{2}(\theta - \theta_0)^2 \quad (5)$$

where $k_\theta = 20\epsilon_h/(\text{rad})^2$ and $\theta_0 = 105°$ (= 1.8326 radians). The bond angle forces do not appear to play a major role in the folding and unfolding of the protein. Notice that the spring constant is relatively high and therefore the deviation of the bond angles from θ_0 is not expected to be large. In view of this, it would have been reasonable to have simply enforced the constraint that all bond angles are equal to θ_0. However, it has been shown that for molecular dynamics that enforce bond angle constraints it is numerically more efficient to use a potential term to describe the constraint.[39] The spring constant k_θ used in Eq. (5) adequately satisfies this criterion.

Dihedral Angle Potential. The general form of the potential for the dihedral angle φ describing the rotation of three successive bonds involving four connected beads is well known. The potential is conveniently represented as

$$V(\varphi) = A(1 + \cos\varphi) + B(1 + \cos(3\varphi)) \quad (13)$$

The above potential has three minima, one at $\varphi = 0$ corresponding to the *trans* configuration, and two slightly higher minima at $\varphi = \pm\cos^{-1}[(3B - A/12B)^{1/2}]$ corresponding to the two *gauche* states. For all the residues except those in the bend regions, $A = B = 1.2\epsilon_h$. For this choice of parameters the difference between the energies of the *trans* and *gauche* states is $1.75\epsilon_h$. For the residues in the bend region we choose $A = 0$ and $B = 0.2\epsilon_h$. The residues involved in the weaker dihedral angle potential (i.e., $A = 0$, $B = 0.2\epsilon_h$) are identified in Figure 2. It is clear that the bonds in the bends must adopt the

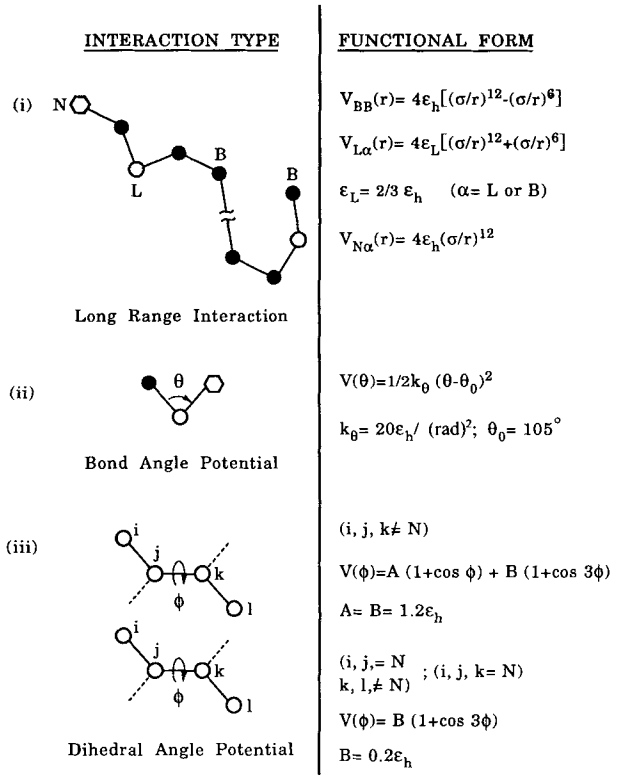

Figure 2. A sketch of the various interactions that are allowed in our model. The associated functional form of the potential and the parameters are also provided.

gauche (or approximately *gauche*) conformations. Our choice of the parameter values in the bend regions (mostly involving the neutral residues) achieves this objective by having two major effects. First, it makes all the three minima shallower than when $A \neq 0$ and B is larger. This enhances the kinetics of bend formation by reducing the barrier between the various states. In addition setting $A = 0$ makes the *trans* and *gauche* states of equal depth, thereby enhancing the thermodynamics of bend formation. Since the major objective of our investigation is to study the nature of various low temperature minima, we have used the simplest physically plausible potentials to ensure that the chain folds into well-defined β-barrel structures. In Figure 2 we provide a summary of the allowed interactions in our model as well as the parameters of the potentials.

The lowest energy structures that emerge using the model described above have the desired shape of a β-barrel. A proper balance between the short-range repulsive forces and the relatively long-range (hydrophobic) attractive forces is necessary to ini-

tiate folding into a desired conformer. This implies that a balance between the dihedral angle potential, which tends to stretch out the molecule into a state in which all bonds have the *trans* configuration, and the attractive hydrophobic potential is crucial to induce folding into a β-barrel like structure upon cooling to low temperatures. If the attraction is too weak, the molecule remains in a more or less elongated state with the bonds in nearly all *trans* configuration, even at low temperatures. If the attractive forces are too strong, we find that the chains fold into a dense globule-like structure (for a description of the relationship between the dihedral angle potential and the collapsed states, see Ref. 40).

It is interesting to contrast our model to that used by Skolnick et al. in their Monte Carlo simulation studies of the equilibrium folding to unique β-barrel structures.[19-23] The energy of a given configuration of the heteropolymer chain was taken to be a sum of three terms: (a) For a given sequence of n beads occupying the various sites of the diamond lattice, each of the $n-3$ interior bonds was allowed to have one of three rotational states. This term is a discrete version of our continuous dihedral angle potential with the three conformations corresponding to the *trans* and two *gauche* states, respectively. (b) Also included were interactions between pairs of nonbonded nearest neighbor residues to mimic the hydrophobic and hydrophilic interactions. This roughly corresponds to the long-range forces used in our model. (c) A crucial element in the model of Skolnick et al. is the "cooperativity" term, which allows for conformational coupling between nonbonded but near conformational states. They found in one particular example that folding of the polypeptide chain does take place even in the absence of this term, but most of their study corresponds to the nonzero value of the cooperativity term. As we stated earlier, the origin of this term is hard to justify. In addition to the fact that our model is a continuum model, which has very interesting physical consequences (such as the emergence of metastable minima) that are discussed below, the absence of a complicated cooperativity term clearly distinguishes our model from the interesting but distinct one considered by Skolnick et al.[19-21]

Simulation Methods

It is well known that one encounters the ubiquitous problem of multiple minima in the simulation of the folding of a protein from the denatured state to the folded state (for a review of the multiple minima problem in the context of protein folding, see Ref. 41). Several different algorithms, most of which are based on the Monte Carlo technique, have been used to overcome this problem.[42,43] One of the aspects of interest to us is the dynamics of the folding (and unfolding) process. Consequently, the simulation technique we have used is a modified molecular dynamics method based on the velocity form of the Verlet algorithm.[44] We have also used a damping term with an appropriately chosen friction coefficient. To compensate for energy loss due to friction, the beads in the backbone is allowed to experience a Gaussian random force. The resulting algorithm is generally referred to as "noisy molecular dynamics" or a type of "stochastic dynamics." There are several reasons for the use of the friction in the molecular dynamics algorithm: (a) In the conventional constant temperature MD method the rescaling (or reassignment) of the velocities greatly perturbs the trajectory. While this is not a serious problem when the free energy profile does not exhibit a multivalley structure, this is not the case in the simulation of the folding process. (b) The atoms in a protein are embedded in a solvent, and consequently the appropriate reduced dynamics is stochastic rather than ballistic. (c) Most importantly, the use of the friction and stochastic terms enhances the rate of exploration of configuration space. This aspect is not as relevant when the polypeptide chain is in the denatured state, but once the protein is in any one of the many metastable conformations the relaxation within that valley is expected to be enhanced with the algorithm we have used. The main purpose of the random force is to control the temperature of the system in the most physically reasonable way.

The algorithm used in our simulations is derived for the low friction case. Recall that there are constraints in the model of the polypeptide chain used in our simulation. For example, the bond length between the primary C_α skeleton is fixed. This constraint is incorporated into the equations of motion using the RATTLE algorithm.[45]

For simplicity, we describe the algorithm for one cartesian component of the position and velocity of a single residue. The mass m of the particle and the Boltzmann constant k_B are both set to unity. For a single particle the equation of motion is

$$\ddot{r} = F_c + \Gamma - \zeta\dot{r} \equiv F \tag{7}$$

where F_c is the configurational force, given by the negative gradient of the potential energy with respect to the position of the particle, Γ is the random force with white noise spectrum, and ζ is the friction coefficient. The above equation is numerically in-

tegrated using the velocity form of the Verlet algorithm.[43] In this algorithm with time step h, the position and velocity at time $t + h$ given those at time t are expressed through second order in h as

$$r(t + h) = r(t) + h\dot{r}(t) + \tfrac{1}{2}h^2F(t) \quad (8)$$

and

$$\dot{r}(t + h) = \dot{r}(t) + \frac{h}{2}[F(t) + F(t + h)] \quad (9)$$

For the standard energy conserving molecular dynamics, $F(t + h)$ is a function of the positions alone and is independent of the velocities at time $t + h$. Thus, once $r(t + h)$ is calculated all the quantities on the right-hand side of Eq. (9) for $\dot{r}(t + h)$ are known. However, when friction is included, i.e., $\zeta \neq 0$, $F(t + h)$ depends on $\dot{r}(t + h)$, and the equation for $\dot{r}(t + h)$ must be solved self-consistently. The result through second order in h is

$$\dot{r}(t + h) = \dot{r}(t)\left[1 - \frac{h\zeta}{2} + \frac{1}{4}(h\zeta)^2\right] + \tfrac{1}{2}h\,[F_c(t) + F_c(t + h) + \Gamma(t + h)] \quad (10)$$

Equation (10) is valid only in the limit $h\zeta$ is small.

As usual, the strength of the random force Γ is related to ζ by the fluctuation-dissipation theorem

$$\langle \Gamma(t)\Gamma(t') \rangle = 2\zeta T\delta(t - t') \quad (11)$$

where T is the temperature of the medium or the bath. In the simulation one uses the discrete version of the equation of motion, and thus a coarse-grained random force, as opposed to the impulsive force given by the above expression, must be used. The coarse grained random force is defined as

$$\bar{\Gamma} = \frac{1}{h}\int_{t-1/2h}^{t+1/2h} \Gamma(s)ds \quad (12)$$

It is easy to show that the statistics of $\bar{\Gamma}(t)$ are given by

$$\langle \bar{\Gamma}_t \rangle = 0 \quad (13)$$

$$\langle \bar{\Gamma}_t\bar{\Gamma}_{t'} \rangle = \frac{2\zeta T}{h}\delta_{tt'} \quad (14)$$

where $\delta_{tt'}$ is the Kronecker delta. This results in the distribution for the random force at a given time step having the Gaussian form

$$P(\Gamma) = (h/4\Gamma\Pi T\zeta)^{1/2}e^{-h\Gamma^2/4T\zeta} \quad (15)$$

The value of the random force at a given step for a specified value of h and T is sampled from the above distribution using the standard Box–Müller algorithm. The noisy molecular dynamics method offers dynamical information, thus allowing us to monitor the actual kinetics of the folding process. Most importantly, the possibility of exploring long-lived metastable minima, which is at the heart of this study, is hard to achieve using the Monte Carlo method on lattice models with discrete states. Thus we believe that the relaxation behavior of the heteropolymer obtained by the "noisy" molecular dynamics should mimic the kinetics of in vitro folding.

RESULTS AND DISCUSSION

High Temperature Structures

The molecular dynamics algorithm described in the previous section was used to equilibrate a polypeptide chain at an elevated temperature. Analysis of the structure shows that the chain is a disordered random coil and remains in an unfolded state. At these temperatures the entropy gain in the unfolded state is greater than the energy gain due to favorable contacts between the hydrophobic residues in the folded state. A snapshot of the chain at high temperature is given in Fig. 1 of our previous paper.[18] We find that there is no element of the secondary structure present and in fact the chain undergoes large-scale conformational fluctuations inconsistent with any long-lived secondary structure. One of the advantages of molecular dynamics simulation over real experiments is that one can arrange the thermodynamic conditions such that the chain configurations are truly random. This problem is often delicate in various in vitro experiments in which denaturation is often achieved by using an appropriate reducing agent. Some of the denaturation procedures can lead to aggregation of protein molecules (see, for example, Ref. 46).

Low Temperature Structures

In order to follow the dynamics of the folding process, we generated the low temperature structures by three different quench techniques. In our simulations, the control parameter that determines the state of the folded chain is the temperature. The transition from the unfolded to folded state was induced by changing the temperature. The resultant low temperature structures obtained by the different quench techniques are discussed below.

Quench from a High Temperature. In this process the high temperature denatured state is directly quenched to $T = 0$ K instantaneously. After the sudden quench, in which the kinetic energy is removed, the energy of the heteropolymer chain is minimized so that a local energy minimum is obtained. The structures obtained by the direct quench usually had no resemblance to the β-barrel structures. In analogy with similar quench experiments on liquids, these low temperature states can be characterized as defective "glassy" proteins.[18,31,32]

Structures from Slow Cooling. In order to obtain ordered low temperature structures and to follow the kinetics of the folding process, we also performed several simulations for various initial conditions by "slowly" (on a simulation time scale) lowering the temperature starting from a typical denatured state. If the metastability hypothesis is valid, one would expect to obtain several different minima in which the structures would resemble a β-barrel with the three bends in the desired locations. It should be emphasized that cooling rates obtained in computer simulations are much higher than those found in typical temperature jump experiments. It is also clear that the precise structures that are formed will depend on the rate of cooling, and consequently the structures obtained by fast quench and adiabatic cooling will be different. The objective of this exercise to ensure that upon slow cooling the heteropolymer chain does indeed fold into a β-barrel structure.

The process of slow cooling was achieved by linearly decreasing the bath temperature with time, starting from a value well above the approximate temperature T_F (see following section on the unfolding–folding transition), at which the chain undergoes a transition from the unfolded state to a folded state. The temperature was decreased from the high temperature value (typically 1.2–2.0 in reduced units) to zero in a time period ranging from 500 to 2500 τ. The time dependence of temperature is given by $T(t) = T_0 - Rt$, where R is the cooling rate. In this process we have obtained structures ranging from the desired β-barrel to other "misfolded" structures.[‡] A typical snapshot of a β-barrel structure obtained by slow cooling is given in Figure 3. We analyzed the nature of the β-barrel structures formed using a crude distance matrix analysis, the elements of which are distances between a pair of residues.[49] The folded β-barrel structures had elements of both parallel and antiparallel strands. This is found to be the case for the sequence with $n = 58$ as well. Such structures do occur in real proteins in which the direction of adjoining strands is determined by the degree of overall hydrophobicity of each of the strands.[50,51] An example of a "misfolded" structure is one that resembles a β-barrel with one of its four strands unfolded. It should be noted that a few of the non-β-barrel structures obtained using the slow cooling method have energies less than those of some of the higher energy β-barrel structures. The precise structure and the associated energy seem to depend on initial conditions. In particular, we have found that identical cooling schedules from different initial conditions can result in distinct low temperature structures.

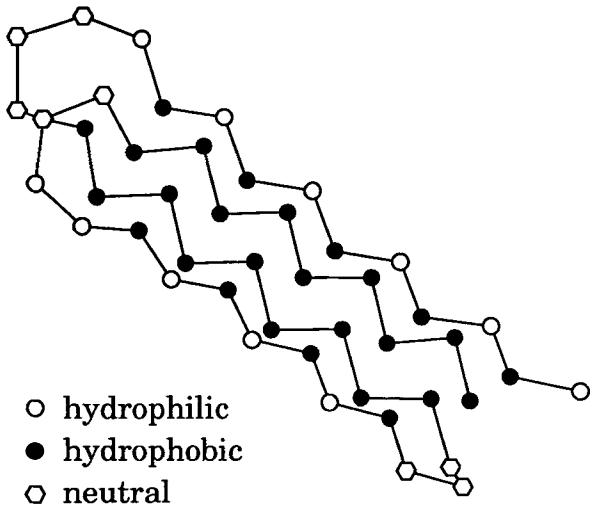

Figure 3. Typical β-barrel structure of the chain below the unfolding–folding transition temperature. The configurations were obtained by a process of slow cooling. The nature of the various residues for this chain consisting of 46 beads is shown by the various symbols.

Structures Obtained by Simulated Annealing. It is clear from the structures obtained by the slow cooling method that the chain, for the set of parameters chosen, could adopt a well-defined β-barrel structure depending on the initial conditions. In order to probe other possible minima in which the heteropolymer chain folds into a β-barrel-like structure, we used a combination of simulated annealing and steepest descent quench methods[50–53] to enhance the sampling of conformation space. A trial β-barrel structure was constructed "manually" such that all bond

[‡] These misfolded structures are to be contrasted with the incorrectly folded structures that were examined in the context of a hypothesis concerning the structure of proteins having sequence homology by co-workers.[47,48] In these studies the side chains of two protein molecules were interchanged and the energy was minimized. They found that the two incorrectly folded proteins have more nonpolar groups exposed to the solvent. The misfolded states in our simulations typically have folds in the "wrong" place and in some instances the hydrophobic residues were found on the outside.

angles away from the bend regions were set at the equilibrium minimum values and all the bonds away from the bends were made to be in the *trans* configuration. For convenience, each of the hairpin turns was made planar, resulting in initially strained bond angles in the bend region. The energy of this structure was minimized by performing "dynamics" with the velocity of all the residues being set to zero after each time step. This is equivalent to the steepest descent quench because the system traverses the path in the free energy landscape to minimize the gradient.[50,51] The structure that resulted from this had the desired folded conformation, which we denote as $\beta 1$. To test the metastability hypothesis further, we searched for other β-barrel-like structures with energies less than that of the $\beta 1$ structure. Starting from the $\beta 1$ structure, the temperature of the system was raised, keeping it well below the unfolding transition temperature as determined from the slow cooling runs. The system was then repeatedly annealed for lengths of time which varied from 100τ to 300τ. Subsequently, the system was quenched to a local minimum, and the structural and energetic aspects of the minima were determined. This method of searching for the local minima is similar to the simulated annealing Monte Carlo technique.[54] This procedure was repeated several times. The initial structure for a given run was taken to be the previous minimum conformation explored by the chain. In this process we obtained eight β-barrel structures. The various structural characteristics, namely the gross appearance and the radius of gyration, of these eight minima were almost identical. However, they differed in their energies, implying that almost all these are *truly distinct* metastable minima. It should be emphasized that we did not by any means exhaust all the minima that can be obtained using this procedure. However, the number of minima corresponding to the folded structure is not expected to scale exponentially with the number of residues. We also wish to stress that we have by no means shown that the simulated β-barrel structure with the lowest energy ($-49.57\ \epsilon_h$) corresponds to a global minimum. The main point of our hypothesis is sufficiently illustrated by the two methods of locating several metastable folded structures.

The discovery that there could exist a collection of minima in inhomogeneous systems is not new. They have been found experimentally[30] as well as in molecular dynamics simulation studies probing the conformational substates in myoglobin.[55,56] In the latter studies[55,56] the conformational substates were found to differ in relative orientation of the helices coupled with side-chain rotations in such a manner that preserved the close packing of the protein interior. Our simulation studies show that many metastable minima, which are not merely conformational substates, exist when our model chain folds into ordered structures. In fact, we argue (in the fifth section) that these minima correspond to tier O of the classification of states suggested by Frauenfelder et al.[30] The important aspect for protein folding is that the ordered structures of the backbone of the chain (secondary structure) in all the minima may be indistinguishable. Thus, insofar as the enzymatic activity of a protein is essentially determined by the conformation, any one of the folded structure might be efficient. The catalytic efficiency may depend on the precise structure in a given minimum.[30]

The energetics of the various structures obtained by the different techniques described above can be further characterized by the "inherent structure" method.[50-53] In this method, the system is described by the structure and energy it attains after the thermal energy is removed by a steepest descent quench. The energies of the various structures obtained by the different quench techniques described above form a "spectrum" that can be plotted and analyzed.[57] A summary of these computations is shown in Figure 4 in which we present the energy spectrum of the various inherent structures obtained in our simulations. The three columns represent the quenched β-barrel structures, the quenched structures that are obtained from slow cooling, and the minima that result from a direct quench from the denatured state. It is clear that the polypeptide chain can adopt conformations that do not have the desired folded structure but have lower energies than

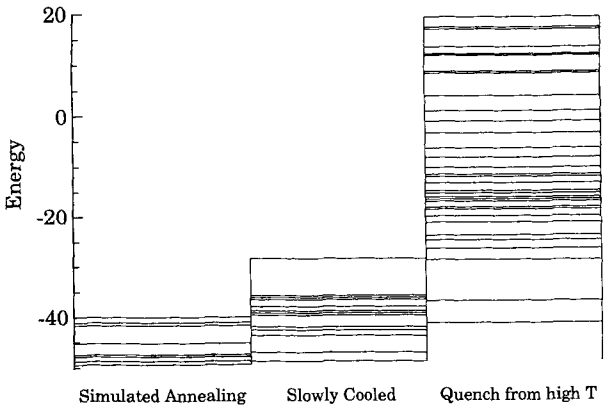

Figure 4. Spectrum of energies of the various quench structures obtained using several simulation techniques. The explanation of the three columns is given in the text. T is the temperature.

certain higher energy β-barrel structures. These are defective folded structures that can presumably acquire the desired β-barrel conformation if annealed for a long time. It should be emphasized that in no instance was a non-β-barrel structure found that was lower in energy than the lowest energy β-barrel structure. However, some of the energies in the second column correspond to structures indistinguishable from β-barrels, demonstrating that our model does indeed fold into an ordered structure under adiabatic cooling conditions. If the initial conditions, which are determined by the environment, are such that the conformation of the chain maps into one of the "defective" minima, one would not obtain the required folded structure. The time for escape out of such a free energy valley would depend on the barrier height separating the minima and the folded state, and could greatly exceed the biological time for the folding process. It is tempting to speculate that the "wrong" choice of initial condition in a real protein could lead to the formation of defective folded states—an example being the formation of kinks along the backbone. It is also interesting to compute the rms fluctuations in the energy of the low energy β-barrel structures.[§] This was computed using the energies of the β-barrels obtained by simulated annealing. We find that $\sigma = \sqrt{\langle \epsilon^2 \rangle - \langle \epsilon \rangle^2}$ for these low energy structures is found to be 3.3 ϵ_h.

Unfolding–Folding Transition

The transition from the unfolded state of a protein to the folded state is often described as a phase transition in a finite system. In the description of the collapse transition in synthetic systems, the "order parameter" that typifies this transition is usually the mean square end-to-end distance or the radius of gyration (see, for example, Ref. 58). The radius of gyration $\langle R_g^2 \rangle$ exhibits different behavior depending on the temperature or, equivalently, the nature of the solvent. Following the usual characterization of synthetic polymers, we have used the temperature dependence of $\langle R_g^2 \rangle$ as well as the average energy of the peptide chain to monitor the thermodynamics of the folding transition. In Figure 5, a plot of the radius of gyration of the chain

$$\langle R_g^2 \rangle = \frac{1}{2n^2} \sum_i \sum_j \langle r_{ij}^2 \rangle \qquad (16)$$

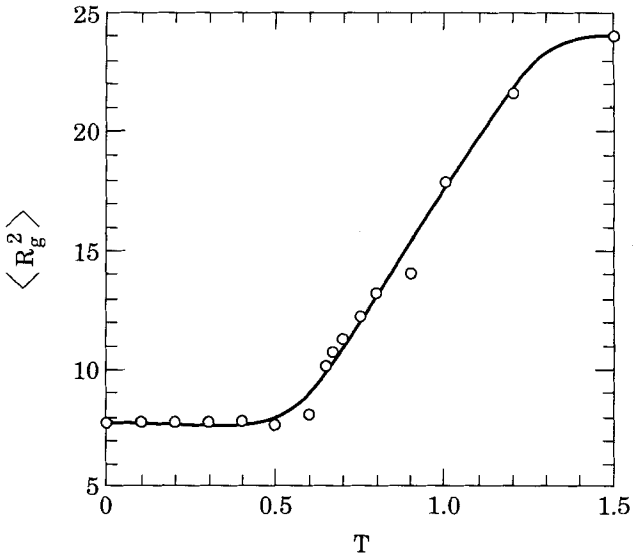

Figure 5. A plot of the radius of gyration $\langle R_g^2 \rangle$ as a function of temperature for the $n = 46$ case. The solid line is drawn by smoothly connecting all the points. The transition from an extended to a compact structure is at $T_F \sim 0.65$.

as a function of temperature is exhibited. It is clear that at high temperatures the chain is in an unfolded state, yielding a large value for $\langle R_g^2 \rangle$. For temperatures less than about 0.65 ϵ_h/k_B the chain size is relatively small, corresponding to a more compact structure. From Figure 5 the transition from an extended (denatured) state to a compact (folded) state takes place around $T_F \simeq 0.65\ \epsilon_h/k_B$. A further confirmation of this transition temperature can be obtained from a plot of the internal energy E (shown in Figure 6) and specific heat C (shown in Figure 7) as a function of temperature. It is clear from Figure 6 that energy is a continuous function of temperature, implying that the folding transition in this model is an "effective" higher order transition that is characteristic of small systems. The specific heat plot shown in Figure 7 is not meant to be quantitative. The transition temperature at $T \simeq T_F$ should be taken as a transition midpoint because since we have a finite system the transition should occur over a range of temperatures. Figure 7 is meant to give a rough idea that even in this finite system one does have a transition from an extended state to a more compact state. The specific heat clearly has a rounded peak at $T = T_F \simeq 0.65\ \epsilon_h/k_B$, which represents a transition from an unfolded state to one of several folded structures. For $T > T_F$ the chain is disordered.

[§] We are grateful to Prof. M. Karplus for suggesting that we compute the rms fluctuations in the low energy β-barrel structures.

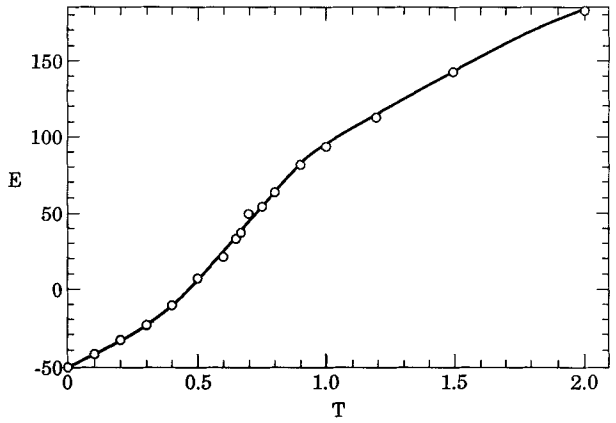

Figure 6. Energy E as a function of temperature for the chain with $n = 46$. This function is a continuous function of temperature. As a guide to the eye the data points have been connected by a smooth line. The continuity of the energy curve implies that the transition from an unfolded structure to a folded one corresponds to an effective higher order transition characteristic of small systems.

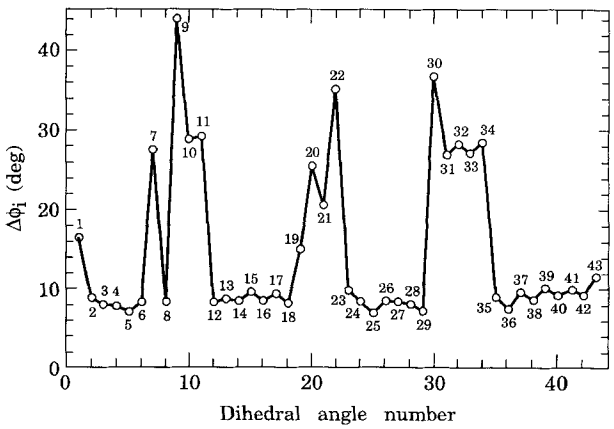

Figure 8. Decomposition of the fluctuation in the total dihedral angle into the contribution from various dihedral angle numbers. The temperature is fixed at $T = 0.4\epsilon_h/k_B$, which is below the transition temperature T_F. The equation is $\Delta\phi_i(T) = \langle(\varphi_i - \varphi_i^0)^2\rangle^{1/2}$, where i is the dihedral angle number. The fluctuations are measured with respect to the ideal zero temperature β-barrel structure.

Fluctuations in the β-Barrel Minima

In this section we introduce various measures to probe the relative fluctuations in various β-barrel structures. The various measures we have utilized to differentiate between two distinct structures are insensitive to certain basic symmetry operations such as considering a structure and its mirror image. Since the Hamiltonian is symmetric under the operation $r_i \rightarrow -r_i$ for all i's we expect that the energies of a given structure and the mirror image of that structure to be identical. To analyze the nature of the fluctuations in the various low energy minimum structures, we have calculated the mean deviations in dihedral angle from the "idealized" β-barrel structure. This structure corresponds to the β-barrel structure with the lowest energy (cf. Figure 3). It would be more instructive to use the structure corresponding to the global minimum (for methods of finding global minima specifically for proteins, see Refs. 42 and 59), if this were known, as the appropriate reference structure. Starting from this initial structure at $T = 0$, we slowly heated the system, allowed it to equilibrate, and computed the quantity

$$\Delta\varphi(T) = \langle 1/N_c \sum_c (\varphi_c(T) - \varphi_c^0)^2\rangle^{1/2} \quad (17)$$

where φ_c^0 is the value of the dihedral angle in the ideal zero temperature β-barrel structure and $\varphi_c(T)$ is the corresponding value at the temperature T. The summation is over the $N_c(=(n-3))$ dihedral angles in the system and the angle brackets denote a statistical mechanical average. A plot of $\Delta\varphi(T)$ as a function of temperature displayed elsewhere[18] shows that $\Delta\varphi$ can be as large as 25° even below T_F. Above T_F these fluctuations are even larger, and indicate that compared to the more compact folded form at $T = 0$, the high temperature structure is random and open.

It is possible to decompose $\Delta\varphi(T)$ into contributions from the various dihedral angles. For example, one can write

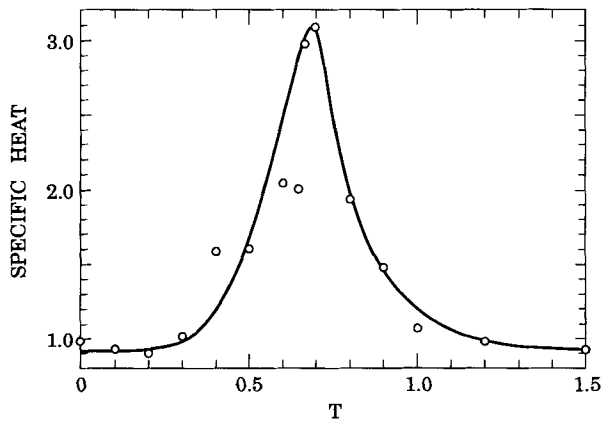

Figure 7. A rough estimate of the specific heat $C = dE/dT$ as a function of temperature. It has been obtained by numerically differentiating the energy vs T curve. The solid line is drawn by connecting the points. The peak at $T_F = 0.65\ \epsilon_h/k_B$ indicates the transition temperature.

$$\Delta\varphi(T) = \frac{1}{N_c}\sum_i \Delta\varphi_i(T)$$

$$= \frac{1}{N_c}\sum_i \langle(\varphi_i - \varphi_i^\circ)^2\rangle^{1/2} \quad (18)$$

where the summation denotes the contribution from the dihedral angle i, which involves the residues $i - 1$, i, $i + 1$, and $i + 2$. The value of the dihedral angle found in the idealized β-barrel structure at $T = 0$ is denoted by φ_i°. The results of the computation of $\Delta\varphi_i(T)$ as a function of the dihedral angle number at $T = 0.4$ is exhibited in Figure 8. The largest fluctuations occur for residues involved in the dihedral angle numbers (7, 9, 10, 11, 20, 21, 22, 30, 31, 32, 33, 34) involving residues that make up the loop. On the other hand, the dihedral angles involving residues in the interior of the heteropolymer, which is largely made up of hydrophobic species, show much smaller fluctuations.

Another measure of fluctuations in various minima is the relative compactness in the different minima. If one assumes that the lowest energy "idealized" β-barrel structure represented in Figure 4 is maximally compact (which is often taken to be the case with native forms of proteins), then it is desirable to predict the dynamics of approach to the compact structure. For example, if the protein is trapped in one of the several metastable minima the time scale needed for achieving maximal compactness is an interesting issue. We have used the dynamic generalization of the quantity measuring static fluctuations introduced earlier,[18] namely

$$\Delta_\alpha(t) = \left\{\frac{1}{n-1}\sum_{i=1}^{n}[(X_i^\alpha(t))^2 - (X_i^\circ(t))^2]^2\right\}^{1/4} \quad (19)$$

as a possible way of assessing the dynamics of approach of the coordinates in the αth structure to an idealized reference structure. In Eq. (19), $X_i^\circ(t)$ is the ith residue in the αth β-barrel structure and $X_i^\circ(t)$ is the corresponding value in the idealized β-barrel structure. Notice if $\alpha \neq 0$ and if there is no transition from the minimum labeled α to the idealized structure during the observational time, then the long time limit of $\Delta_\alpha(t)$ is expected to be nonzero. In Figure 9 a plot of $\Delta_\alpha(t)$ as a function of time for three temperatures is given. It is clear that $\Delta_\alpha(t)$ for $T = 0.4$ has not decayed to zero on the time scale of our simulation. Since the long time limit of $\Delta_\alpha(t)$ is nonzero, this implies that α and 0 are truly distinct minima. Even at $T = 1.5\ \epsilon_h/k_B$ and $T = 2.0\ \epsilon_h/k_B$,

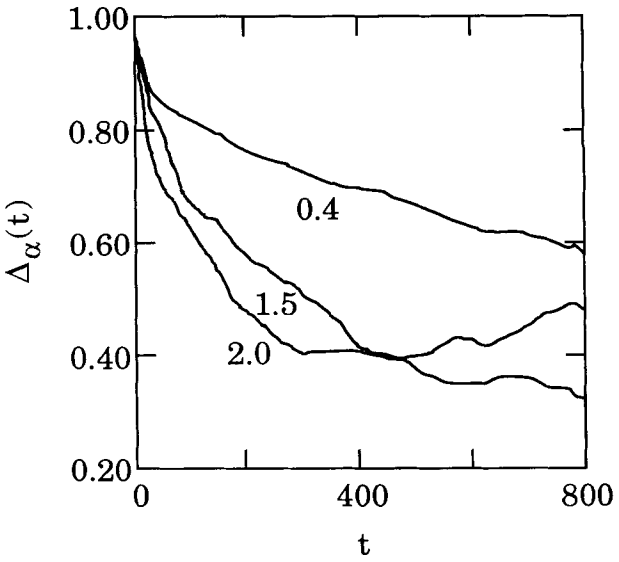

Figure 9. The measure of compactness $\Delta_\alpha(t)$ given by Eq. (20) as a function of t at different temperatures. The numbers adjacent to each curve indicate the temperature in reduced units ϵ_h/k_B. Two lowest energy β-barrel structures generated by the simulated annealing method (see Figure 5) were used to calculate $\Delta_\alpha(t)$. The approximate scaling $\Delta_\alpha \sim 1/k_c t$, where k_c, is the rate at which maximal compactness is obtained, is evident. The rate constant k_c is dependent on temperature and decreases as the temperature is lowered.

the transition from one minimum to another is quite slow. The approximate time scale for achieving compactness comparable to that in the idealized structure appears to be very slow. After an initially fast decay in $t \simeq (50\text{–}75)\tau$ (which corresponds to a rapid collapse to the structure similar to that found in the idealized β-barrel structure) the system apparently encounters a barrier that has to be overcome to achieve a greater compactness. Based on an analogy with our previous findings in structural glasses,[60,61] we suggest that the decay of the measure for achieving maximal compactness should be described by the relation

$$\Delta_\alpha(t) \sim \tau_c/t. \quad (20)$$

for t greater than a transient time ($\sim 75\tau$). In Eq. (21), τ_c is the rate at which maximal compactness is achieved.

Experimental Evidence

There are some experimental results that are consistent with our findings. (a) The earliest indication

of the existence of metastable states was found in the protein ovalbumin. In a series of experiments, Smith and Back found that when ovalbumin was warmed under alkaline conditions a conformational change to a more stable form, identified as S-Ovalbumin, occurred.[62–65] This more stable form exhibited very similar structural characteristics to those of ovalbumin. The percentage of S-Ovalbumin depended largely on the method of preparation; the initial conditions determine the kinetics of formation of the more stable form. In the language of the present paper, we conclude that ovalbumin is in a distinct metastable state that undergoes a transformation to another minimum (S-ovalbumin) with very little structural change. The rate of binding of the two minima to metal anion is found to vary in accord with our assertions.[66] (b) In a series of experiments, Frauenfelder et al. have been studying the kinetics of binding of O_2 and CO to myoglobin (Mb).[30] The results have been convincingly interpreted in terms of conformational substates. According to this picture, the conformational states of Mb and MbCO are arranged in a heirarchical fashion. The states are arranged into various tiers, with states (three in number) in tier 0 having the largest barrier between them. The states in 0 furcate into another manifold of states and so on. These authors also assert that the structural features of the three states in tier 0 are similar and they bind to CO with different rates. In terms of the hypothesis in our paper, the three substates found in tier 0 in MbCO correspond to the three distinct metastable minima. Since detailed kinetic studies of folding of proteins are just becoming available, the generality of these results have not yet been established.

SUMMARY AND CONCLUSIONS

In this paper we have discussed in detail the possibility that globular protein molecules in nature are generically in one of the many possible metastable states. We have shown using computer simulations on a highly simplified representation of a protein that one can indeed have many minima with differing energies in which the protein acquires very similar structural characteristics. It would be desirable to obtain further experimental data on hen egg ovalbumin that clearly seems to be an ideal candidate for testing in detail certain of the ideas presented here.

Many experiments, as well as some simulation studies, have described the protein folding process as being highly cooperative and hence describable in terms of an all-or-none picture. Our studies suggest that a hierarchical structure with intermediates for the folding process is also a likely possibility (for a further discussion of this point, see Ref. 67). Both of these pictures have received experimental support. In our studies we have shown that the polypeptide chain undergoes large structural fluctuations when $T \approx T_F$. These structural fluctuations, which can persist for a sufficiently long time, can be thought of as intermediates. However, these fluctuating structures are highly unstable, and hence experimental characterization of these intermediates may be difficult. If the lifetimes of these fluctuating structures are too small to be experimentally determined, then the folding of a protein will appear to be an all-or-none process. Since the concentration of intermediates is almost always small, it is likely that experiments that probe the dynamics of the folding process can often be explained in terms of an all-or-none process. It is very difficult to obtain evidence for long lived metastable states in discrete lattice Monte Carlo simulations[#]. If the picture we have described here is indeed correct, then both cooperativity (leading to an all-or-none picture) and hierarchical organization can occur in the folding process of a complex enough protein. In our simulation studies we have seen that below T_F the chain spontaneously folds into a β-barrel structure even though there is no evidence of any structural remnant at $T \approx T_F$. It therefore follows that the rate-determining process for the formation of a folded structure occurs late in the folding process. The conformation of the chain at the transition state would have much of the structure found in the folded state of the protein. These observations are in rough accord with the detailed kinetic studies of the folding of bovine pancreatic trypsin inhibitor.

The scenario of a late transition state in the folding process and the formation of fluctuating structures at $T \approx T_F$ suggests the following dynamics. For time scales less than that defined by the rate-determining step, the kinetics can be nonexponential, which is a reflection of the several fluctuating intermediate structures. At longer times, the kinetics should be exponential. If the lifetime of the intermediates is too small to be detectable, then the kinetics will almost always appear to be exponential. This appears to be the case in certain experimental studies where the folding process is essentially all or none.

[#] By construction in discrete state lattice Monte Carlo simulations, there is little possibility of observing any long-lived metastable intermediates. It is for this reason Skolnick and coworkers (see Refs. 19–23) always observe an all-or-none process in their simulation studies.

In this article, we have also introduced a measure to describe the rate at which a folded structure achieves maximum compactness. Chan and Dill have argued that the origin of internal architecture in proteins is a natural consequence of compactness in these molecules.[68] Our analysis here suggests that after a fast initial decay the rate of approach to a maximally compact structure can be extremely slow. This again suggests that a simple all-or-none picture may not always be adequate. The time dependence of the measure of compactness suggested here is in principle testable by monitoring the time evolution of radius of gyration, which can be obtained from the dynamic structure factor of the protein.

We are indebted to several colleagues for useful discussions. In particular, we wish to thank Robert Zwanzig for encouragement and for several incisive remarks during the initial stages of this project. We owe a special debt of gratitude to Professor W. Kauzmann, who pointed out the references on hen egg ovalbumin. In addition, we wish to acknowledge useful correspondence with Professor Kauzmann and with Professor J. Knowles. We are also grateful to Jonathan Widom for a discussion about the experimental ways of measuring the time required for maximum compactness to be reached in proteins. This work was supported in part by a grant from the National Science Foundation under grant number NSF-CHE-86-57356. Additional support from the Henry and Camille Dreyfus Foundation is gratefully acknowledged. The computations were made possible by a generous grant of supercomputer time by the San Diego Supercomputer Center.

REFERENCES

1. Creighton, T. E. (1984) in *Proteins: Structure and Molecular Properties,* W. H. Freeman, San Francisco.
2. Stryer, L. (1988) in *Biochemistry,* 3rd ed. W. H. Freeman.
3. Gierasch, L. M. & King, J., Eds. (1990) *Protein Folding: Deciphering the Second Half of the Genetic Code,* American Association for the Advancement of Science, Washington, DC.
4. Ghelis, C. & Yan, J. (1982) *Protein Folding,* Academic Press, New York.
5. Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, O. J., Georgopoulos, C. P., Hendrix, R. W. & Ellen, R. J. (1988) *Nature* **333,** 330.
6. Godobinoff, P. G., Gatenby, A. A. & Lorimer, G. H. (1989) *Nature* **337,** 44.
7. Epstein, C. J., Goldberger, R. F. & Anfinsen, C. B. (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28,** 439.
8. Anfinsen, C. B. (1973) *Science* **181,** 223.
9. Anfinsen, C. B. & Scheraga, H. A. (1975) *Adv. Protein Chem.* **29,** 205.
10. Wolynes, P. G. (1991) *Cargese Lectures,* to be published.
11. Paine, G. H. & Scheraga, H. A. (1985) *Biopolymers* **24,** 1391.
12. Paine, G. H. & Scheraga, H. A. (1986) *Biopolymers* **25,** 1547.
13. Levitt, M. (1980) in *Protein Folding,* Jaenicke, R. Ed., North Holland, Amsterdam.
14. Levinthal, C. (1968) *J. Chim. Phys.* **65,** 44.
15. Wetlaufer, D. B. & Ristow, S. (1973) *Ann. Rev. Biochem.* **42,** 135.
16. Karplus, M. & Weaver, D. L. (1976) *Nature* **260,** 404.
17. Karplus, M. & Weaver, D. L. (1976) *Biopolymers* **18,** 1421.
18. Honeycutt, J. D. & Thirumalai, D. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3526.
19. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Biopolymers* **28,** 1059.
20. Kolinski, A., Skolnick, J. & Yaris, R. (1987) *Biopolymers* **26,** 937.
21. Sikorski, A. & Skolnick, J. (1989) *Biopolymers* **28,** 1097.
22. Skolnick, J. & Kolinski, A. (1990) *Science* **250,** 1121.
23. Sikorski, A. & Skolnick, J. (1990) *J. Mol. Biol.* **213,** 183.
24. Brooks, C. L., Petitt, M. & Karplus M. (1988) *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics,* Wiley, New York.
25. McCammon, J. A. & Harvey, S. C. (1988) *Dynamics of Proteins and Nucleic Acids,* Cambridge University Press, Cambridge.
26. Mezard, M., Parisi, G. & Virasoro, M. A. (1987) *Spin Glass Theory and Beyond,* World Scientific, Singapore.
27. Angell, C. A. (1988) *Phys. Chem. Solids* **49,** 863.
28. Frauenfelder, H., et al. (1990) *J. Phys. Chem.* **94,** 1024.
29. Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E., Sanke, T., Shyamasunder, E. & Young, R. D. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 5000.
30. Frauenfelder, H., Parak, F. & Young, R. D. (1988) *Ann. Rev. Biophys. Biochem.* **17,** 451.
31. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524.
32. Bryngelson, J. D. & Wolynes, P. G. (1987) *J. Phys. Chem.* **93,** 6902.
33. Wolynes, P. G. (1990) in *Spin Glasses and Biology,* Stein, D. L., Ed., World Scientific Press, Singapore.
34. Levitt, M. & Sharon, R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8557.
35. Guo, Z. & Thirumalai, D., to be published.
36. Edwards, S. F. (1965) *Proc. Phys. Soc.* **85,** 613.
37. Oono, Y. (1985) *Adv. Chem. Phys.* **61,** 301.
38. Dill, K. A. (1990) *Biochemistry* **29,** 7133.
39. van Gunsteren, W. F. & Berendsen, H. J. C. (1977) *Mol. Phys.* **34,** 1311.

40. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *J. Chem. Phys.* **85,** 3585.
41. Gibson, K. D. & Scheraga, H. A. (1987) in *Structure and Expressions: Nucleic Acids and Proteins,* Sarma, R. H. & Sarma, M. H. Eds., Adenine, Guilderland, NY.
42. Li, Z. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci.* **84,** 6611.
43. Scheraga, H. A. & Paine, G. H. (1986) *Ann. NY Acad. Sci.* **482,** 60.
44. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. (1982) *J. Chem. Phys.* **76,** 637.
45. Andersen, H. C. (1983) *J. Comp. Phys.* **52,** 24.
46. Schulz, G. F. & Schirmer, R. H. (1979) *Principles of Protein Structure,* Springer-Verlag, New York.
47. Novotny, J., Bruccoleri, R. & Karplus, M. (1984) *J. Mol. Biol.* **177,** 787.
48. Novotny, J., Rashin, A. A. & Brucolleri, R. E. (1988) *Proteins* **4,** 19.
49. Harel, T. F., Kuntz, I. D. & Crippen, G. M. (1983) *Bull. Math. Biol.* **45,** 665.
50. Ptitsyn, O. B. & Finkelstein, A. V. (1980) *Quart. Rev. Biophys.* **13,** 339.
51. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167.
52. Stillinger, F. H. & Weber, T. A. (1982) *Phys. Rev.* **25,** 978.
53. Stillinger, F. H. & Weber, T. A. (1987) *Science* **225,** 318.
54. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983) *Science* **220,** 671.
55. Elber, R. & Karplus, M. (1987) *Science* **235,** 318.
56. Noguti, T. & Go, N. (1989) *Protein* **5,** 97.
57. Honeycutt, J. D. & Andersen, H. C. (1987) *J. Phys. Chem.* **91,** 4950.
58. de Gennes, P. G. (1985) in *Scaling Concepts in Polymer Physics,* Cornell University Press, Ithaca, NY.
59. Head-Gordon, T., Stillinger, F. H. & Arrecis, J. (1991) AT&T Bell Laboratory Preprint, February.
60. Thirumalai, D., Mountain, R. D. & Kirkpatrick, T. R. (1989) *Phys. Rev. A* **39,** 3563.
61. Thirumalai, D. & Mountain, R. D. (1990) *Phys. Rev. A* **42,** 4574.
62. Smith, M. B. (1964) *Aust. J. Biol. Sci.* **17,** 261.
63. Smith, M. B. & Back, J. F. (1965) *Aust. J. Biol. Chem.* **18,** 365.
64. Smith, M. B. & Back, J. F. (1968) *Aust. J. Biol. Chem.* **21,** 539.
65. Smith, M. B. & Back, J. F. (1968) *Aust. J. Biol. Chem.* **21,** 549.
66. Goux, W. J. & Venkatasubramanian, P. N. (1986) *Biochemistry* **25,** 84.
67. Richardson, J. S. & Richardson, D. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation,* Fasman, G. D., Ed., Plenum Press, New York.
68. Chan, H. S. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6388.
69. King, J. (1989) *Chem. Eng. News* **67,** 32.
70. Richards, F. M. (1991) *Scientific American,* p. 54, (January, 1991).
71. Friedrichs, M. S. & Wolynes, P. G. (1989) *Science* **246,** 371.
72. See also Stein, D. L. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3670.
73. Shakhnovich, E. J. & Gutin, A. M. (1989) *J. Phys. A* **22,** 1647.

Received April 3, 1991
Accepted October 14, 1991