Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 3526–3529, May 1990
Biophysics

Metastability of the folded states of globular proteins

J. D. HONEYCUTT AND D. THIRUMALAI

Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742

Communicated by Robert Zwanzig, February 26, 1990 (received for review November 11, 1989)

ABSTRACT **The possibility that several metastable minima exist in which the folded forms of a polypeptide chain have similar structural characteristics but different energies is suggested. The validity of this hypothesis is illustrated with the aid of simulation methods on a model protein that folds into a β-barrel structure. Some implications of this hypothesis such as the existence of multiple pathways with intermediates for protein folding are discussed.**

The early experimental studies of the reversible denaturation of proteins laid the foundation for the hypothesis that the folding of polypeptide chains is a thermodynamically controlled process (1). This hypothesis gained further supporting evidence from the studies of Anfinsen and coworkers (2) who found that ribonuclease in the denatured state not only refolds to its native structure with the correct disulfide bridging but also regains complete enzymatic activity. Based on these studies, the so-called thermodynamic hypothesis was advanced, which asserts that the three-dimensional folded structure of a polypeptide chain in a specified environment is invariably the one with the lowest Gibbs free energy (3, 4). The thermodynamic hypothesis has been used as the justification for computing the conformation of proteins using energy minimization methods (5). On the other hand, the kinetic control of protein folding suggests that an energy barrier hinders the protein from forming the state of the lowest free energy, thus trapping it in a metastable state (6). The basis for the formation of the folded state of protein molecules by a kinetically controlled process has its genesis in a calculation by Levinthal (7), who argued that the time for the random search of all conformations of even moderately sized polypeptide chains to attain the state of lowest free energy would be astronomically large and hence would be incompatible with biological time. This idea led to the picture that the protein folds into a unique metastable state, which would be the one with the lowest free energy among all the kinetically accessible structures.

In this report we present a different hypothesis based on the notion of kinetic control for the folding of a polypeptide chain. The hypothesis, referred to as the metastability hypothesis, can be stated as follows: 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 (i.e., they exhibit similar equilibrium bond angle and dihedral angle distributions that are computed using those configurations belonging to a given minimum) but have different energies from one another. Almost all of the folded states are metastable and the transition from one minimum to another is infrequent. The various structures typically differ in certain dihedral angles especially in the bend regions. The particular state into which the protein folds depends on the initial condition. In other physical systems, it is not uncommon for two different states to have similar statistical structural characteristics with differing energies. For example, metal–metalloid alloys in their liquid and glassy states have similar structure factors but their energies can be very different (8). The formation of metastable folded conformations by a kinetically controlled mechanism has been known in the context of pleating of polymer molecules in their crystallization (9, 10). The above hypothesis is different from the previously advocated kinetic controlled process in that we do not suggest that folding occurs by a directed specific pathway that drives the chain into a unique metastable state. The postulate that there are several metastable states in which the polypeptide acquires similar structures implies that there are multiple pathways for the folding process (11). If this is the case then the specific pathway chosen by a polypeptide chain depends on the initial condition of the environment, namely, solvent ionic strength, presence of an appropriate chaperon molecule, etc. The plausibility of such a hypothesis is demonstrated by our molecular dynamics simulation of a model protein molecule that exhibits several different β-barrel-shaped structures differing in the energies in the folded state. The various folded structures were obtained by a combination of a simulated annealing technique (12), the steepest descent quench method (13, 14), and slow cooling.

The model heteropolymer representing the polypeptide chain is one containing as few parameters as possible (unpublished results). For our purposes we chose to consider only the backbone of the chain; side chains are not explicitly included. This makes the model somewhat unrealistic but we feel it is sufficient to illustrate the principle behind the metastability hypothesis. The only guide we have used to construct the model is that the various terms in the Hamiltonian representing the peptide chain be calculable (at least in principle) using quantum mechanical or statistical mechanical methods. In our simulations we consider n beads representing the various residues, with $n = 46$. Some of these are chosen as hydrophobic, some as hydrophilic, and others as neutral. The neutral residues are expected to occur at the various bends of the β-barrel structure. The bond lengths between successive beads are constrained to be of equal fixed length. Three forms of interactions were used to obtain the Hamiltonian for the polypeptide chain: (*i*) a long-range force (range referring to the distance along the backbone of the chain) mimicking the hydrophobic attraction; (*ii*) a harmonic potential constraining the bond angle between three successive beads; (*iii*) a potential term for the (dihedral angle) rotation about the central bond connecting four successive residues. The Hamiltonian for a chain with n beads is taken to be a sum of the three terms.

The interaction between the hydrophobic residues, which is a potential of mean force, is modeled using the Lennard–Jones potential. The potential between a hydrophilic residue and either another hydrophilic or a hydrophobic residue is taken to be purely repulsive with the range being somewhat larger than a simple r^{-12} repulsive potential. The neutral residues interact with other residues including one another through a purely repulsive r^{-12} potential. A standard form for the dihedral angle potential is chosen where the trans state is the most stable and the two gauche states are slightly higher in energy. In the bend region, which involves the neutral

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

residues, the parameters in the dihedral angle potential are adjusted so that the three minima—namely, the trans and the two gauche states—have the same energy with a reduced barrier between the three states. This facilitates the formation of the bends in the β-barrel structure. This model is similar in spirit and in some details to that used by Skolnick *et al.* (15, 16) in their lattice Monte Carlo simulation of the β-barrel structure but differs in certain crucial ways, the most obvious of which is that ours is a continuum model.

The simulations were done using a modified molecular dynamics method based on the velocity form of the Verlet algorithm (17) with a frictional force and Gaussian random force added. The main purpose of the random force is to control the temperature of the system in a physically reasonable way. The RATTLE algorithm introduced by Andersen (18) was used to account for the bond length constraint. The analysis of the underlying topography of the free-energy hypersurface for our model was done in two ways. In the first method a high-temperature state of the polypeptide chain was prepared. It was found that the chain is in a completely denatured state. A typical snapshot is presented as Fig. 1. The process of folding was initiated by performing several "slow" coolings of the chain from the initially denatured state. A slow cooling run was performed by linearly decreasing the bath temperature from a value well above the unfolding transition temperature to zero over a specified period of time $t_{max} = N\tau$, where τ is the natural time unit in the problem and N typically varies from 500 to 2500. For our problem τ corresponds to 1–2 psec. The cooling rate depends on the duration of the simulation and can be made to vary within limits imposed by computer time. The slowest cooling runs yielded structures that were visually indistinguishable from a β-barrel conformation. Faster coolings led to defective structures differing noticeably from a β-barrel. A typical snapshot of a β-barrel structure obtained by slow cooling is given in Fig. 2. The slow coolings from a denatured state were used to develop several metastable structures, many of which were native-like. In addition, some initial conditions yielded structures that had lower energies than some of the β-barrel-like structures but differed considerably in their shapes. These could be characterized as having defects of some kind (19) and, with long annealing times, these structures presumably could make transitions to the desired folded structures.

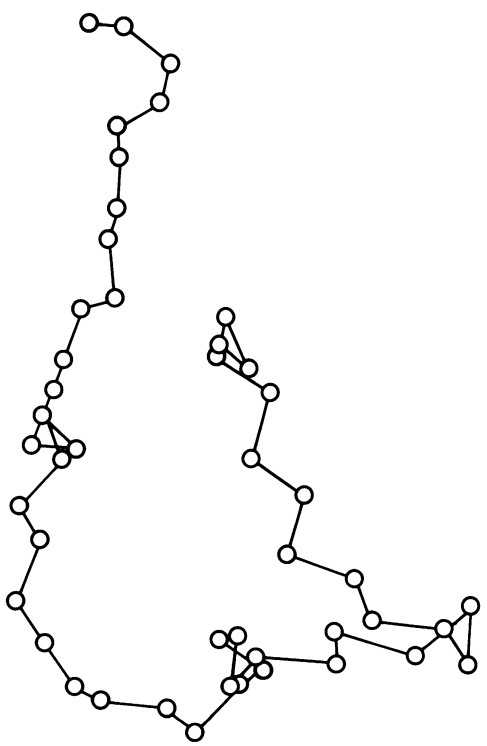

FIG. 1. Snapshot of the polypeptide chain at high temperature.

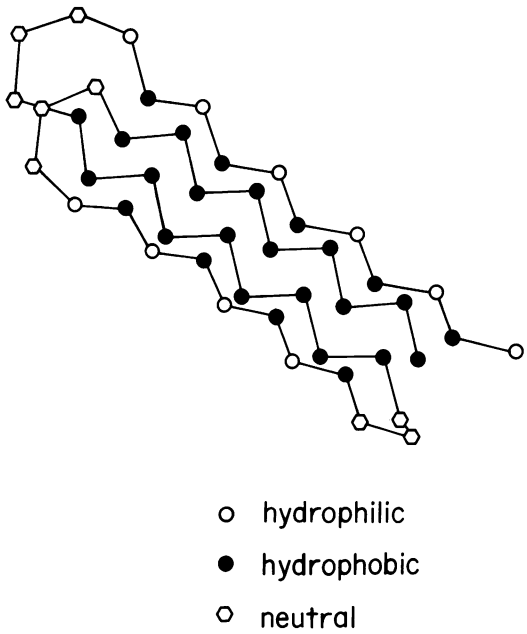

FIG. 2. Typical β-barrel structure of the chain below the unfolding–folding transition temperature. This configuration was obtained by a process of slow cooling. The nature of each of the residues is shown by the various symbols.

To probe the existence of other possible minima that have β-barrel-like structures, we performed the following calculation to enhance the sampling of the configuration space. A trial β-barrel structure was constructed "manually" such that all bond 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 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 (13, 14). The structure that resulted from this had the desired folded conformation that we denote as β1. To test the metastability hypothesis further, we searched for other β-barrel-like structures with energies less than that of the β1 structure. Starting from the β1 structure the temperature of the system was raised, keeping it well below the unfolding transition temperature as determined from slow cooling runs. The system was then repeatedly annealed for lengths of time that 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 equivalent to the simulated annealing Monte Carlo technique (12). 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. The nature of the various β-barrel structures was investigated using the distance matrix (20), the elements of which are distances between pairs of residues. The plots of the distance matrix reveal that all the β-barrel

structures have the characteristic signatures of parallel and anti-parallel β-strands. The various structures are also visually similar in having the characteristic overall shape of a β-barrel with all the hairpin turns in the proper place. They differ in certain bond and dihedral angles, especially in the bend regions, and consequently have different energies. The differences in the energies of the various structures imply that all these are truly distinct metastable states.

We note that the discovery of multiple minima in inhomogeneous systems like proteins is not new. They have been found experimentally (21) as well as in the recent molecular dynamics studies of the conformational substates of myoglobin (22). In the latter study the various conformational substates differ in the rotations of the side chains (22). There is practically no change in the coordinates of the backbone of the chains. Our simulations have shown that many metastable minima, which are not merely conformational substates, exist when our model chains fold into ordered structures. The more important observation for protein folding is that the ordered structures of the backbone of the chain in all the minima may be statistically indistinguishable. Thus, insofar as the enzymatic activity of a real native protein is essentially determined by the conformation, any one of these metastable structures might be sufficient. The precise structure that is obtained in our studies is clearly dependent on the initial conditions and, hence, on the pathway.

To further characterize the various structures obtained in our simulations, we have used the "inherent structure" technique (13, 14). In this method, a system is characterized by the structure and energy it attains after the thermal energy is removed by a steepest descent quench. We have performed (*i*) quenches of the model β-barrel structures obtained from β1 by simulated annealing, (*ii*) quenches of structures obtained by cooling from a high temperature, and (*iii*) direct quenches of chains in the denatured state. The energies of each of the quenched structures form a "spectrum" that can be plotted and analyzed (23).

A summary of those computations is shown in Fig. 3 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 obtained from simulated annealing, the quenched structures obtained from slow cooling, and the minima obtained from a direct quench from the denatured state. Some of the values in the second column are for structures that are indistinguishable from β-barrels, both visually and in terms of their distance matrices (unpublished results). The formation of these structures requires slow cooling rates. Other values in the second column correspond to structures that do not have hairpin turns at the correct locations and can best be classified as distorted (or defective) versions of the β-barrel structures. The energies of a few of these are actually lower than some of the values in the first column. Thus, it follows that the chain can adopt "defective" conformations that actually have lower energies than certain β-barrel structures. Presumably, these defective structures could anneal in time to ordered structures, but the time scale may be longer than that which can reasonably be simulated. The local minima represented in the third column do not even qualitatively resemble β-barrels and should be viewed as random structures. If the initial conditions, which are determined by the environment, are such that the conformation of the chain maps into one of these minima, one does not obtain the folded β-barrel structure. The time for escape out of such a free-energy valley would depend on the barrier height separating the minimum and the folded state and might exceed the biological time. 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 making it inefficient in its physiological activity.

Although we have found that certain statistical characteristics of the various β-barrel structures corresponding to the different minima are virtually identical, it is of interest to compute the fluctuations in these structures. To analyze the nature of these fluctuations, we obtained the mean deviation in the dihedral angle from the idealized folded β-barrel structure. The latter was taken to be the one with the lowest energy found in our simulations. Starting from this initial structure at $T = 0$, we slowly heated the system, allowed it to equilibrate, and computed the quantity

$$\Delta\phi = \left\langle \frac{1}{N_c} \sum_c (\phi^c - \phi_o^c)^2 \right\rangle^{1/2}, \quad [1]$$

where ϕ_o^c is the value of the dihedral angle in the ideal zero temperature β-barrel and ϕ^c is the corresponding value at the temperature of the system. The summation is over the N_c ($= n - 3$) dihedral angles in the system and the angular bracket denotes a statistical mechanical average. In Fig. 4, we show $\Delta\phi$ as a function of temperature. This figure shows that $\Delta\phi$ can be as large as 25° even below the folding–unfolding transition temperature. Thus, the polypeptide chain can tolerate large fluctuations in the backbone conformation and still preserve the folded structure. To compare static differences in the various β-barrel structures, we also computed the quantity

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

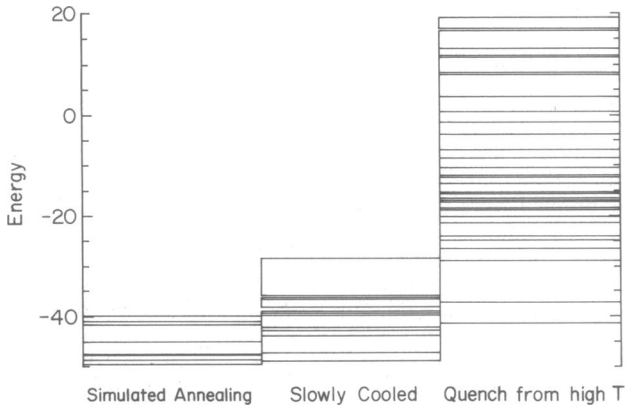

FIG. 3. Spectrum of energies of the various quench structures obtained using several simulation techniques. The explanation of the three columns is given in the text. T, temperature.

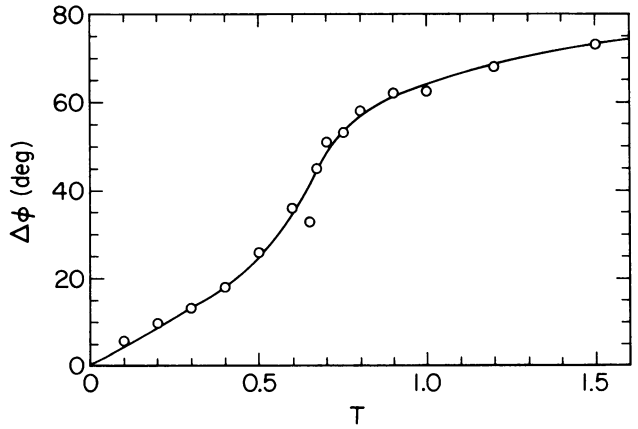

FIG. 4. Fluctuation in the dihedral angle in degrees as a function of temperature (T). The fluctuation is measured with respect to the ideal zero temperature β-barrel structure.

where X_i^α is the coordinate of the ith residue in the αth β-barrel structure, and x_i^0 is the corresponding value in the minimum energy structure. One end of the chain was taken to be the origin of the coordinate system. The quantity Δ_α is the statistical fluctuation in the secondary structure from its ideal value. The values in the different β-barrel minima varied from 0.64σ to 3.83σ, where σ ($\approx$3 Å) is the effective range of the hydrophobic attraction. The higher free-energy β-barrel structures are slightly less compact than the ideal folded conformer. Our model study suggests the possibility that large-scale fluctuations in the backbone can still preserve the folded structure.

We should emphasize that the validity of the metastability hypothesis has been illustrated using a simple model and the relevance for real proteins remains to be shown. Proteins have much more intricate structure than has been assumed in our model and hence it is much harder to test our hypothesis for a realistic model of proteins. Nevertheless, our study introduces the possibility that these metastable states are relevant for protein folding. Despite this, our results are not inconsistent with the detailed experimental studies of the kinetics of the folding transition in bovine pancreatic trypsin inhibitor (24) and *Staphylococcus aureus* (25). The bovine pancreatic trypsin inhibitor kinetic study suggested that there are metastable states that have native-like structures. In addition the authors of these studies obtained intermediates with local conformations not found in the native proteins. In the refolding pathway of *S. aureus*, two metastable moieties that have secondary structures similar to the native proteins have been identified (25). The experimental demonstration of multiple valleys in which proteins exhibit structures (and stability) resembling the native conformation adds additional support for the hypothesis advanced here.

We are grateful to R. Zwanzig for several enlightening remarks. We thank R. Pastor for useful comments on the manuscript. This work was supported in part by a grant from the National Science Foundation. The computations were supported by a generous grant of supercomputer time by the San Diego SuperComputer Center.

1. Mirsky, A. E. & Pauling, L. (1936) *Proc. Natl. Acad. Sci. USA* **22,** 439–447.
2. Epstein, C. J., Goldberger, R. F. & Anfinsen, C. B. (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28,** 439–449.
3. Anfinsen, C. B. (1973) *Science* **181,** 223–230.
4. Anfinsen, C. B. & Scheraga, H. A. (1975) *Adv. Prot. Chem.* **29,** 205–300.
5. Levitt, M. (1980) in *Protein Folding*, ed. Jaenicke, R. (North-Holland, Amsterdam).
6. Wetlaufer, D. B. & Ristow, S. (1973) *Annu. Rev. Biochem.* **42,** 135–158.
7. Levinthal, C. (1968) *J. Chim. Phys.* **65,** 44–45.
8. Cargill, G. S., III (1975) *Solid State Phys.* **30,** 227–320.
9. Frank, F. C. & Tosi, M. (1961) *Proc. R. Soc. London Ser. A* **263,** 323–339.
10. Zwanzig, R. & Lauritzen, J. I. (1968) *J. Chem. Phys.* **48,** 3351–3360.
11. Harrison, S. C. & Durbin, R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4028–4030.
12. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983) *Science* **220,** 671–680.
13. Stillinger, F. H. & Weber, T. A. (1982) *Phys. Rev. A* **25,** 978–989.
14. Stillinger, F. H. & Weber, T. A. (1984) *Science* **225,** 983–989.
15. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Proc. Natl. Acad. Sci. USA* **85,** 5057–5061.
16. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Biopolymers* **28,** 1059–1095.
17. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. (1982) *J. Chem. Phys.* **76,** 637–649.
18. Andersen, H. C. (1983) *J. Comput. Phys.* **52,** 24–34.
19. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
20. Havel, T. F., Kuntz, I. D. & Crippen, G. M. (1983) *Bull. Math. Biol.* **45,** 665–720.
21. Frauenfelder, H., Parak, F. & Young, R. D. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17,** 451–479.
22. Elber, R. & Karplus, M. (1987) *Science* **235,** 318–321.
23. Honeycutt, J. D. & Andersen, H. C. (1987) *J. Phys. Chem.* **91,** 4950–4963.
24. Creighton, T. E. & Goldenberg, D. P. (1984) *J. Mol. Biol.* **179,** 497–526.
25. Mitchinson, C. & Pairs, R. H. (1985) *J. Mol. Biol.* **184,** 331–342.