Theor Chem Acc (2003) 109:8–21
DOI 10.1007/s00214-002-0390-6

Theoretical
Chemistry Accounts

Feature article

Study of the dynamics of protein folding through minimalistic models

Goundla Srinivas, Biman Bagchi

Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560 012, India

Received: 28 April 2002 / Accepted: 11 August 2002 / Published online: 13 November 2002

Abstract. Starting with the Levinthal paradox, a brief introduction to the protein folding problem is presented. The existing theories of protein folding, including the folding funnel scenario, are discussed. After briefly discussing different simulation studies of model proteins, we discuss our recent work on the dynamics of folding of the model HP-36 (the chicken villin headpiece) protein by using a simplified hydropathy scale. Special attention has been paid to the statics and dynamics of contact formation among the hydrophobic residues. The results obtained from this simple model appear to be surprisingly similar to several features observed in the folding of real proteins. The account concludes with a discussion of future problems.

Keywords: Protein folding – Free energy – Landscape – HP-36

1 Protein folding problem

Proteins are essential for life. The first functionally important action of a protein is to fold correctly into its unique native state. Proteins must fold correctly into their native state in order to perform their predetermined biological activity. Misfolded proteins not only fail to function but they can also cause diseases. For example, misfolded forms of the prion protein (PrP) are the infectious particles that cause a group of diseases known as spongiform encephalopathies [1]. To be specific, the conformational conversion of the PrP from its normal soluble helical conformation to an insoluble β-sheet state (misfolded state) causes the prion disease and infectivity [1]. β-Amyloid protein deposits, called plaques, are found in the brains of people with Alzheimer's disease. The reason for the formation of these deposits is an area of active research [2]. It has been suggested very recently that misfolded protein monomers coalesce into full-blown aggregates which are toxic [3].

Protein folding is obviously a very complicated time-dependent process involving large-scale changes in the conformation and the volume [4] of the heteropolymer. The objective of this account is to articulate our recent work and understanding of some aspects of this complex problem through the study of minimalistic models. Such models have been used extensively by many groups and have proven to be useful in generating and testing broad ideas.

Clearly, a satisfactory understanding of the protein folding problem requires the solution of a large number of associated problems, many of which are of interest in their own right. Nevertheless, the ultimate goal is to predict the three-dimensional structure of the native state, correct to atomic resolution, once the primary sequence is given. Many different approaches have been adopted, from statistical analysis of existing analogous structures to the use of genetic algorithms, to predict the tertiary structure. The focus of the present review is on the theoretical, primarily statistical mechanical and computational, approach to the problem.

Although the details of the process of protein folding are not yet fully understood, significant progress has been made in recent years (a number of excellent reviews are available [5–9]). Most of the earlier studies were based on the widely accepted belief that the folded proteins exist in the global minimum free-energy state. This is known as the Anfinsen hypothesis. Many years ago, Anfinsen [10] predicted that small globular proteins can fold in the absence of any catalytic biomolecules. Small proteins typically fold rapidly and reliably to a unique native state from any one of the vast number of initial unfolded conformations. The rapidity of folding, the complexity of protein structure, and the fact that each molecule in its extended state can take a microscopically different path to the finally folded state make protein folding a very difficult problem for both experimental and theoretical approaches [11–15].

Correspondence to: B. Bagchi
e-mail: bbagchi@sscu.iisc.ernet.in

1.1 Levinthal's paradox

If the native state of the protein is indeed the most stable, meaning the lowest free-energy state, then the question naturally arises: How does the protein get there? This question is of vital importance in the understanding of the protein folding problem and it was first elegantly coined by Levinthal in the following fashion: "How long does it take for an extended, unfolded protein to reach its final, folded native state, through a random search of its configurations?". In order to find an obvious (but misleading!) answer to this question, let us consider a protein with 101 amino acids. For simplicity, we restrict the analysis to the case in which each bond can have only three states (conformations). Even in this simple case, there exist 3^{100} possible states for the protein. Among these, since the native state is unique, only one state corresponds to the native state. Thus, in order to find the correct native state, Levinthal assumed that the protein has to randomly search each of these configurations. Even at the rate of 10^{13} new configurations per second, it requires 10^{27} years (a cosmologically long time) to reach the native state. Nevertheless, in reality, proteins fold into their stable native state within seconds or less. This observation is the celebrated Levinthal paradox [16].

The previous example reveals that random searches are not an effective way for the proteins to find the correctly folded native state. More importantly, Levinthal's paradox suggests that the protein explores no more than a tiny fraction of the total number of states and may even adopt a kinetically accessible, metastable configurations that may not have the lowest free energy! Thus, the final functionally active state of a folded protein may be realized through kinetic or thermodynamic control. However, the final state must be sufficiently stable to survive long enough to perform biological functions.

1.2 The blind watchmaker

Dawkins [17], the well-known British neo-Darwinianist, considered another paradox: How long does it take for a monkey to type the following line from Hamlet's remark, "Me think it is like a weasel"? This sentence contains 28 characters (including five spaces). Since there are 26 letters and a space, to reach the correct statement by random typing, the monkey has to try out 27^{28} key strokes. This is again a very big number! On the other hand, if we do not allow the monkey to change those letters that are already correctly typed in place, even by random typing, Hamlet's remark can be reached in only a few thousand key strokes! Thus, this paradox can be resolved by replacing a random search by a biased, "enlightened" search!

1.3 Resolving Levinthal's paradox

Both the examples already mentioned reveal that to reach a final well-defined state (folded state in protein folding, Hamlet's remark in *The blind watchmaker*) random searches are not effective ways. In particular, the latter example suggests that the biased searches are more effective than completely random searches. Then, naturally one raises the following questions: "Is there any bias in protein folding to resolve the Levinthal paradox?"; "How much bias is required in order to get a biologically meaningful time scale?". In a simple but revealing study of this problem, Zwanzig, Szabo and Bagchi (ZSB) [18] presented a simple model where they considered an initial configuration of a protein with $N+1$ amino acids, having S "incorrect contacts". They gave an expression to calculate the folding time, τ, as

$$\tau(S) \approx \left(\frac{1}{Nk_0}\right)\left(1+\frac{k_0}{k_1}\right)^N \quad , \tag{1}$$

where k_0 and k_1 represent the rate at which a "correct contact" can become an "incorrect contact" and an "incorrect contact" can become a "correct contact", respectively. When $k_0 = k_1$, we get the time required by a fully random search and this grows as 2^N and indeed is very large for $N > 100$. However, for a biased search $k_1 > k_0$, and the ratio k_0/k_1 is small, which gives rise to a drastic reduction in the time required to find the native state from a randomly chosen extended state. For a fully biased search ($k_0 = 0$), Eq. (1) takes the following form:

$$\tau(S) = \left(\frac{1}{k_1}\right)\sum_{j=1}^{S}\frac{1}{j} \quad . \tag{2}$$

By this approach, ZSB demonstrated that a reasonable energy bias (of the order of a few k_BT) against locally unfavorable configurations reduces Levinthal's time to a biologically significant size, like a second or so, for the time required to fold the protein.

This analysis provides an example of competition between the entropy and the energy – the entropic cost is reduced by the energetic gain during the folding. The gain in energy (which translates into lowering of free energy) acts as the driving force. This suggests that the free-energy surface for folding essentially resembles a multidimensional funnel [19–24] where one has a large entropy (many configurations) at higher energy at the top of the funnel while much fewer states with lower energy in the lower part of the funnel. The protein sits at the bottom of the funnel. The examples just discussed provide a direct link to the funnel-like picture. While the random searches amount to spending a lot of time in the nearly flat surface, characterized by lots of small minima, in a region which is away from the folding pathway (in the rugged landscape), the biased searches amount to traveling down a smooth funnel. Indeed, by considering the "correctness" of the configuration (in the ZSB model), Zwanzig [25] later showed that one can obtain a funnel-like landscape where the final native state is separated from the disordered state by an entropic barrier. A detailed description of the energy landscape picture is given in later sections.

2 Protein folding theories

The ultimate goal of any theoretical approach of protein folding is to predict the tertiary structure from any given sequence of amino acids. This lofty goal, of course,

includes the prediction of the secondary structures as well. Progress towards achieving this goal has been slow. Early statistical mechanical theories of protein folding considered single-domain small proteins [26, 27] because of their simplicity. Attempts are now currently being made to understand the secondary structure formation.

2.1 The statistical field theory of heteropolymer collapse

Bryngelson and Wolynes (BW) proposed a theory by "statistically" extending the generalized Flory theory [28]. This theory consists of two parts, namely heteropolymer collapse and an ordering of the residues into the native state.

According to this theory, the free energy of a Gaussian heteropolymer consisting of monomers of the same size with radius σ and end-to-end distance R, confined in a volume V can be expressed as a function of its linear size (R) and temperature, T, as

$$F(T,R) = -NT\log V - 3T\log\left(\frac{R}{R_0}\right) + \frac{3}{2}T\left(\frac{R}{R_0}\right)^2 + \left[\frac{1}{2}T - z\left(\bar{k} + \frac{\Delta\kappa^2}{2T}\right)\right]\frac{N^2\sigma^3}{R^3} + \frac{1}{6}T\frac{N^3\sigma^6}{R^6} \quad , \tag{3}$$

where $R_0^2 = Nl^2$ ($l = 2\sigma$) and $-\bar{\kappa}$ is the average energy of interaction. The minimization criterion for the free energy leads to

$$\frac{\partial F}{\partial R} = -\frac{3T}{R} + \frac{3T}{R_0^2}R - \frac{3}{2}\left(1 + \frac{\Delta T}{T}\right) \times (T - T_\theta)\frac{N^2\sigma^3}{R^4} - \frac{TN^3\sigma^6}{R^7} = 0 \quad , \tag{4}$$

where T_θ denotes the θ temperature.

This approach can be extended to protein folding by assuming an intrinsic energetic preference of the monomers for the native conformation. If the interaction energies are defined for the primary structure as $-\epsilon$, for the secondary structure as $-J$ and for the tertiary structure as $-\kappa$, for the residues involved in the interaction in their native state, this theory provides a simple two-parameter model for protein folding. These two parameters are η, the generalized packing fraction, and ρ, the fraction of amino acid residues in the native state. The free energy can be written in terms of T and $\eta(\rho)$ as

$$\begin{aligned} F(T,\rho) = N\Big\{ & -\epsilon\rho - J\rho^2 - z\bar{\eta}(\rho). \\ & \times\left[\left(\bar{\kappa} + \frac{\Delta\kappa^2}{2T}\right) + \left(\kappa - \bar{\kappa} - \frac{\Delta\kappa^2}{2T}\right)\rho^2\right] \\ & + \rho T\log\rho + T(1-\rho)\log\rho + T(1-\rho) \\ & \times\log\left(\frac{1-\rho}{V}\right) + T\left[\frac{1-\bar{\eta}(\rho)}{\bar{\eta}(\rho)}\log[1-\bar{\eta}(\rho)]\right] + T\Big\} \\ & + \frac{3T}{2}\left(\frac{\bar{R}(\rho)}{R_0}\right)^2 - 3T\log\left(\frac{\bar{R}(\rho)}{R_0}\right). \end{aligned} \tag{5}$$

The model of BW predicts a first-order phase transition from the extended to the ordered state [29]. The free-energy surface shows two minima, which correspond to the extended and the native state. The two minima are separated by a barrier. The model could thus capture the cooperative nature of the transition. Another interesting point is that starting from the disordered state, there are several alternative pathways to the native minimum.

2.2 The statistical thermodynamic theory of thermal stabilities of globular proteins

In spite of the previously mentioned implications, BW theory is not sufficiently sophisticated to predict the tertiary structure for a given sequence of amino acids – it is a simple two-parameter theory meant to catch the essence of the folding process.

A theory similar in spirit to BW was proposed by Dill et al. [27]. This theory is also a two-parameter theory, with the packing density and the fraction of hydrophobic residues in the core as the two parameters. The free-energy barrier to the folding was calculated by constructing a thermodynamic pathway consisting of the following two steps.

1. Random collapse of the unfolded chain without any structuring so that the hydrophobic and hydrophilic residues are randomly distributed throughout a condensed globular structure.
2. Formation of a compactly folded form by the subsequent reconfiguration leading to an arrangement of hydrophilic residues in the surface which surrounds a core of hydrophobic residues.

Thus, according to this theory, the driving force for the protein folding is the force originating from the hydrophobic interactions. As a result, the net change in free energy can be written as

$$\begin{aligned} \Delta F_1 &= \Delta F_h\Phi_1 + \Delta F_{\text{conf1}} \quad \text{(for step 1)}, \\ \Delta F_2 &= \Delta F_h\Phi_2 + \Delta F_{\text{conf2}} \quad \text{(for step 2)} \quad , \end{aligned} \tag{6}$$

where Φ_i denotes the fraction of hydrophobic residues in the sequence. Thus, the total change in free energy of folding is

$$\Delta F = \Delta F_1 + \Delta F_2 \tag{7}$$

and on the basis of protein models (to be described later) the free-energy functions appearing in the previous expressions can be derived. This theory also predicts a first-order phase transition between the folded and the extended states of the polymer.

Thus, the theories of BW and Dill et al. could capture some of the essence of the protein folding problem. These initial theories were followed by a series of studies which vastly improved our understanding of protein folding [16, 18, 20, 21, 25, 30–36]. Recently Portman and coworkers [37, 38] presented a more detailed microscopic theory of protein folding rates. By studying the effect of chain stiffness on the fine structure of the free-energy profile, they found that increasing the persistence length of the chain tends to smooth the free-energy

profile. By neglecting the non-native contacts and trapping effects, they could obtain the reaction coordinates and folding rate prefactors for specific proteins with known native structures.

3 Energy landscape picture

From the numerical and analytical studies of various models Wolynes, Onuchic and others [19, 20, 39–42] introduced and elaborated the concept of an energy landscape for protein folding. In particular, following the general consideration of protein folding, Leopold et al. [19] introduced the concept of a "protein folding funnel", a kinetic mechanism to understand the self-organizing principles of the protein sequence–structure relationship. Crudely speaking, a folding funnel is a collection of geometrically similar collapsed structures, one of which is in the global minimum. As pointed out by Dill [5], the hydrophobic collapse is probably the initial step in the protein folding; however, secondary structures, like helices and hair-pin bends, may occur simultaneously. Thus, folding can occur simultaneously along different coordinates. According to this latter development, the folding kinetics is determined by an energy landscape and for foldable proteins this resembles a funnel with a free-energy gradient towards the native structure. The introduction of the concept of a folding funnel provided a much needed breakthrough in the understanding of the pathways of protein folding.

The funnel-like energy landscape picture emphasizes the importance of a global overview of the protein's energy surface in understanding the folding process. This viewpoint will be of great help if folding occurs through organizing an ensemble of structures rather than through only a few uniquely defined sequentially arranged structural intermediates. In such a case, a statistical description of the energy landscape can, in principle, be used to describe the protein folding. Such a description can be built by using polymers, as explained later.

4 Driving forces in protein folding

In order to understand the protein folding problem, it is essential to study the forces that drive the protein folding. Many years ago Kauzmann [43] studied the hydrophobic interactions in a great detail. He concluded that the hydrophobic interactions are one of the major driving forces in protein folding. Later studies supported his ideas [44]. There are several experimental facts to support this view:

1. A protein is less affected by mutation of protein surfaces than hydrophobic cores [45].
2. Hydrophobic interactions are the strong interactions among amino acids in water [46].
3. Hydrophobic interactions are the crucial structure-determining forces [47, 48].

According to this view, folding cooperatively and closely resembles the collapse of a polymer in a poor solvent. Other forces, such as hydrophilic forces, are weaker but they can affect the stability [5]; however, the role of hydrophilic forces in protein folding is still not completely known.

Guided by the hydrophobic interactions, a minimal model of globular protein behavior can be constructed from a side-chain-centric perspective (solvation forces are dominant) rather than from a backbone centric are (nearest-neighbor amino acid interactions dominant). This leads to the possibility of designing polymers that can fold and perform protein-like functions, even without a peptide backbone. Thus, to a first approximation, a polymer can be modelled by a binary sequence of hydrophobic/polar monomers [49].

The minimal model can serve as a guide to obtain general features. At the same time, it is important to note that the minimal models cannot provide the balance of forces in real proteins. It should also be noted that in reality, since the stability of a protein is a small difference of large interactions, all the interactions contribute to the structure, thermodynamics and kinetics.

On the basis of these ideas a variety of models and real protein simulation studies have been carried out. In the following a few important classes of protein simulations are described in brief.

5 Protein simulation methods

5.1 Lattice models

Early simulation studies attempted to understand the protein folding by simulating the lattice models which are the simplest possible (and certainly rather crude) models to study the complex protein folding problem [50]. Later the use of lattice-based approaches were extended to predict the protein tertiary structure [6, 51–53]. Generally, in this model amino acid residues constituting the protein are represented by two types of beads (hydrophobic and hydrophilic). Each bead occupies one lattice site. The movement of beads on the lattice is governed by certain rules [46]. In this method, the approach to the native state is driven by the energy due to the favorable hydrophobic contacts between the appropriate residues. Thus, the folding pathway is essentially guided by a single order parameter, ρ_n, the fraction of the native topological contacts present in any given configuration. A statistically better and smooth folding path can be obtained by incorporating the Go-like potential [54].

The lattice models are protein-like in the sense that they fold to a unique native structure from an astronomically large number of possible initial conformations and do so rapidly, reproducibly and reversibly. The advantages of these models are twofold. First, the thermodynamic driving force for folding is an explicit part of the model, so the origin of the stabilizing forces can be separated from the problem of folding mechanism. Second, it is straightforward to study a large collection of folding events by direct simulation, with complete access to the structural details of every conformation that the polymer samples.

5.2 Off-lattice simulations

Although the lattice models are protein-like, they often omit features that are critical for understanding protein function. As a result, extracting useful information of the folding process from the abundance of detail provided by these numerical experiments remains a challenge.

Subsequent development found success in the study of model proteins. For example, off-lattice Langevin dynamic simulation studies [22, 55, 56] proved more promising in studying the dynamics of protein folding. In this method, protein is represented by a necklace of beads of different kinds. The solvent molecules surrounding the protein are implicitly represented through the solvent-averaged frictional force. This model allows the protein conformation to be mimicked through the application of geometrical constraints both on the relative separation and the orientation. Recent studies of this class demonstrated the folding of a model protein into a β-barrel structure [55]. A stochastic kinetic model for titin unfolding has been presented and studied by kinetic Monte Carlo off-lattice simulations [57]. These studies suggested the existence of several metastable minima in which the folded forms of the protein have similar structural characteristics but different energies.

5.3 All-atom simulations

While the lattice and off-lattice simulations consider model proteins where the amino acids are replaced by spheres of certain hydrophobicity, all-atom simulations explicitly study the interaction between all the atoms present in a real protein [58–62]. For example, an all-atom Monte Carlo simulation of a small peptide demonstrated many characteristic features of the folding process and supported the energy landscape and the funnel concept [60]. Notable among the all-atom simulations is the study of Kollman and Duan [58], who carried out the first ever 1-μs simulation of a protein in aqueous solution. They studied the thermostable chicken villin headpiece subdomain, a 36-residue protein (commonly known as HP-36 protein) in aqueous solution by explicit representation of water molecules. They found a nativelike structure with two pathways. Recently, this method was successfully applied to study both the thermodynamics and kinetics of the folding of a small peptide [60, 61].

6 Polymer collapse: relation to protein folding

It is a dream for experimental biologists to watch the protein folding through various conformational changes to reach the final native state. This is, however, still a difficult task owing to the biologically fast folding time of proteins. On the other hand, simulation studies provide a qualitative folding picture by modeling complex proteins as simple polymer chains. In the following, an attempt is made to understand the complex protein folding problem through the study of polymer collapse.

Recent simulation studies showed that the collapse of a homopolymer chain qualitatively resembles protein folding [63]. The variation in the radius of gyration of the polymer chain and the total energy are shown as a function of reduced time in Fig. 1. This figure shows that during the collapse, the size of the polymer decreases continuously without facing barriers. In other words, the collapse of a homopolymer is a continuous process. This model homopolymer is used to reproduce the experimentally observed bimodality [64, 65] in the energy-transfer efficiency between the donor–acceptor pair embedded at suitable locations along the backbone, during the folding [66].

7 Folding and unfolding of a single-domain model protein

As mentioned earlier, the folding of an extended, unfolded protein to its unique three-dimensional folded native state has attracted a great deal of interest [16, 67, 68]. Recent theoretical studies have suggested that the size, stability and topology of a protein influence the folding rate and mechanisms [26, 68, 69, 70]. This brings relevance to the funnel picture [21].

In order to obtain an energy landscape one needs to study the folding of individual members of the ensemble of proteins, for a long time. Given the complexity of protein folding and the amount of computational time it requires, this is a difficult task to achieve by conventional simulation techniques such as molecular dynamics/Monte Carlo or all-atom simulation [58] or even by Langevin dynamic simulations [55]. Lattice models are, of course, easier to study, but they are not appropriate in this case since each protein ultimately reaches the global minimum. These difficulties can be avoided by using the Brownian dynamics (BD) simulation technique. The BD

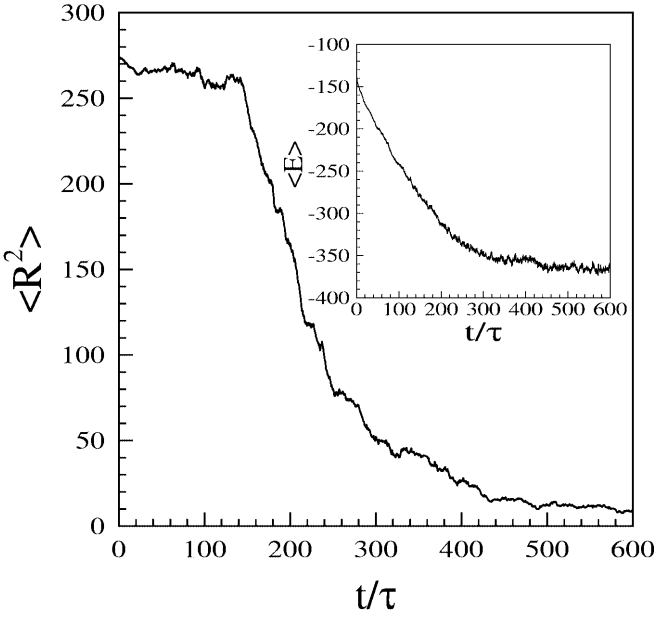

Fig. 1. Variation in the mean square radius of gyration and the energy (*inset*) during the folding of a model homopolymer obtained from Brownian dynamics (*BD*) simulations as a function of reduced time (adopted from Ref. [63])

simulation has the advantage that it does not include the detailed description as all-atom simulations, thereby reproducing the qualitative features by the implicit representation. In the following, we report the construction of an energy landscape of a model HP-36 protein.

8 Model of HP-36

Motivated by the success of the single homopolymer models [63, 66], a simple model of the chicken villin headpiece (HP-36) protein is proposed and studied in detail by off-lattice the BD simulation method [71, 72]. A representative stable structure of this protein (obtained from the Protein Data Bank) is shown in Fig. 2. This particular protein is the smallest monomeric polypeptide characterized, consisting of only naturally occurring amino acids that autonomously fold into a unique and thermostable structure without disulfide bonds or ligand binding [73]. A simulation study of model protein folding was carried out by using the hydropathy scale of amino acids. The diverse interactions among the amino acid residues are categorized into three classes by introducing a simplified hydrophobic scale. The simulations incorporate all the six different inter- and intra-amino acid interactions.

HP-36 protein is modelled as a necklace of different kinds of beads. Each bead in the sequence represents the corresponding amino acid in the protein sequence. There are 36 beads in the chain, since the number of residues in the original protein sequence (MLSDEDFKAV FGMTRSAFAN LPLWKQQNLK KEKGLF) is 36.

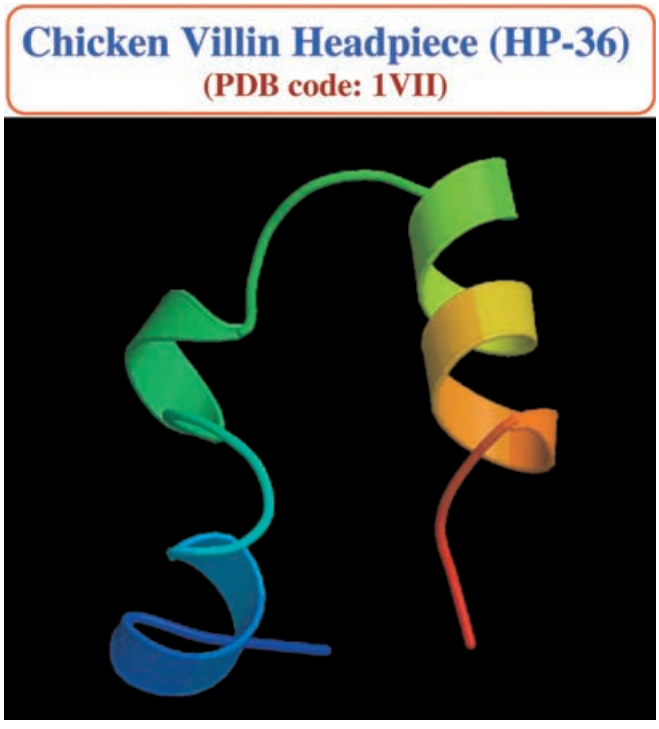

Fig. 2. One of the stable structures of thermostable chicken villin headpiece subdomain, a 36-residue (HP-36) protein (PDB code: 1VII)

All the beads are assumed to be of same mass and size. The beads in HP-36 interact via a site–site Lennard-Jones potential. Neighboring beads are connected via harmonic springs. The total potential energy of the chain is given by [71]

$$U = U_{\mathrm{b}} + U_{\mathrm{LJ}} + U_{\mathrm{s}} \ , \tag{8}$$

where U_{b} represents the bonding potential,

$$U_{\mathrm{b}} = \sum_{i=2}^{N} \kappa(|\ \mathbf{r}_i - \mathbf{r}_{i-1}\ |)^2 \ , \tag{9}$$

where κ represents the bond strength. The interaction between nonbonded beads is represented by the Lennard-Jones-like potential,

$$U_{\mathrm{LJ}}(r) = \epsilon_{i,j}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \ , \tag{10}$$

where σ is the Lennard-Jones collision diameter and $\epsilon_{i,j}$ represents the interaction strength. N is the number of beads in the chain, $\mathbf{r}_i$ is the position of bead i and $r_{ij} = |\ \mathbf{r}_i - \mathbf{r}_j\ |$. The stiffness is introduced through the bending potential U_{s},

$$U_{\mathrm{s}} = \mathscr{S}(\cos\theta - 1)^2 \ . \tag{11}$$

The unit of time, τ, is b^2/D_0, where the single bead diffusion coefficient is denoted by D_0. $\epsilon^* = \epsilon/k_{\mathrm{B}}T$, where $k_{\mathrm{B}}T$ is the thermal energy. The length is scaled by l, the bead diameter.

As discussed earlier, one of the major driving forces of protein folding in aqueous media is the hydrophobic/hydrophilic nature of amino acids. This can be best represented by the hydropathy scale [44, 74, 75]. This scale arranges the standard free energies of transfer from aqueous solutions to pure liquid hydrocarbons and provides a measure of hydrophobicity. In a certain sense, hydropathy scale provides a quantitative measure of the liking of a particular amino acid for water. Depending on the hydropathy values, all the amino acids present in the HP-36 sequence are categorized into three classes [71]: hydrophobic, weakly hydrophilic and strongly hydrophilic. The classification of the amino acids is presented in Table 1. The classification is done according to the following criterion. If the hydropathy value is positive, the amino acid is hydrophobic. On the other hand, among the hydrophilic amino acids (negative hydropathy value) if the hydropathy value is smaller than -2.5, it is strongly hydrophilic, otherwise it is weakly hydrophilic. A schematic representation of the hydrophobic scale is presented in Fig. 3a. A pictorial representation of the color code of the hydropathy values of both the original sequence and the simplified sequence due to the present categorization is shown in Fig. 3b.

Table 1. Classification of the amino acids constituting the HP-36 protein, according to the hydropathy values

Amino acid	Category
AFLMPV	Hydrophobic
GSTW	Weakly hydrophilic
DEKNQR	Strongly hydrophilic

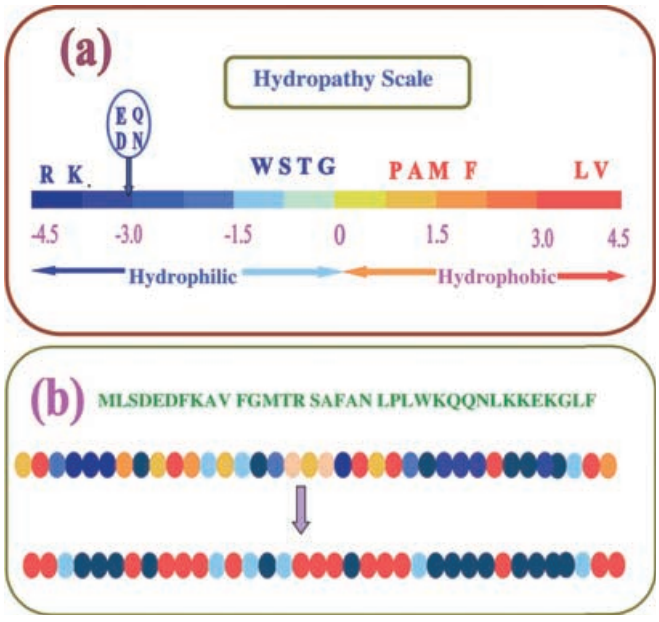

Fig. 3a, b. Schematic representation of modeling of the HP-36 protein (shown in Fig. 2) by using the hydropathy values. **a** Schematic representation of the hydropathy scale. The hydrophilic nature decreases from *blue* to *red*. **b** Pictorial representation of the color code of the hydropathy values of both the original sequence and the simplified sequence used in the present study (adopted from Ref. [71])

Note that the transfer of the hydropathy scale to the intermolecular potential is to be understood as a "solvent-averaged potential". This can also be considered as a potential of mean force, well known in colloids. Also, such a transfer of hydrophobicity to an interatomic potential was perhaps first done by Dill [46] in lattice simulations. The interaction strength parameter values for the six different interactions are listed in Table 2 . As can be seen from Table 2, interaction between two strongly hydrophilic groups is least favored because water will shield them, while that between two hydrophobic groups is strongly attractive. Thus, these potentials are all water-averaged potential.

8.1 Folding study

For each trajectory, an initial configuration is selected from the Monte Carlo generated equilibrium configurations at $\epsilon^* = 0.1$. The temperature of the initial configuration is then instantaneously reduced by 0.1ϵ, after 2.5×10^5 BD steps. Five such quenches, each with a gap of 2.5×10^5 step, are incorporated to facilitate the folding. Further simulations for 2.5×10^6 BD steps are carried out (subsequent to the quenching) to obtain the final configuration. Such a procedure is repeated for the model proteins with 1000 different configurations.

At each time step, the time-dependent total energy, the root-mean-square end-to-end distance, R^2, and the radius of gyration, R_g, were all monitored to follow the progress of the folding transition. The results presented here are the average over 600 trajectories with different initial configurations. More details on the simulation scheme can be found in a similar study on homopolymers [63, 66].

8.2 Unfolding study

It is well known that a protein can be denatured or unfolded from its native state by adding salts (e.g. guanadenium chloride) or chemical agents (e.g. urea). This is sometimes called cold denaturation. It is believed that these agents modify the interactions of water at the protein–water interface. Recent computer simulation studies [76] seem to suggest that the role of urea is to provide an energetically favorable environment of the hydrophobic groups in water. This provides the required driving force for unfolding. In the context of the present hydropathy scale, an aqueous solution containing urea makes the hydrophobic groups less attractive to each other. Therefore, to motivate the unfolding of the folded protein, the interaction among the different residues (polymer beads) is changed to reflect the altered scenario in the presence of urea in solution. This gives rise to a nice unfolding of the folded state, whose dynamics are described later. The modified interaction parameters for this case are listed in Table 3.

9 Folding features of HP-36

Typical snapshots of model HP-36 folding are shown in Fig. 4 as observed in BD simulations. The central picture corresponds to the initial configuration, while the four other structures represent configurations with different energies. In the initial configuration the hydrophobic and hydrophilic groups are at arbitrary locations (as in the unfolded HP-36); however, the final configurations show that the model protein folds into a compact minimum-energy structure by forming a hydrophobic

Table 2. Interaction parameter, $\epsilon_{i,j}$, value for the six different interactions in the folding of model HP-36 protein

Nature of the interaction	$\epsilon_{i,j}$
Hydrophobic–hydrophobic	2.0ϵ
Weakly hydrophilic–weakly hydrophilic	0.3ϵ
Strongly hydrophilic–strongly hydrophilic	0.3ϵ
Hydrophobic–weakly hydrophilic	1.0ϵ
Hydrophobic–strongly hydrophilic	0.8ϵ
Strongly hydrophilic–weakly hydrophilic	0.3ϵ

Table 3. $\epsilon_{i,j}$ value for the six different interactions in the unfolding of model HP-36 protein

Nature of the interaction	$\epsilon_{i,j}$
Hydrophobic–hydrophobic	0.3ϵ
Weakly hydrophilic–weakly hydrophilic	1.0ϵ
Strongly hydrophilic–strongly hydrophilic	1.0ϵ
Hydrophobic–weakly hydrophilic	0.8ϵ
Hydrophobic–strongly hydrophilic	0.8ϵ
Strongly hydrophilic–weakly hydrophilic	1.0ϵ

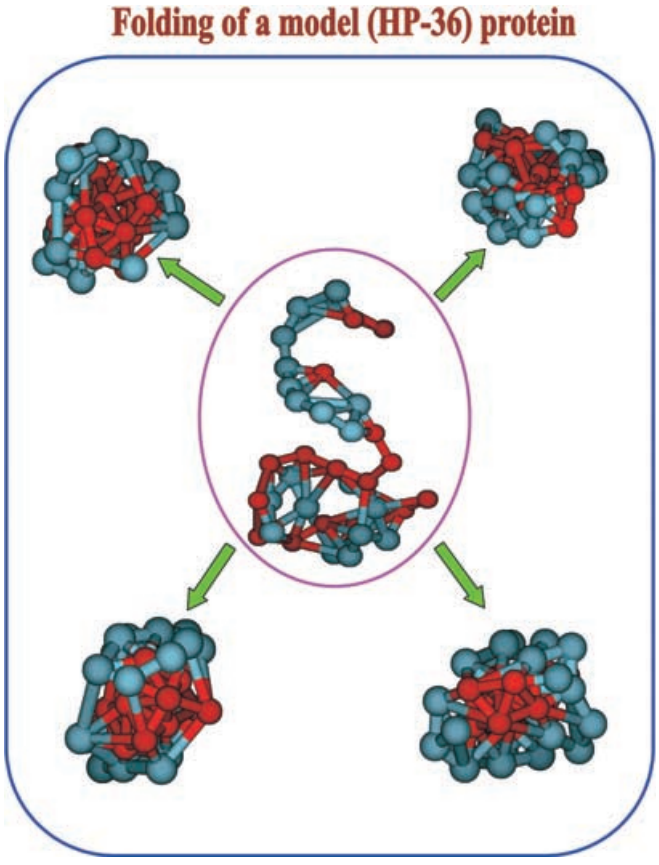

Fig. 4. Snapshots of a few conformations of model HP-36 as observed in BD simulations. The configuration in the *center* corresponds to the initial configuration, while the rest of the configurations represent the different minimum-energy configurations. Note the formation of a hydrophobic core within the hydrophilic outer surface in all the final configurations. (adopted from Ref. [71])

core within the hydrophilic outer surface. These look surprisingly close to the schematic figures given by Dill [5] representing the folding of a model protein (Fig. 7 of Ref. [5]). As emphasized by Dill [5], the main difference between the collapse of a protein and a homopolymer is that while the former is accompanied by an increase in chain compactness, the later involves an additional reconfiguration of the hydrophobic and polar residues to reach a compact native state. It is important to note that Dill's own simulations were based on lattice models. Hence, this picture was not possible to reproduce so vividly. From Fig. 4 it is evident that the simplified model used in our study represents the complex folding aspect, at least qualitatively.

9.1 Characterization of nativelike state

A folding/collapse study of 600 extended configurations has been carried out. Among these, only six (i.e. 1%) of the configurations are found to have attained a native-like conformation. The native-like configurations are characterized by the energy criterion and the number of specific hydrophobic contact pairs formed. The minimum-energy criterion employed is to consider those configurations which have an energy below a certain low-energy cutoff (-65ϵ) – the lowest-energy state (with 20 hydrophobic contacts) has an energy of -72ϵ. This cutoff itself is determined by the lowest-energy state obtained in our simulations. Thus, the five other states selected as nativelike states are those close in energy to the lowest-energy configuration. The second criterion is based on the number of hydrophobic contacts – the lowest-energy native state is also found to have the largest number (20) of hydrophobic contacts. The nativelike states are chosen to have at least 15 contacts.

The stability of the native configuration is verified by carrying out further simulations of the native state for the 5×10^6 steps. The energy is shown as a function of the reduced time in Fig. 5a, while that of the radius of gyration is shown in Fig. 5b. This figure reveals the stability of the nativelike conformation obtained in the presented study.

9.2 Construction of the funnel

In a simulation study, the energy landscape can be obtained by distributing the energies corresponding to

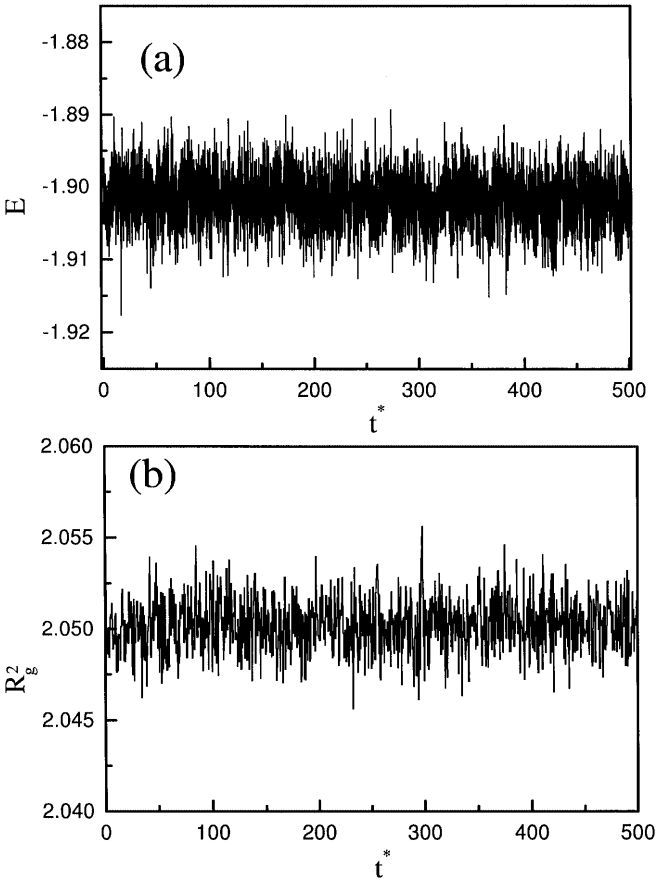

Fig. 5. Fluctuations in **a** the energy and **b** the radius of gyration of the native state as a function of reduced time. This figure demonstrates the stability of the native configuration obtained from BD simulations

the various final states obtained in simulations. The evolution of a funnel-like energy landscape, as observed in a simulation study, is shown in Fig. 6. The configurations shown in the figure correspond to the native state, misfolded, unfolded and metastable states (described in the figure). Most of the folded configurations are found to have the hydrophobic core (red) surrounded by the hydrophilic outer surface (blue). This figure demonstrates the ability of a minimalist model (HP-36) to obtain the qualitative features of protein folding. The funnel-like picture appearing in Fig. 6 is actually a three-dimensional plot (or histogram) – the number of hydrophobic contacts is shown on the figure. In this figure the minima are characterized by the larger number of hydrophobic contacts than the maxima which are characterized by the smaller number of hydrophobic contacts. As already mentioned, the funnel is rugged, which is reflected in the time evolution of the radius of gyration and the energy of the polymer subsequent to the quenching (described later).

9.3 Pair contacts

9.3.1 Equilibrium distribution of contacts

Subsequent to the temperature quenching, a fully extended initial configuration gradually folds into a compact rigid structure by forming the native contacts. The probability distribution that a specific (i, j) pair contact is separated by a distance R is defined as [77]

$$G_{i,j}(R) = \left\langle \delta(R - R_{ij}) \right\rangle \ , \tag{12}$$

where R_{ij} is the distance between the hydrophobic residues i and j in the final collapsed state. In Eq. (12), averaging is defined by

$$\langle x \rangle = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} x^n \ , \tag{13}$$

where $\mathcal{N}$ is the total number of configurations which include the subsets of either nativelike configurations or configurations that are far from the native state or all the configurations.

The equilibrium pair time correlations, $G_{ij}(R)$, for various pairs of hydrophobic residues are plotted in Fig. 7. Figure 7b shows the enlarged version of the first peak of the pair correlation functions shown in Fig. 7a. The distribution of the contacts with respect to bead 1 ($i = 1$) is plotted. The notation of pair correlations between the various residues is indicated in the figure. As can be seen from the figure, the pair correlation functions provide considerable insight into the pair dynamics. As shown in Fig. 7b the largest peak corresponds to the contact pair 9–12, while the pair 1–16 shows the

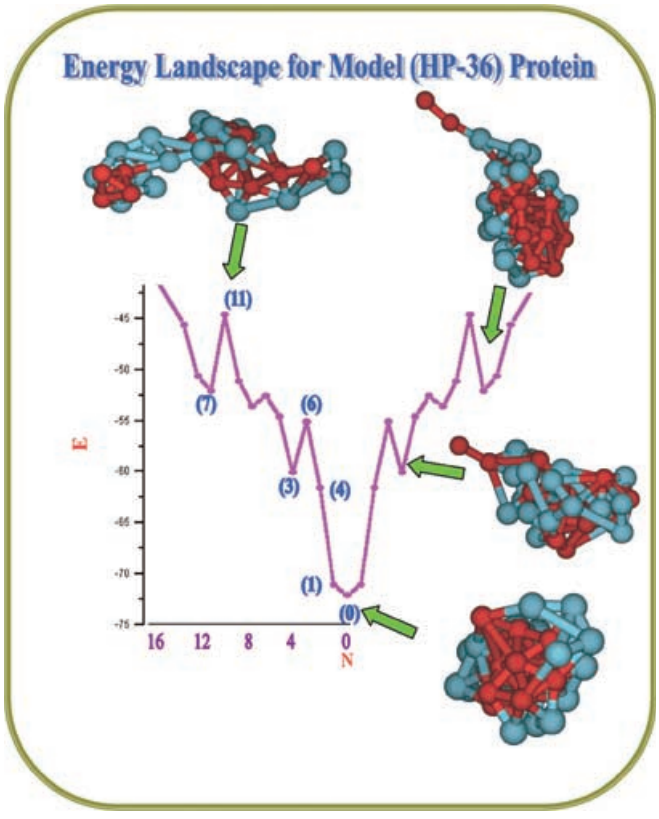

Fig. 6. Energy landscape (the funnel) for the model HP-36 protein obtained from BD simulations. The configurations corresponding to various energy states (unfolded, transition and native state) are also shown. The x-axis denotes the number of configurations with energy E (adopted from Ref. [71])

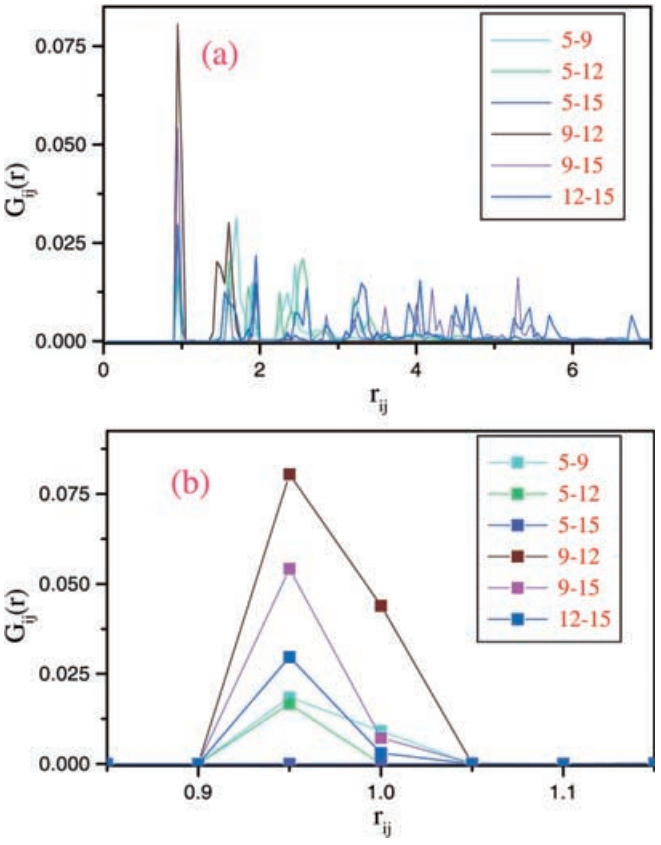

Fig. 7a, b. Pair distribution functions between the various hydrophobic residues. **a** Full $G_{ij}(r)$, **b** the closeup of the first peak in $G_{ij}(r)$

smallest peak. From the HP-36 protein sequence (Fig. 3a) it can be seen that among the hydrophobic residues studied, 9 and 12 are separated by just three monomers, while the hydrophobic residues 1 and 16 are separated by 34 monomers. In other words, the pair 9–12 involves local interaction, while the interaction between 1–16 is nonlocal. The rest of the peaks shown in the figure (9–15, 12–15, 5–9, 5–12 and 5–15) have the peak heights proportional to the respective sequence separation in the HP-36 sequence.

9.3.2 Dynamics of contact-pair formation

The unfolded protein obviously folds by forming pair contacts between the specific beads. The dynamics of tact pair formation provide a better window to study the protein folding; however, the dynamics are much more involved. For the present purpose, it is sufficient to monitor the distance between the specific pairs as a function of time. The study of the pair correlations reveals information on the contact formation between the specific hydrophobic residues. The dynamics of each of these pairs are studied in detail. The contact-pair time correlation function is defined as [77]

$$C_{i,j}(t) = \frac{\langle R_{i,j}(t)\rangle - \langle R_{i,j}(\infty)\rangle}{\langle R_{i,j}(0)\rangle - \langle R_{i,j}(\infty)\rangle} \quad . \tag{14}$$

The dynamics are followed by studying the folding from time $t = 0$ to $t = \infty$, the time required for all the temperature quenches. The pair correlation, $C_{i,j}(t)$, is shown in Fig. 8 as a function of the reduced time for all the previous mentioned contacts. The average trend of the approach to the native state is shown by symbols, while that for the individual pair is shown by full lines. The representation of the curves is shown in the figure. From this figure it is clear that the pair 9–12 is the fastest to attain its final position in the native conformation, while the pair 1–16 is the slowest. The rest of the pairs (9–15, 12–15, 5–9, 5–12 and 5–15) show a similar trend to that shown in Fig. 7. Close observation of pair 9–12 in the HP-36 sequence (Fig. 3) shows that these two hydrophobic beads are surrounded by hydrophobic beads on either side. Moreover, the hydrophobic blocks containing 9 and 12 are separated by just one hydrophilic bead. The fast contact formation of the 9–12 pair reveals the importance of the sequential arrangement of the amino acid residues in the protein in determining the folding rate. This is clearer from Fig. 9, where the folding rate (inverse of the folding time) is plotted as a function of the sequence separation (in terms of monomers). The folding time of a pair of hydrophobic residues shows a strong dependence on the sequence separation. Note that at short separations the folding

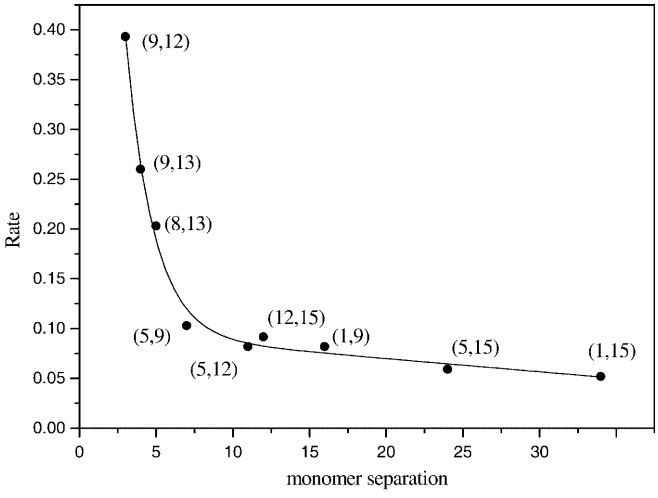

Fig. 9. Rate of native contact formation between the hydrophobic residues as a function of monomer separation. The hydrophobic beads corresponding to a given monomer separation are indicated

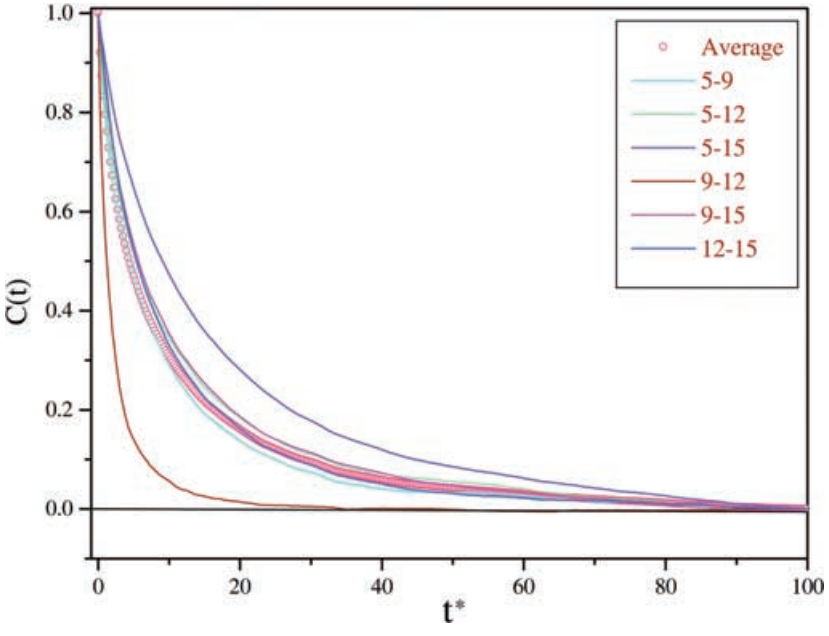

Fig. 8. Pair correlation function $C_{i,j}(t)$ defined by Eq. (14) for various pairs of hydrophobic residues as a function of reduced time. The pair with the smallest sequence separation (9–12) in the HP-36 folds faster (*brown*), while the pair with the longest separation (1–16) is the slowest (*red*) to reach the native contact. The *circles* represent the result obtained by averaging over all seven contact pairs

rate shows an exponential decay with increasing monomer separation. This reveals that the initial part of the folding is dominated by the local interactions, while the nonlocal interactions come into play in the later stages. This reveals that the protein folding involves two stages. First, the formation of the pair contacts between the residues owing to the local interactions. Second, the pair contacts mediated by the nonlocal interactions. However, folding continues for a relatively long time even after the formation of the (local) pair contacts. Thus it is clear that the protein folding involves the formation of the local contacts, which is a fast and perhaps the first step, and the formation of the nonlocal contacts, which is a relatively slow step. The rate-determining step in the protein folding is the formation of the nonlocal contacts. This is in accordance with the study of the Grantcharova et al. [33].

The average $C_{i,j}(t)$ is shown in Fig. 10 as a function of the reduced time in two very different situations. The full line shows the result obtained by averaging over five nativelike configurations selected in the vicinity of the native state. In the same figure, the result obtained by averaging over five configurations that are far from the native state is shown by the dashed line. While the average $C_{i,j}(t)$ is found to show an exponential-like decay (not shown), $C_{i,j}(t)$ in these two cases exhibits very different behavior. This comparison reveals that the formation of nativelike configurations results from the fast formation of the initial contacts (as already described, these are the local contacts) followed by the slow (nonlocal) contacts, which results in a long-time tail in $C_{i,j}(t)$. On the other hand, the folding of the protein into the energetically unfavoured configurations (far from the native state) results from the formation of the arbitrary pair contacts, resulting in a monotonic decay of $C_{i,j}(t)$, which resembles the collapse of a simple homopolymer [72]. Thus, the formation of the "correct" pair contacts plays a vital role in the protein folding, which is in accordance with the study of Zwanzig [25].

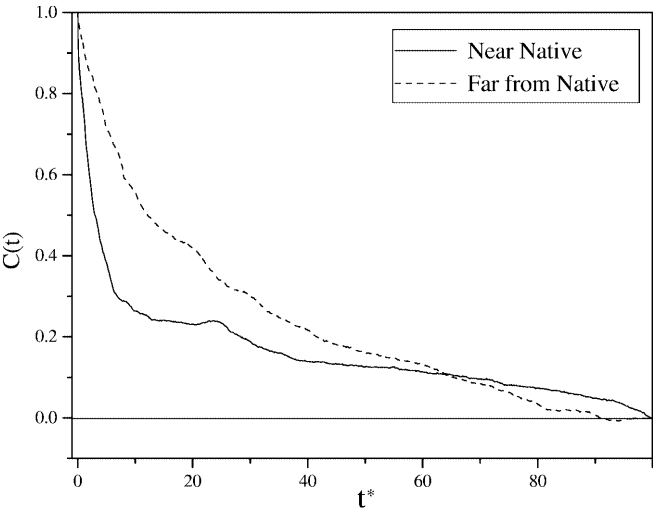

Fig. 10. Pair correlation function calculated by averaging over five configuration near the native state (*full line*) and that for five configurations far from the native state (*dashed line*). The native configuration results from the initial fast contact formation followed by slower rearrangement. On the other hand, the energetically unfavorable configurations originate owing to the slow monotonic contacts, as in case of the collapse of simple homopolymers

9.3.3 Dynamics near the native state

The equilibrium and the dynamic approach to the native conformation are described here. $C_{i,j}(t)$ for the six different contacts are plotted in Fig. 11 as a function of the reduced time. The representation of this figure is same as that in Fig. 10. This figure reveals that most of the pairs (except for pair 5–15, which is 25 residues apart), show a faster approach towards the native state.

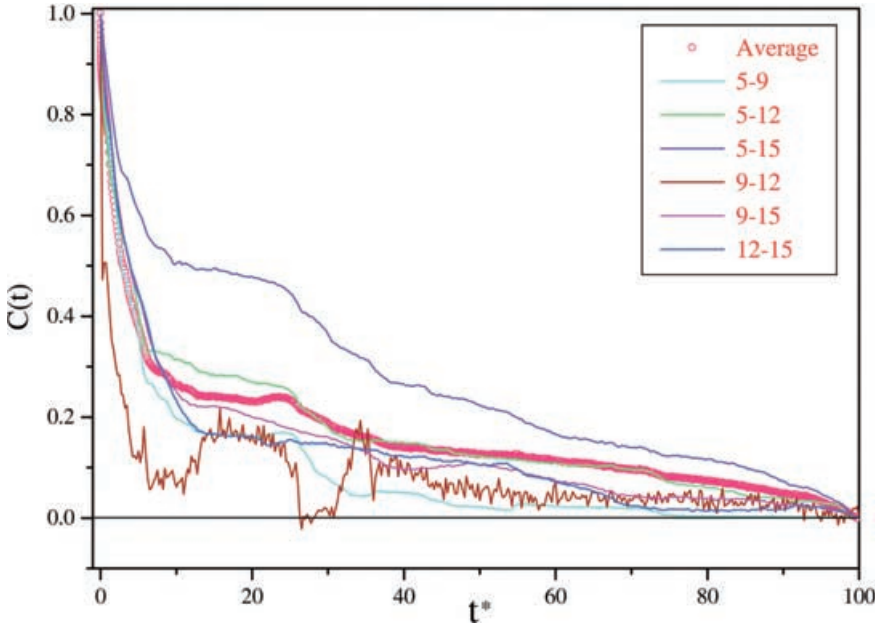

Fig. 11. Pair correlation functions for the individual contacts for the native configuration. As shown, most of the contacts form very fast and the native state is attained by the readjustment of these initial contacts

9.4 Dynamics of folding and unfolding

The native configuration (corresponding to the minimum in the energy funnel) is chosen for the unfolding study. In the study of unfolding, the interaction energies among the various amino acids are changed instantaneously at time $t = 0$. Snapshots of the unfolding of HP-36 protein found in BD simulations are shown in Fig. 12. As shown in this figure, the completely folded initial configuration gradually unfolds by breaking the native contacts to reach the fully extended state. To emphasize this point, the number of topological contacts is plotted in Fig. 13 as a function of time. For comparison, the variation in N_{topo} during the folding is also shown in the inset. As shown in main figure, N_{topo} is a maximum at $t = 0$, which corresponds to the native state. In the unfolding case, the variation in number of topological contacts (main figure) shows the opposite trend to that of the folding (inset). On the whole, N_{topo} decreases during the unfolding, while it increases in the case of folding. This figure demonstrates the role of topological contacts and the importance of the hydrophobic forces in protein folding/unfolding.

The increase in energy during unfolding is shown in Fig. 14. In the inset the same is shown for the mean square radius of gyration. It is interesting to note the oscillatory dynamics, also recorded for topological contact formation in Fig. 13. Such oscillations seem to indicate that the polymer, when unfolding, faces a barrier. It is observed that the oscillations in the energy and in the mean square radius are larger for folding. Of course these oscillations may be very much dependent on the effective potentials used in the simulation model, but they indicate the presence of barriers along folding and unfolding, except that there seem to be more barriers during folding and also the pathway seems to be more complicated. In particular, the dynamics during folding show considerably more oscillations in the final stage of folding which is relatively smooth for unfolding.

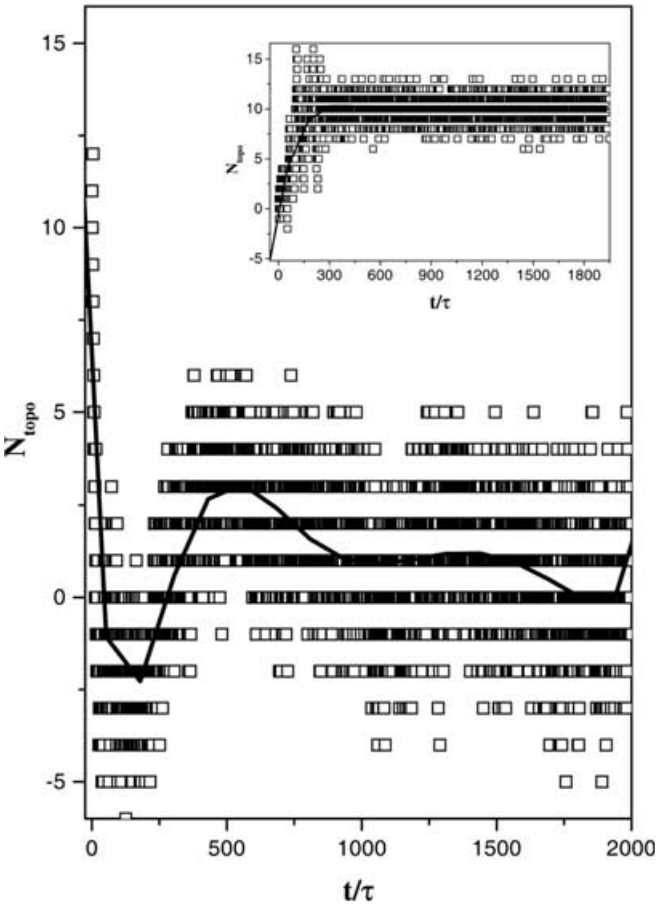

Fig. 13. Variation in the number of hydrophobic topological contacts as a function of time showing the result for unfolding. The *inset* represents that for the folding. During the unfolding, N_{topo} shows the opposite trend to that observed in the case of folding

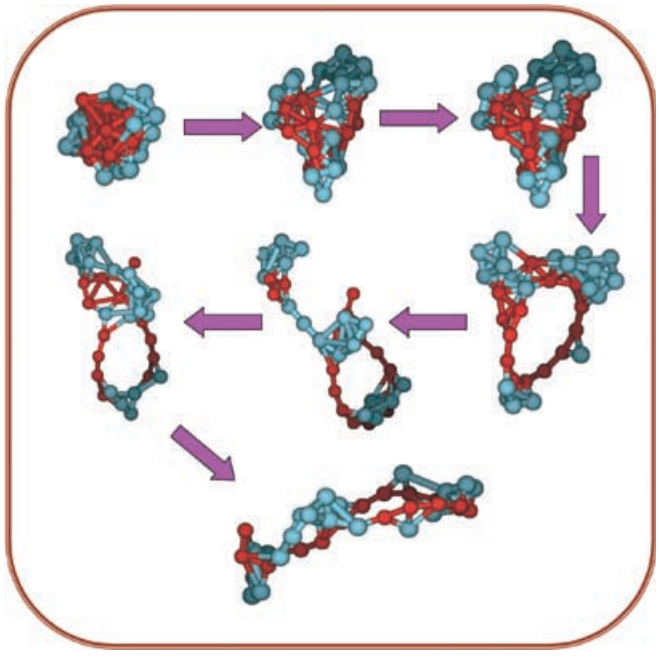

Fig. 12. Snapshots of the unfolding of model HP-36 protein as observed in BD simulations

9.5 Protein folding funnel versus that of RNA/DNA folding

The folding of DNA/RNA can be considerably different from that of the proteins. Folding of DNA hairpins has recently been studied in some detail as the hairpins are the commonly encountered structural intermediates in DNA/RNA folding [78–82]. For the DNA hairpins, the free-energy landscape is rough at the outer most edges. This is a consequence of the fact that in this case, prior to the hairpin formation, there is a large probability for the misfolding. Once a correct "native contact" is formed, the rest of the chain folds by instantaneous "zipping". This leads to a smooth and downhill pathway in the later stages. On the other hand, for the proteins, in the initial stages the formation of "nativelike" contacts is energetically favored. In the later stages, the roughness of the energy surface increases owing to competition between

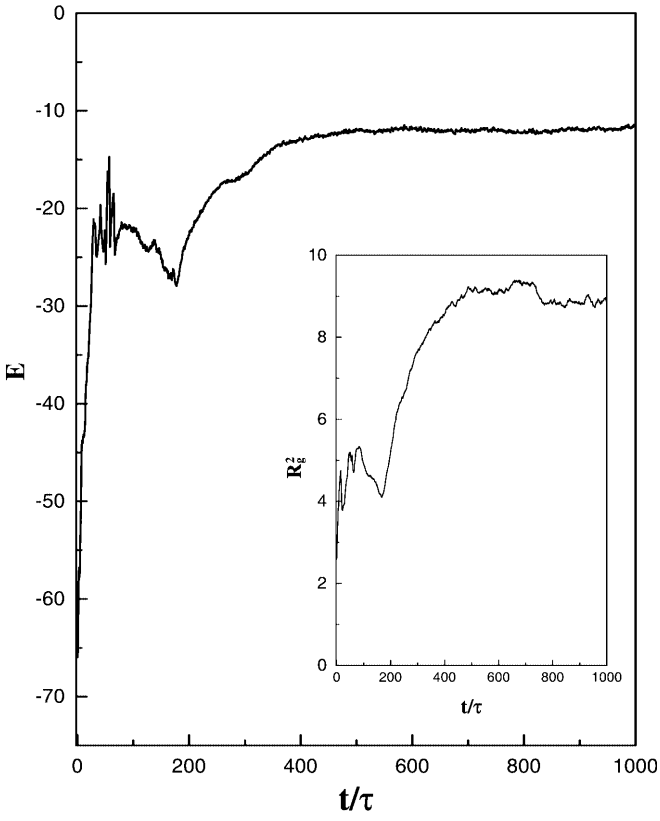

Fig. 14. Time dependence of energy and mean square radius of gyration during the unfolding of a model HP-36 protein obtained from BD simulations

the energy and structural entropy. Moreover, in this case, the formation of a few "native contacts" does not always result in the formation of a native conformation. The cost of making a wrong contact increases as the folding proceeds. This leads to an energy-landscape picture with increasing roughness as the protein approaches the native state. The folding scenario of proteins and DNA and RNA thus appears to be rather different.

10 Conclusion

The study of the pair correlations during the folding revealed that the folding of a protein involves two stages: fast local contact formation which is followed by the slower nonlocal contacts. The protein folding rate is mainly controlled by the latter contacts. This is further confirmed by comparing the $C_{i,j}(t)$ for the states near and far from the native states.

Given the complexity of the real protein folding problem, the ability of such a minimalist model (as used here and also elsewhere) to capture many of the essential features of protein folding is quite encouraging. The present study suggests that it may be possible to obtain semiquantitative information on the folding mechanism, folding rates and also the stability by modeling more complex proteins in a similar way.

Studies exploring the sensitivity to the potential employed and also on generalizing the set of potentials to accommodate more realistic potentials will reveal more information. Also, the study of the formation of the specific contacts that characterize the native state may provide deeper insight. The present studies seem to suggest that a simple minimalistic model as adopted here should be generalized at least to include the difference in size of the amino acid residues.

There are still many unsolved or poorly understood theoretical problems. For example, the relationship between helix formation and the rigidity of the heteropolymer chain has not been fully explored. A homopolymer chain with a simple Lennard–Jones type interaction always seems to have a toroidal and not a helical shape as the most stable state. Clearly, specific interactions among the amino acid residues play an important role in stabilizing the helix. In addition, the rigidity of the backbone also needs to be considered. Most of the theoretical studies have not included even such simple factors as the variation of the size of the amino acids and the hydrophobicity along the chain in any realistic fashion. Even from a polymer theory perspective, these are interesting problems worth further investigation. In addition, it is now known that in those proteins which are involved in enzymatic activity, certain hydrophobic patches remain exposed to water, such as the tryptophan, methionine and phenylalanine in subtilisin carlsberg (PDB code 1SBC). How such arrangements are stabilized, despite apparently strong bias against the hydrophobic groups on the surface, remains to be understood.

On the experimental side, a lot more needs to be learned from dynamic studies. For example, it would be nice to have flourescence resonance energy transfer experiments done during folding, by using, for example, single-molecule spectroscopic techniques. Even studies of denatured proteins can reveal information about the dynamics of conformational transition between different states of proteins.

Acknowledgements. We thank Samir Pal and Arnab Mukherjee for many discussions and Arun Yethiraj for helpful suggestions. The financial support from DST, India, is gratefully acknowledged. G. S. thanks CSIR for a research fellowship.

References

1. Horwich AL, Weissman JS (1997) Cell 89: 499
2. Walsh DM, Klyubin I, Fadeeva JV, Cullen WK, Anwyl R, Wolfe MS, Rowan MJ, Selkoe DJ (2002) Nature 416: 535
3. Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson CM, Stefani M (2002) Nature 416: 507
4. Levitt M (1991) Curr Opin Struct Biol 1: 224
5. Dill KA (1990) Biochemistry 29: 7133
6. Skolnick J, Kolinski A (1989) Annu Rev Phys Chem 40: 207
7. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Annu Rev Phys Chem 48: 545
8. Brooks CL III, Gruebele M, Onuchic JN, Wolynes PG (1998) Proc Natl Acad Sci USA 95: 1037
9. Shakhnovich EI (1997) Curr Opin Struct Biol 7: 29
10. Anfinsen CB, Haser E, Seli M, White FH (1961) Proc Natl Acad Sci USA 47: 1309
11. Kim PS, Baldwin RL (1990) Annu Rev Biochem 59: 631 and references therein

12. (a) Zhou Y, Hall CK, Karplus M (1996) Phys Rev Lett 77: 2822; (b) Zhou Y, Karplus M, Wichert JM, Hall CK (1997) J Chem Phys 107: 10691
13. Shakhnovich EI, Gutin AM (1990) Nature 346: 773
14. Sali A, Shakhnovich EI, Karplus M (1994) Nature 369: 248
15. Dinner RA, Karplus M (1998) Nature Struct Biol 5: 236
16. Levinthal C: In: Debrunner P, Tsibris JCM, Munck E (eds) Proceedings of a Meeting held at Allerton House, Monticello, Il. Mossbauer spectroscopy in biological systems. University of Illinois Press, Urbana, p 22
17. Dawkins R (1987) The blind watchmaker. Norton, New York, pp 46–50
18. Zwanzig R, Szabo A, Bagchi B (1992) Proc Natl Acad Sci USA 89: 20
19. Leopold PE, Montal M, Onuchic JN (1992) Proc Natl Acad Sci USA 89: 8721
20. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND (1995) Proc Natl Acad Sci USA 92: 3626
21. Onuchik JN (1998) Proc Natl Acad Sci USA 94: 7129
22. Nymeyer H, Garcia AE, Onuchic JN (1998) Proc Natl Acad Sci USA 95: 5921
23. Schoemaker BA, Portman JJ, Wolynes PG (2000) Proc Natl Acad Sci USA 97: 8868
24. Goldstein RA, Luthey-Schulten Z, Wolynes PG (1992) Proc Natl Acad Sci USA 89: 8029
25. (a) Zwanzig R (1995) Proc Natl Acad Sci USA 92: 9801; (b) Zwanzig R (1997) Proc Natl Acad Sci USA 94: 148
26. (a) Bryngelson JD, Wolynes PG (1987) Proc Natl Acad Sci USA 84: 7524; (b) Bryngelson JD, Wolynes PG (1989) J Phys Chem 93: 6902
27. Dill KA, Alonso DOV, Hutchinson K (1989) Biochemistry 28: 5439
28. de Gennes PG (1990) Scaling concepts in polymer physics. Cornell University Press, Ithaca
29. Sali A, Karplus M, Shaknovich EI (1994) J Mol Biol 235: 1614
30. Wolynes PG (1997) Proc Natl Acad Sci USA 94: 6170
31. Schoemaker BA, Wang J, Wolynes PG (1997) Proc Natl Acad Sci USA 94: 777
32. Shakhnovich EI, Gutin M (1999) Proc Natl Acad Sci USA 90: 7195
33. Grantcharova V, Alm VJ, Baker D, Horwich AL (2001) Curr Opin Struct Biol 11: 70
34. Plaxco KW, Simons KT, Baker D (1998) J Mol Biol 277: 985
35. Laurents DV, Baldwin RL (1998) Biophys J 75: 428
36. Brockwell DJ, Smith DA, Radford SE (2000) Curr Opin Struct Biol 10: 16
37. Portman JJ, Takada S, Wolynes PG (1998) Phys Rev Lett 81: 5237
38. (a) Portman JJ, Takada S, Wolynes PG (2001) J Chem Phys 114: 5069; (b) Portman JJ, Takada S, Wolynes PG (2001) 114: 5082
39. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Proteins 21: 167
40. Hansmann UHE, Onuchic JN (2001) J Chem Phys 1601: 115
41. Socci ND, Onuchic JN, Wolynes PG (1996) J Chem Phys 104: 5860
42. Miller MA, Wales DJ (1999) J Chem Phys 111: 6610
43. Kauzmann W (1959) Adv Protein Chem 14: 1
44. Tanford C (1980) The hydrophobic effect, 2nd edn. Wiley, New York
45. (a) Lim WA, Sauer RT (1992) Biochemistry 219: 319; (b) Mathews BW (1993) Biochemistry 62: 139
46. Dill KA (1999) Protein Sci 8: 1166
47. Predki PF, Agarwal V, Brunger AT, Regan L (1996) Nature Struct Biol 3: 54
48. Kurodu Y, Hamada D, Tanaka T, Goto Y (1996) Folding Design 4: 255
49. Roy S, Ratnaswamy G, Boice JA, Fariman D, McLendon G, Hecht MH (1997) J Am Chem Soc 116: 5302
50. (a) Yue K, Dill KA (1993) Phys Rev E 48: 2267; (b) Chan HS, Dill KA (1990) Proc Natl Acad Sci USA 87: 6368
51. Skolnick J, Kolinski A (1990) Science 250: 1121
52. Kolinski A, Galazaka W, Skolnick J (1996) Proteins 26: 271
53. Pande VS, Rokhsar DS (1999) Proc Natl Acad Sci USA 96: 1273
54. Taketomi H, Ueda Y, Go N (1975) Int J Pept Protein Res 7: 445
55. (a) Honeycutt JD, Thirumalai D (1990) Proc Natl Acad Sci USA 87: 3526; (b) Honeycutt JD, Thirumalai D (1992) Biopolymers 32: 695
56. (a) Guo Z, Thirumalai D, Honeycutt JD (1992) J Chem Phys 97: 525; (b) Klimov D, Thirumlai D (1996) Proteins 26: 411
57. Makarov DE, Hansma PK, Metiu H (2001) J Chem Phys 114: 9663
58. Duan Y, Kollman PA (1998) Science 282: 740
59. Boczko EM, Brooks CL III (1997) Science 269: 393
60. Hansmann UHE, Masuya M, Okamato Y (1997) Proc Natl Acad Sci USA 94: 10652
61. Hansmann UHE, Okamato Y (1997) J Comput Chem 18: 920
62. Pande VS, Rokhsar DS (1999) Proc Natl Acad Sci USA 96: 9062
63. Srinivas G, Yethiraj A, Bagchi B (2001) J Phys Chem B 105: 2475
64. Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, Chemla DS, Dawson PE, Schultz PG, Weiss S (2000) Proc Natl Acad Sci USA 97: 5179
65. Weiss S (1999) Science 283: 676
66. Srinivas G, Bagchi B (2002) Phys Chem Commun 5: 59
67. Zhou Y, Hall CK, Karplus M (1996) Phys Rev Lett 77: 2822
68. Alm E, Baker D (1999) Proc Natl Acad Sci USA 96: 11305
69. Riddle DS, Gratcharova VP, Alm E, Ruczinski I, Baker D (1999) Nature Struct Biol 6: 1016
70. Plaxco KW, Guijarro JI, Morton CJ, Pitkeathly M, Campbell ID, Dobson CM (1998) Biochemistry 37: 2259
71. Srinivas G, Bagchi B (2002) J Chem Phys 116: 0000
72. Srinivas G, Bagchi B (2002) Curr Sci 82: 179
73. McKnight JC, Doering DS, Matsudaria PT, Kim PS (1996) J Mol Biol 260: 126
74. Stryer L (1995) Biochemistry, 4th edn. Freeman, New York
75. Nelson DL, Cox MM (1993) Lehninger principles of biochemistry, 3rd edn. Worth, New York
76. (a) Cheng Y-K, Rossky PJ (1998) Nature 392: 696; (b) Cheng Y-K, Sheu WS, Rossky PJ (1999) Biophys J 76: 1734
77. Srinivas G, Bagchi B (2002) J Phys Chem B (submitted)
78. Ansari A, Kuznetsov SV, Shen Y (2001) Proc Natl Acad Sci USA 98: 7771
79. Scheffler IE, Elson EL, Baldwin RL (1968) J Mol Biol 36: 291
80. Gralla J, Crothers DM (1973) J Mol Biol 73: 497
81. Bonnet G, Krichevsky O, Libchaber A (1998) Proc Natl Acad Sci USA 95: 8602
82. Wallace MI, Ying L, Balasubramanian S, Klenerman D (2000) J Phys Chem B 104: 11551