

Foldability and the funnel of HP-36 protein sequence: Use of hydropathy scale in protein folding

Goundla Srinivas and Biman Bagchi*

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

Brownian dynamics simulation study of the folding of a model thermostable chicken villin head piece subdomain, a 36-residue protein (HP-36), is carried out 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 intraamino acid interactions. The model protein reproduces some of the qualitative features of the complex protein folding, including the funnel-like energy landscape. Although there are several states near the minimum of the folding funnel, we could identify a stable native configuration. In addition, the study reveals a correlation between the contact order, topology, and the stability.

I. INTRODUCTION

The study of protein folding has been a subject of great importance over the past several decades.^{1–20} Protein folding is a complex problem. This is not only due to the frustration that results from the diverse interactions among the constituting amino acids but also to the fact that the folded native configuration depends both on sequence and the amino acid content. Numerous theoretical studies have suggested that the size, stability, and the topology of a protein influence the folding rate and mechanisms.^{2,7} The recently emerging concept in this field is the free energy landscape guided folding.²¹

Many of the theoretical studies have been directed to single domain small proteins.^{2,18} For example, the early statistical mechanical theories by Dill and co-workers¹⁹ and by Bryngelson and Wolynes² were based on the idea of heteropolymer collapse and reordering among hydrophobic and hydrophilic residues. They provide a two order parameter model for protein folding, namely η , the generalized packing fraction and ρ , the fraction of amino acid residues in the native state. This model predicts a first order transition from globule to coil when a protein molecule is denatured by temperature variation. In terms of η and ρ the free energy function for the collapse transition can be obtained as²

$$\frac{F}{NT} = -\frac{1+\rho^2}{T} \eta + \rho \log \rho + (1-\rho) \log \frac{(1-\rho)}{N} + 1 + \left(\frac{1}{\eta} - 1\right) \log(1-\eta) + \frac{3}{2} N^{-4/3} \eta^{-2/3} + \frac{2}{3N} \log \eta. \quad (1)$$

The theories of Dill and of Bryngelson and Wolynes 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.^{3–5,11–14,21–24} For example Zwanzig *et al.*⁴ showed

that a small energy bias (on the order of a few $k_B T$) against the locally unfavorable configurations can reduce the Levinthal's time to a biologically significant size. Later Zwanzig⁵ presented a simple model of protein folding kinetics based on the "correctness" of the folded structure and found that the folding time has a maximum near the folding transition temperature and a minimum at a lower temperature. More recently Wolynes and co-workers^{25,26} have presented a 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 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 structure.

On the other hand, from the numerical and analytical studies of various models Wolynes,¹¹ Onuchic,²¹ and others^{27–31} introduced and elaborated on the concept of energy landscape. According to this latter development, the folding kinetics is determined by an energy landscape and for foldable proteins this resemble a funnel with a free energy gradient toward the native structure. The introduction of the concept of the folding funnel provided a much needed breakthrough in understanding the pathways of protein folding.

Progress has also been made over the past few years in linking the experimental and theoretical approaches to protein folding.^{7,22–24,31–33} The main outcome of these studies is that the topology is a very significant determinant of protein folding rates. A newly emerging concept in protein folding is contact order parameter; average sequence separation between the contacting residues. Baker and co-workers²² observed a correlation between the folding rate and the contact order. Later studies found an improved correlation between the folding rate and the relative contact order; the average sequence distance between all pairs of contacting residues normalized by the total sequence length

*Author to whom correspondence should be addressed; electronic mail: bbagchi@sscu.iisc.ernet.in

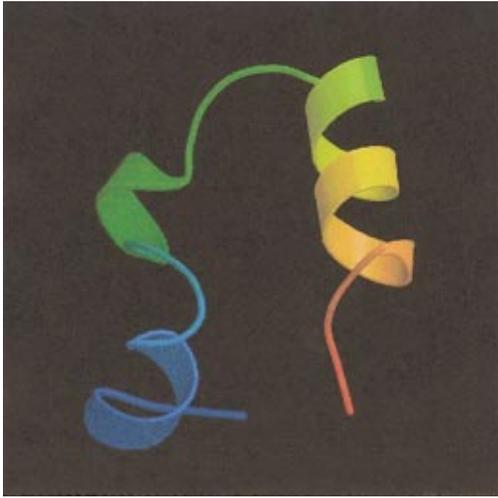


FIG. 1. (Color) One of the stable structures of thermostable chicken villin headpiece subdomain, a 36-residue (HP-36) protein (PDB code:1VII).

$$CO = \frac{1}{LN} \sum^L \Delta S_{i,j}, \quad (2)$$

where $\Delta S_{i,j}$ is the sequence separation in residues, between the contacting residues i and j and N is the total number of residues in the protein. Note that the summation is over all the contacts L .

Earlier simulation studies attempted to describe the protein folding through the lattice model: the simplest possible model to study the complex protein folding problem.¹⁹ Later the use of lattice-based approaches pioneered the prediction of the protein tertiary structure.^{34,35} The approach to the native state in lattice model simulations is driven by the energy due to the 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 smoother folding path is obtained by incorporating the Go-like potential.³⁶ For example, Gutin *et al.*³⁷ obtained a multidimensional energy landscape by carrying out the folding study of lattice models.

Subsequent development found success in the study of model proteins. For example, the off-lattice Langevin dynamic simulation studies^{16,38} proved more promising in studying the dynamics of protein folding. Recent studies of this class demonstrated the folding of a model protein into a β -barrel structure. These studies suggested the existence of several metastable minima in which the folded forms of protein have similar structural characteristics but different energies. A stochastic kinetic model for titin unfolding has been presented and studied by kinetic Monte Carlo simulations.³⁹ By incorporating the realistic bond angles of amino acids and dihedral angles, Levitt^{40,41} pioneered the off-lattice simulations to closely monitor the protein folding. Recently Clementi *et al.*⁴² and Clementi and Onuchic⁴³ demonstrated the role of the native state topology in determining the folding mechanism by studying small globular proteins.

While the lattice and off-lattice simulations study the model proteins, all-atom simulations explicitly consider the

interaction between all the atoms present in real protein.^{8,44–46} 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 funnel concept.⁴⁵ Notable among the all-atom simulations is the study of Kollman and Duan,⁸ who carried out the first ever 1 μ s simulation of a protein in aqueous solution. They have studied the thermostable chicken villin head piece subdomain, a 36-residue protein (commonly known as HP-36 protein) in the aqueous solution by explicit representation of water molecules. They found a native-like structure with two pathways.

In this work, we study the HP-36 protein using the hydrophathy value of the constituting amino acids. The complex interactions among and between the various amino acids are simplified by categorizing them into three different classes. As a result the HP-36 protein is represented as a necklace of three different kinds of beads with a total of six different interactions. The chain stiffness is also included in the chain Hamiltonian.

By studying this minimalist model, we find that the model HP-36 protein shows some of the qualitative features of folding that are common to the real proteins. (a) It folds into various collapsed and ordered states. (b) It shows a funnel-like energy landscape. (c) It exhibits a correlation between the topology and the stability and (d) distribution of energies and the topological contacts supported the funnel-like energy landscape picture.

The rest of the paper is organized as follows. In Sec. II we introduce a simplified model for the HP-36 protein. In Sec. III, the simulation details are described. Section IV present the details of calculating the probability distribution of energy and topological contacts. The results and discussion is presented in Sec. V. We close this paper with a few conclusions in Sec. VI.

II. MODEL OF HP-36: MAPPING FROM THE HYDROPATHY SCALE INTO INTER-AMINO ACID INTERACTIONS

We have studied protein folding by modeling a thermostable 35-residue subdomain within chicken villin headpiece HP-36. A representative stable structure of this protein (obtained from the protein data bank) is shown in Fig. 1. The choice of this particular protein comes from the fact that it 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. This particular protein sequence was first reported by McKnight *et al.*⁴⁷

In this study, the HP-36 protein is modeled as a necklace

TABLE I. Classification of the amino acids constituting the HP-36 protein, according to the hydrophathy values.

Amino acid	Category
AFLMPV	hydrophobic
GSTW	weakly hydrophilic
DEKNQR	strongly hydrophilic

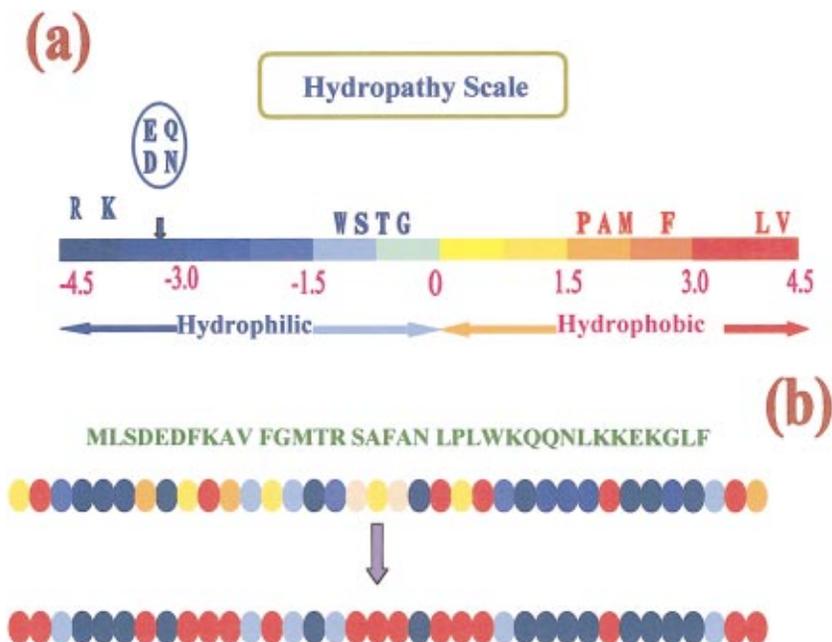


FIG. 2. (Color) Schematic representation of modeling of the HP-36 protein (shown in Fig. 1) by using the hydrophathy values. (a) A schematic representation of the hydrophathy scale. The hydrophilic nature decreases from blue to red. (b) A pictorial representation of the color code of the hydrophathy values of both the original sequence and the simplified sequence used in the present study.

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 LPLWKQQLK KEKGLF) are 36. All the beads are assumed to be of the same mass and size.

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 hydrophathy scale.^{48,49} Depending on the hydrophathy values we have categorized all the amino acids present in the HP-36 sequence into three classes; (i) hydrophobic, (ii) weakly hydrophilic, and (iii) strongly hydrophilic. In Table I the classification of amino acids is presented. The classification is done according to the following criterion. If the hydrophathy value is positive, the amino acid is hydrophobic. On the other hand, among the hydrophilic amino acids (hydrophathy value is negative) if the hydrophathy value is smaller than -2.5 , it is strongly hydrophilic, otherwise weakly hydrophilic. In Fig. 2(a) a schematic representation of the hydrophobic scale is presented. Figure 2(b) shows a pictorial representation of the color code of the hydrophathy values of both the original sequence and the simplified sequence due to the present categorization.

As mentioned above, the interaction between the amino acids and the polar water molecules plays a key role in protein folding. In the present study, the interaction strength has been implicitly incorporated within the interaction parameter ϵ_{ij} (to be described in Sec. III). Due to the above mentioned classification a total of six different kinds of interactions are possible. The following is a list of the interaction strength parameter values for all six different interactions: (i) hydrophobic-hydrophobic = 2ϵ ; (ii) strongly hydrophilic-strongly hydrophilic = 0.3ϵ ; (iii) weakly hydrophilic-weakly hydrophilic = 0.3ϵ ; (iv) strongly hydrophilic-hydrophobic = 0.8ϵ ; (v) weakly hydrophilic-hydrophobic = ϵ ; and (vi) strongly hydrophilic-weakly hydrophilic = 0.3ϵ .

III. SIMULATION DETAILS

As mentioned above, the HP-36 protein is modeled as a necklace of three different kinds of beads; the beads interact via a site-site Lennard-Jones (LJ) potential. Neighboring beads are connected via harmonic springs. Each bead represents an amino acid in the actual protein sequence. The total potential energy of the chain can be written as

$$U = U_b + U_{LJ} + U_s, \quad (3)$$

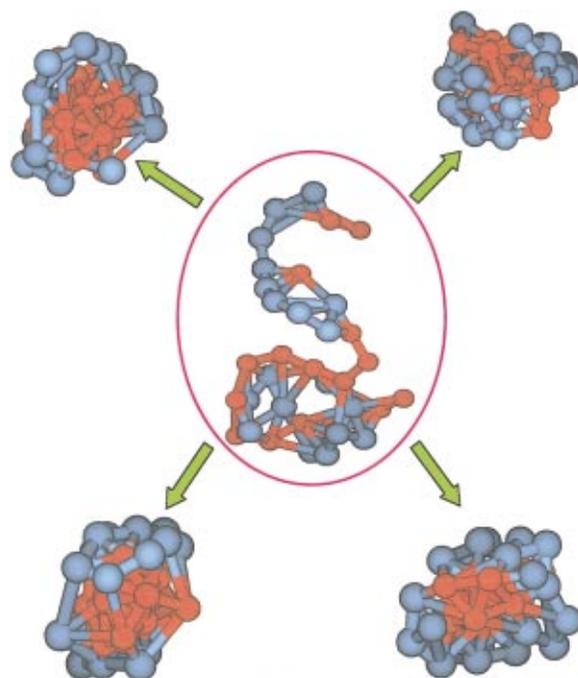


FIG. 3. (Color) Snapshots of various conformations of model HP-36 as observed in Brownian dynamics simulations. The configuration in the central box corresponds to the initial configuration, while the rest of the configurations represent the different minimum energy configurations. Note the formation of hydrophobic core in all the final configurations.

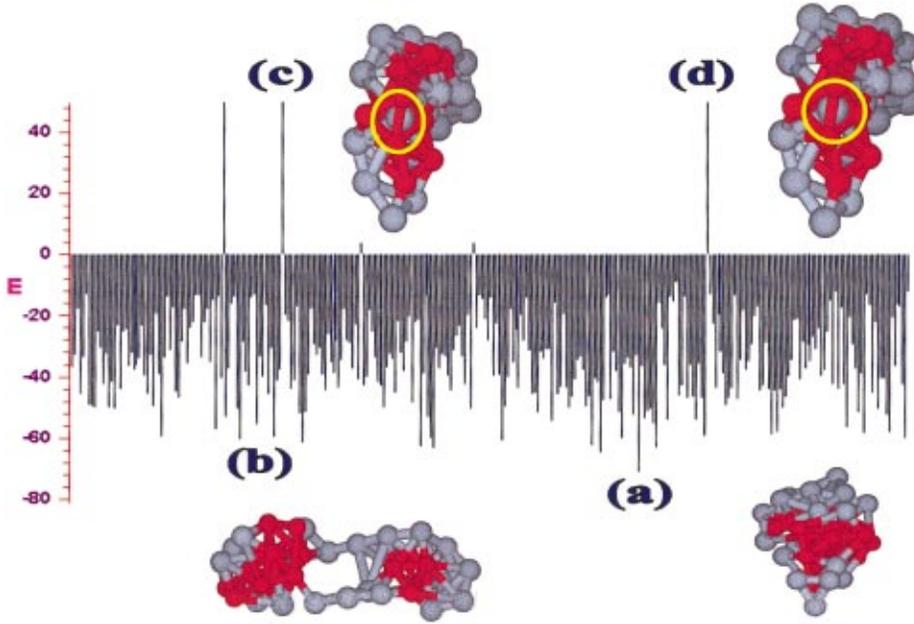


FIG. 4. (Color) Energies of various final energy configurations for the model HP-36 protein obtained from BD simulations is shown. Configurations correspond to the energy values (a), (b), (c) and (d). The hydrophilic bead within the hydrophobic core is highlighted by the yellow circles in structures (c) and (d).

where U_b represents the bonding potential

$$U_b = \sum_{i=2}^N b(|\mathbf{r}_i - \mathbf{r}_{i-1}|)^2, \quad (4)$$

we set $b=9$ in this study. The interaction between non-bonded beads is represented by the Lennard-Jones like potential

$$U_{LJ}(r) = \epsilon_{i,j} \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (5)$$

where σ is the LJ collision diameter and $\epsilon_{i,j}$ represents the interaction strength. N is the number of beads, \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

$$U_s = S (\cos \theta - 1)^2, \quad (6)$$

where N is the number of the beads in the chain $b = \sigma$, and the chain stiffness $S=1$. Note that the stiffness potential depends on the angle formed by three subsequent residues in sequence. However, in the present model we have not utilized the realistic values for the bond angles of the amino acids. This is because our goal is to study the spontaneous development of the native state from the initial unfolded state through hydrophobic collapse by a minimalist model. Such a model has been successfully used by Noguchi and Yoshikawa⁵⁰ to study the polymer collapse and variation in structural morphology. Nevertheless, the off-lattice simulations incorporating the realistic bond angles and including the dihedral terms reveal more information as already shown by Clementi *et al.*⁴² and Clementi and Onuchic⁴³ and also by Levitt.^{40,41}

The time evaluation of the model protein is done according to the following equation of motion.⁵¹

$$\mathbf{r}_j(t + \Delta t) = \mathbf{r}_j(t) + F_j(t)\Delta t + \Delta X^G(t), \quad (7)$$

where $\mathbf{r}_j(t)$ is the position of j th bead at time t and the systematic force on j is denoted by $F_j(t)$. The random Brownian displacement (BD) $\Delta X^G(t)$ is taken from a Gaussian distribution with zero mean and $2\Delta t$ variance. The normalized random numbers are generated by the reshuffling

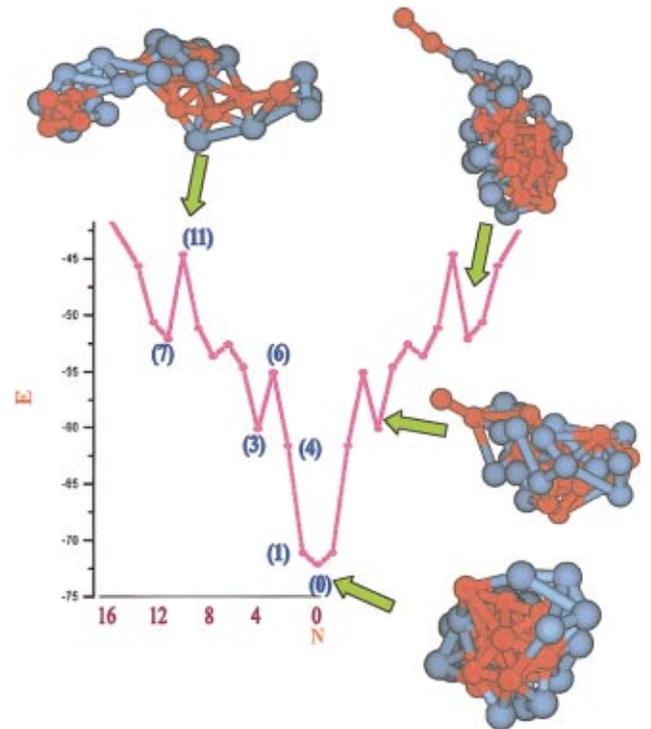


FIG. 5. (Color) Energy landscape (the funnel) for the model HP-36 protein obtained from BD simulations is shown. The distance from the native state Q in terms of topological contacts is indicated for different energy states. Configurations corresponding to various energy states (given in parentheses) (unfolded, transition, and native state) are also shown. The X axis denotes the number of configurations at energy E .

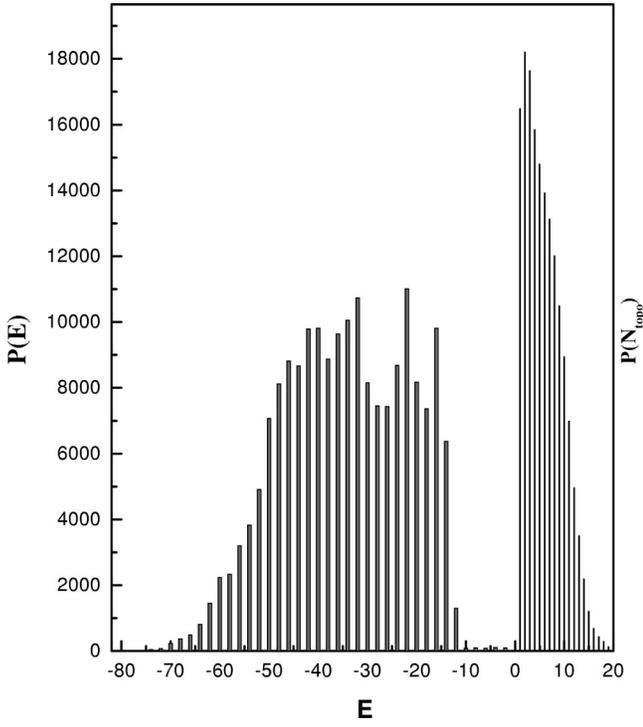


FIG. 6. The probability distribution of energy and the topological contacts between the hydrophobic residues obtained from BD simulations are shown. Positive X axis represent $P(N_{\text{topo}})$ while the energy distribution is shown on the negative X axis.

method.⁵² The time step Δt is chosen as 0.0005τ . For convenience, we define $\epsilon^* = \epsilon/k_B T$, where $k_B T$ is the thermal energy. The unit of time τ is b^2/D_0 . The length is scaled by b , the bead diameter, as usual. A scheme similar to that of Noguchi and Yoshikawa⁵⁰ was employed to investigate the structural transition.

A. 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 steps, have been incorporated to facilitate the folding. Further simulations for 2.5 million 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 500–1000 of such trajectories with different initial configuration. More details on the simulation scheme can be found in a similar study on homopolymers.^{53,54}

IV. DISTRIBUTION OF ENERGY AND TOPOLOGICAL CONTACTS: DEFINITIONS AND METHODOLOGY

The probability distribution of energy ($P(E)$) is defined by the following expression:

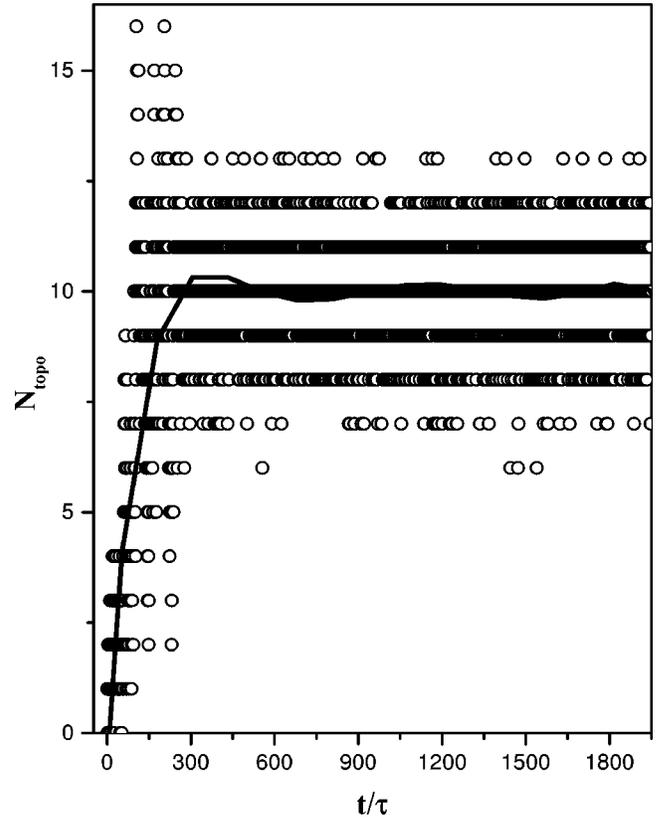


FIG. 7. The time development of the hydrophobic topological contacts for the model HP-36 protein is shown. Symbols represent the number of topological contacts at time τ . The solid line is a fit showing the qualitative nature of the same.

$$P(E) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \theta(E - E_i), \quad (8)$$

where the summation is over all the final configurations with energy E_i .

The above equation is to be understood in the following fashion. In each simulation, after choosing an initial configuration at time $t=0$, the folding is followed until a stable final state is reached. It is important to note that in each simulation only one *protein* is simulated to obtain one final energy and the corresponding hydrophobic topological contacts: the nonbonding contacts between the hydrophobic groups, which are then stored for the calculation of distribution of energies and topological contacts. Thus, in order to obtain a list of final energies and their corresponding hydrophobic contacts (N_{topo}), we had to repeat such simulations for \mathcal{N} number of independent single protein chains sampled from an equilibrium distribution. In other words, we have carried out \mathcal{N} different simulation runs with independent protein configurations (with the same sequence) to obtain \mathcal{N} number of energies and the corresponding N_{topo} values.

After the execution of simulation, the final energy values are used in Eq. (8) to obtain $P(E)$. This essentially provides an array of final energies and hydrophobic contacts corresponding to each final conformation obtained from simulations. A histogram of energy distribution is obtained by spreading $P(E)$ over a bin of width 0.5. In this way, a con-

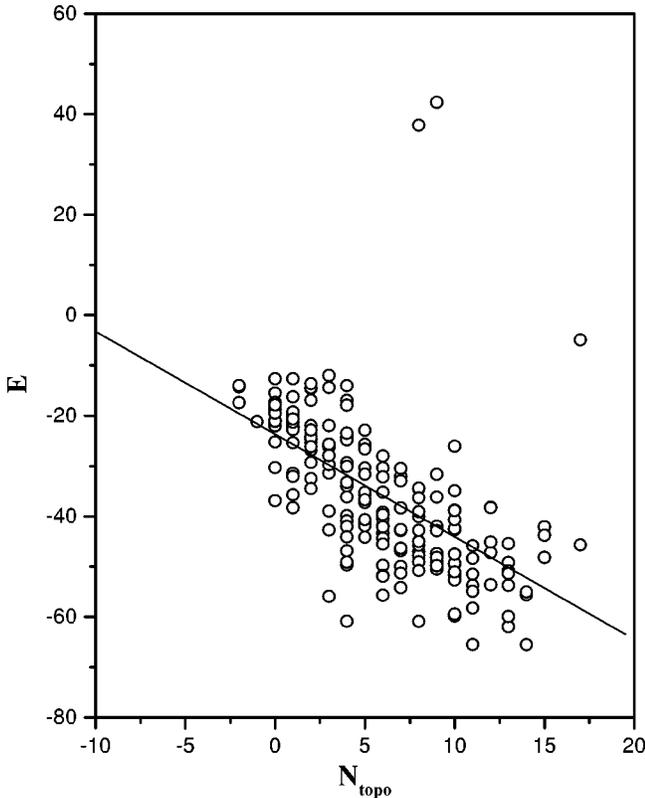


FIG. 8. The relation between the energy and the topological contacts between the hydrophobic residues is shown for proteins with different configuration. The BD simulation results are shown by the symbol while the solid line represents a linear fit. Note that this figure also represents the correlation between the stability and the topology (minimum the energy maximum the stability).

tinuous probability distribution ($P(E)$), can in principle be obtained from Eq. (8) by taking the $\mathcal{N} \rightarrow \infty$ limit—in our case, we get a histogram (Fig. 6).

$P(N_{\text{topo}})$ can be obtained in a way similar to that of $P(E)$. The only difference in this case is that the number of topological contacts (N_{topo}) should be distributed instead of energies into the bins accordingly to form the histogram. In this study we report the results on both energy and topological contact distributions. Similar results on distribution of reaction times and reaction efficiencies in a homopolymer system can be found in our earlier study.⁵⁴

V. NUMERICAL RESULTS AND DISCUSSIONS

The results obtained by performing the off-lattice BD simulations on the model HP-36 protein are presented in this section.

Figure 3 shows typical snapshots of model HP-36 folding as observed in BD simulations. The central picture corresponds to the initial configuration, while the four other structures represent the 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 core within the hydrophilic*

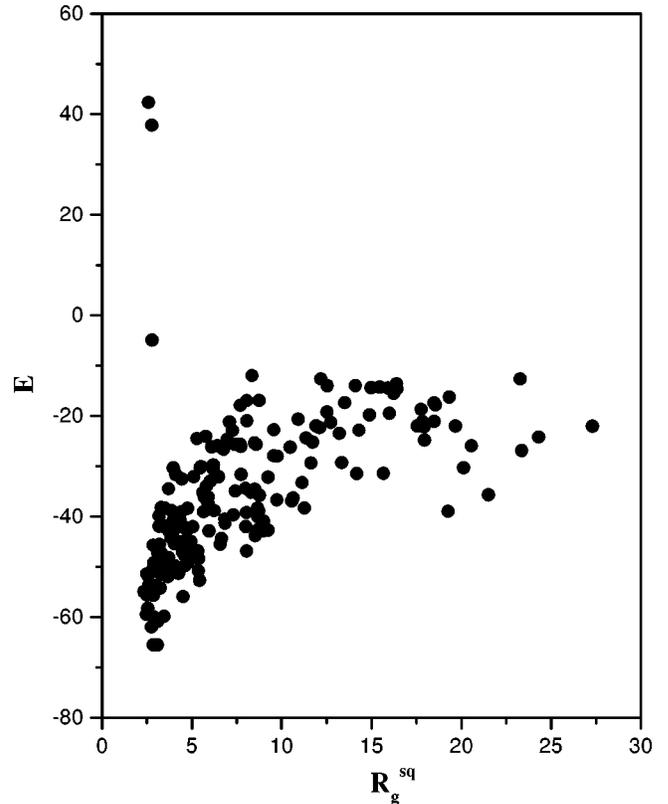


FIG. 9. Similar to Fig. 8, the protein stability is plotted as a function of the mean square radius of gyration of the protein.

outer surface. It is interesting to note that such a simplified model could represent the complex folding aspect, at least qualitatively.

In Fig. 4 the final energies for 500 model proteins of HP-36 are shown. In the same figure a few structures corresponding to different energies are also shown. Note that there exist a small fraction of structures with the positive energies too [e.g., (c) and (d)]. Close observation of such structures reveals that in them there is a hydrophilic bead stuck within the hydrophobic core. We find such configurations lead to the structure with less stability. The energy landscape picture is obtained by distributing all the final configurations into the bins of N_E : the number of final configurations with energy E . The final result, the evolution of funnel-like energy landscape, is shown in Fig. 5. For the sake of clarity, distance from the native state Q in terms of the topological contacts for various energy states is indicated in brackets ($Q=0$ for the native state). In the same figure, configurations corresponding to the native state, misfolded, unfolded, and metastable states are also shown. This figure clearly demonstrates that the present minimalist model is reliable in obtaining the qualitative features of protein folding. This is one of the main results of the present study.

The probability distribution of the energies calculated by using Eq. (8) is plotted in Fig. 6. A similarly obtained distribution of topological contacts ($P(N_{\text{topo}})$) is also shown in the same figure. In this figure the positive X axis represents the number of topological contacts, while the negative X axis corresponds to the potential energy. As can be seen from this figure the two distributions crudely resemble the mirror im-

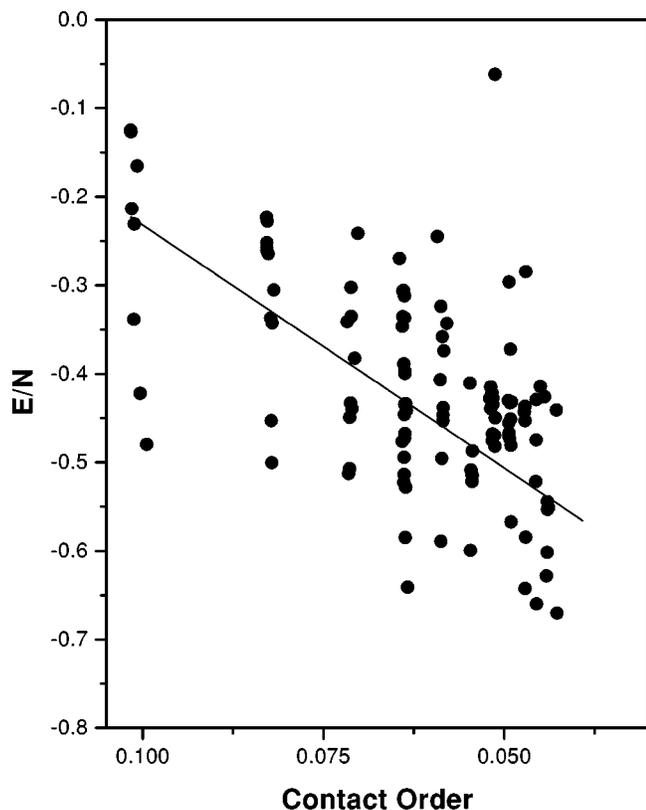


FIG. 10. The stability of protein as a function of the relative contact order is shown. The contact order is calculated by using Eq. (2). Solid line shows the linear fit.

ages. The minimum in energy distribution (at ≈ -73) corresponds to a configuration with the maximum topological contacts ($N \approx 20$). On the other hand, the majority of the configurations are found with less topological contacts and more energy. In other words, the probability to have either the maximum topological contacts or the minimum energy is relatively small. This can be attributed to the frustration. Thus this figure reveals the connection between the topology and the stability (minimum energy implies maximum stability and vice versa). To understand the role of topology in more detail, we have also studied the time development of the topological contacts. This is shown in Fig. 7 where the variation in number of topological contacts is shown for a particular protein over the entire folding time. The symbols show the number of topological contacts at time τ , obtained from simulations. The solid line is a fit, which shows the overall variation in N_{topo} . As can be seen from this figure, more or less there are no topological contacts in the unfolded state (at $\tau=0$). Protein starts folding by forming the topological contacts and reaches the final state with a nearly constant but reasonably large topological contacts. This feature is in accordance with the study of Duan and Kollman⁸ who found that the fraction of native contacts increases during the folding [Fig. 2(b) of Ref. 8]. At this point it is important to note that Duan and Kollman observed that within the unfolded villin head piece subdomain, hydrophobic collapse and helix formation occur in an initial phase followed by conformational readjustments. As mentioned in Sec. IV, since we have not taken the dihedral angles into account, we could observe

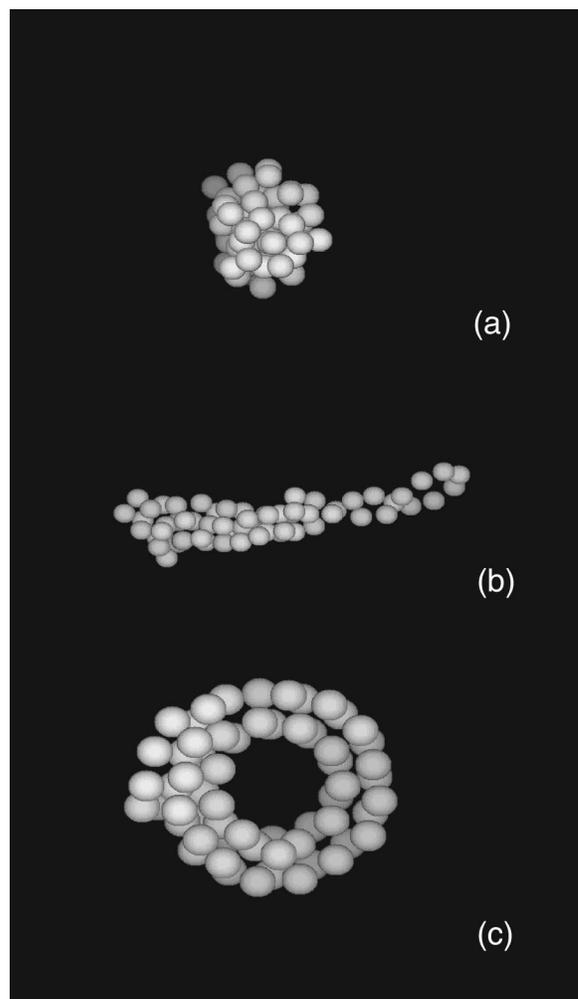


FIG. 11. Snapshots of various morphological structures of a sufficiently large homopolymer chain ($N=60$) as observed in Brownian dynamics (BD) simulations (adopted from Ref. 55). The chain stiffness parameter values are $\mathcal{S}=1$ (spherical), $\mathcal{S}=5$ (rod), and $\mathcal{S}=10$ (toroid) for the structures shown from top to bottom.

the hydrophobic collapse but not the helix formation.

We have also investigated the role of the size, stability, and the topology of the protein in determining the native state. In Fig. 8 the number of topological contacts against the energy are shown. This can be regarded as the correlation between the topology and the stability since the structures with low energy are more stable and vice versa. Similarly the stability is shown against the change in the radius of gyration (protein size) in Fig. 9. These two figures together demonstrate the role of topology and size in driving the protein folding. As shown in Fig. 8, the stability increases with increasing topological contacts between the hydrophobic residues. Figure 9 also reveals that the minimum energy configuration favors a compact structure. Among the compact structures one with the maximum topological contacts corresponds to the global minima.

A. Relative contact order

The concept of topology and its role in protein folding can be better understood in terms of the relative contact order (CO). We have calculated the CO over a number of configu-

rations by using Eq. (2). Figure 10 shows the stability of the protein as a function of the relative CO. Symbols represent the simulation result while the full line is the linear fit. As shown in this figure, there is a nearly linear correlation between the CO and the stability. Such a correlation has also been observed in the recent protein folding experiments.²⁰

B. Effect of chain stiffness

Being described as the merit of the reduction of protein to a simple heteropolymer with a certain amount of stiffness, it is worth mentioning the effect of chain stiffness itself. To this end, we have studied the effect of chain stiffness in a homopolymer by varying the chain stiffness. It is found that by varying the chain stiffness, the polymer collapses into various ordered and/or collapsed structures. In Fig. 11, the snapshots of three such structures obtained by varying the chain stiffness are shown. At low stiffness value ($S=1$) the polymer collapsed into a molten globule [Fig. 11(a)]. Increasing the chain stiffness (to $S=5$) resulted in the formation of rod-like structure, as shown in Fig. 11(b). If the polymer becomes more stiff ($S=10$) it attains a toroidal structure [Fig. 11(c)].

VI. CONCLUSION

We have studied the foldability of a thermostable chicken villin head piece subdomain, a 36-residue protein (HP-36). The complex amino acid interactions are simplified by categorizing them into three different classes with the use of hydropathy scale. The present model could capture qualitative features of the complex protein folding.

The model protein folded into the low energy compact structures. The funnel-like energy landscape for the protein folding is obtained. Although there are several states near the minimum of the folding funnel, we could identify a stable native configuration, well separated from the rest of the configurations. The study of the intercorrelations between size, stability, and topology revealed the importance of topology in determining the protein folding. This is in accordance with the recent studies of Clementi and Onuchic,⁴³ who emphasized the importance of the native state topology in determining the protein folding mechanism. The study of contact order revealed a reasonably good correlation between the stability and the contact order.

Given the complexity of the real protein folding, the ability of such a minimalist model to capture the essential features of protein folding is quite surprising. The present study suggests that it is possible to obtain qualitative information on the folding mechanism, folding rates, and also about the stability by modeling the more complex proteins in a similar way.

Future work will include a more detailed description of the hydrophobic interactions and also the differing sizes of the amino acids along the sequence.

ACKNOWLEDGMENTS

Financial support from CSIR, New Delhi, India and DST India is gratefully acknowledged. G.S. thanks CSIR for a research fellowship.

- ¹Y. Zhou, C. K. Hall, and M. Karplus, Phys. Rev. Lett. **77**, 2822 (1996); Y. Zhou, M. Karplus, J. M. Wichert, and C. K. Hall, J. Chem. Phys. **107**, 10691 (1997).
- ²J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987); J. Phys. Chem. **93**, 6902 (1989).
- ³C. Levinthal, in *Mossbauer Spectroscopy in Biological Systems*, edited by P. Debrunner, J. C. M. Tsibris, and E. Munck (University of Illinois Press, Urbana, IL), p. 22.
- ⁴R. Zwanzig, A. Szabo, and B. Bagchi, Proc. Natl. Acad. Sci. U.S.A. **89**, 20 (1992).
- ⁵R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **92**, 9801 (1995).
- ⁶E. I. Shakanovich and A. M. Gutin, Nature (London) **346**, 773 (1990).
- ⁷E. Alm and D. Baker, Proc. Natl. Acad. Sci. U.S.A. **96**, 11305 (1999).
- ⁸Y. Duan and P. A. Kollman, Science **282**, 740 (1998).
- ⁹R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **94**, 148 (1997).
- ¹⁰H. S. Chan and K. A. Dill, Annu. Rev. Biophys. Biophys. Chem. **20**, 447 (1991).
- ¹¹P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **94**, 6170 (1997).
- ¹²B. A. Schoemaker, J. Wang, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **94**, 777 (1997).
- ¹³J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci, Proc. Natl. Acad. Sci. U.S.A. **92**, 3626 (1995).
- ¹⁴E. I. Shakhnovich and M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1999).
- ¹⁵A. Sali, E. I. Shakanovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994); Nature (London) **369**, 248 (1994).
- ¹⁶J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **87**, 3526 (1990); Biopolymers **32**, 695 (1992).
- ¹⁷K. A. Dill, D. O. V. Alonso, and K. Hutchinson, Biochemistry **28**, 5439 (1989).
- ¹⁸K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
- ¹⁹K. Yue and K. A. Dill, Phys. Rev. E **48**, 2267 (1993); H. S. Chan and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **87**, 6368 (1990).
- ²⁰R. A. Dinner and M. Karplus, Nat. Struct. Biol. **5**, 236 (1998).
- ²¹J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **94**, 7129 (1997).
- ²²V. Grantcharova, E. J. Alm, D. Baker, and A. L. Horwich, Curr. Opin. Struct. Biol. **11**, 70 (2001).
- ²³K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).
- ²⁴D. J. Brockwell, D. A. Smith, and S. E. Radford, Curr. Opin. Struct. Biol. **10**, 16 (2000).
- ²⁵J. J. Portman, S. Takada, and P. G. Wolynes, Phys. Rev. Lett. **81**, 5237 (1998).
- ²⁶J. J. Portman, S. Takada, and P. G. Wolynes, J. Chem. Phys. **114**, 5069 (2001); **114**, 5082 (2001).
- ²⁷J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins **21**, 167 (1995).
- ²⁸P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **89**, 8721 (1992).
- ²⁹U. H. E. Hansmann and J. N. Onuchic, J. Chem. Phys. **1601**, 115 (2001).
- ³⁰N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
- ³¹M. A. Miller and D. J. Wales, J. Chem. Phys. **111**, 6610 (1999).
- ³²D. S. Riddle, V. P. Grantcharova, E. Alm, I. Ruczinski, and D. Baker, Nat. Struct. Biol. **6**, 1016 (1999).
- ³³K. W. Plaxco, J. I. Guijarro, C. J. Morton, M. Pitkeathly, I. D. Campbell, and C. M. Dobson, Biochemistry **37**, 2259 (1998).
- ³⁴J. Skolnick and A. Kolinski, Science **250**, 1121 (1990).
- ³⁵A. Kolinski, W. Galazaka, and J. Skolnick, Proteins **26**, 271 (1996).
- ³⁶H. Taketomi, Y. Ueda, and N. Go, Int. J. Pept. Protein Res. **7**, 445 (1975).
- ³⁷A. M. Gutin, A. M. V. I. Abkevich, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **92**, 1282 (1995).
- ³⁸Z. Guo, D. Thirumalai, and J. D. Honeycutt, J. Chem. Phys. **97**, 525 (1992); D. K. Klimov and D. Thirumalai, Proteins **26**, 411 (1996).
- ³⁹D. E. Makarov, P. K. Hansma and H. Metiu, J. Chem. Phys. **114**, 9663 (2001).
- ⁴⁰M. Levitt, Nat. Struct. Biol. **8**, 392 (2001) and references therein.
- ⁴¹M. Levitt, J. Mol. Biol. **104**, 59 (1976).
- ⁴²C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
- ⁴³C. Clementi and J. N. Onuchic, J. Mol. Biol. **311**, 879 (2001).
- ⁴⁴E. M. Boczko and C. L. Brooks III, Science **269**, 393 (1997).
- ⁴⁵U. H. E. Hansmann, M. Masuya, and Y. Okamoto, Proc. Natl. Acad. Sci. U.S.A. **94**, 10652 (1997).
- ⁴⁶U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **18**, 920 (1997).

- ⁴⁷C. J. McKnight, D. S. Doering, P. T. Matsudaria, and P. S. Kim, *J. Mol. Biol.* **260**, 126 (1996).
- ⁴⁸D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 3rd ed. (Worth, New York, 1993).
- ⁴⁹L. Stryer, *Biochemistry*, 4th ed. (Freeman, New York, 1995).
- ⁵⁰H. Noguchi and K. Yoshikawa, *J. Chem. Phys.* **113**, 854 (2000); **109**, 5070 (1998).
- ⁵¹D. L. Ermak and J. A. McCammon, *J. Chem. Phys.* **69**, 1352 (1978); R. W. Pastor, R. Zwanzig, and A. Szabo, *ibid.* **105**, 3878 (1996).
- ⁵²W. H. Press, S. A. Teulolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran* (Cambridge University Press, Cambridge, 1986).
- ⁵³G. Srinivas, A. Yethiraj, and B. Bagchi, *J. Chem. Phys.* **114**, 9170 (2001).
- ⁵⁴G. Srinivas and B. Bagchi, cond-mat/(0105138).
- ⁵⁵G. Srinivas and B. Bagchi, *J. Chem. Phys.* **116**, 837 (2002).