Contact pair dynamics during folding of two small proteins: Chicken villin head piece and the Alzheimer protein β-amyloid

Arnab Mukherjee and Biman Bagchi[a)]
Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore, India 560 012

The folding of an extended protein to its unique native state requires establishment of specific, predetermined, often distant, contacts between amino acid residue pairs. The dynamics of contact pair formation between various hydrophobic residues during folding of two different small proteins, the chicken villin head piece (HP-36) and the Alzheimer protein β-amyloid (βA-40), are investigated by Brownian dynamics (BD) simulations. These two proteins represent two very different classes—HP-36 being globular while βA-40 is nonglobular, stringlike. Hydropathy scale and nonlocal helix propensity of amino acids are used to model the complex interaction potential among the various amino acid residues. The minimalistic model we use here employs a connected backbone chain of atoms of equal size while an amino acid is attached to each backbone atom as an additional atom of differing sizes and interaction parameters, determined by the characteristics of each amino acid. Even for such simple models, we find that the low-energy structures obtained by BD simulations of both the model proteins mimic the native state of the real protein rather well, with a best root-mean-square deviation of 4.5 Å for HP-36. For βA-40 (where a single well-defined structure is not available), the simulated structures resemble the reported ensemble rather well, with the well-known β-bend correctly reproduced. We introduce and calculate a contact pair distance time correlation function, $C_P^{ij}(t)$, to quantify the dynamical evolution of the pair contact formation between the amino acid residue pairs i and j. The contact pair time correlation function exhibits multistage dynamics, including a two stage fast collapse, followed by a *slow (microsecond long) late stage dynamics* for several specific pairs. The slow late stage dynamics is in accordance with the findings of Sali *et al.* [A. Sali, E. Shakhnovich, and M. Karplus, Nature **369**, 248 (1994)]. Analysis of the individual trajectories shows that the slow decay is due to the attempt of the protein to form energetically more favorable pair contacts to replace the less favorable ones. This late stage contact formation is a highly cooperative process, involving participation of several pairs and thus entropically unfavorable and expected to face a large free energy barrier. This is because any new pair contact formation among hydrophobic pairs will require breaking of several contacts, before the favorable ones can be formed. This aspect of protein folding dynamics is similar to relaxation in glassy liquids, where also α relaxation requires highly cooperative process of hopping. The present analysis suggests that waiting time for the necessary pair contact formation may obey the Poissonian distribution. We also study the dynamics of Förster energy transfer during folding between two tagged amino acid pairs. This dynamics can be studied by fluorescence resonance energy transfer (FRET). It is found that suitably placed donor–acceptor pairs can capture the slow dynamics during folding. The dynamics probed by FRET is predicted to be nonexponential.

I. INTRODUCTION

The nonequilibrium dynamics of folding of an extended protein chain to its unique folded native state[1] is a highly complex problem with many interesting aspects which are currently being intensely investigated by physicists, chemists, and biologists. Recent experimental, theoretical, and computer simulation studies[2–12] have unearthed and explained many fascinating aspects of folding. The paradigm of energy landscape (with the idea of folding funnel directing the dynamics) has provided a new insight into the problem[10–12] and has helped in generating a language to discuss the folding problem.

For any real protein, if we assume the landscape paradigm and the associated folding funnel as the appropriate description of the pathways leading to the native state, the question regarding the dynamics of folding inevitably boils down to a problem of evolving a microscopic framework describing the dynamics of pair contact formation. The dynamics of pair contact formation is highly a nontrivial problem because it is a highly cooperative process, involving connected residues of different shapes, sizes, interactions, etc. In a qualitative sense, one can attempt to understand the initial stages as hydrophobic collapse and secondary structure formation. However, quantitative understanding of even this elementary process is highly nontrivial. As described by Gray and co-workers, the two collapsed states close to the

[a)]Electronic address: bbagchi@sscu.iisc.ernet.in

unique native structure can have very different free energy barrier.[13] While energetic penalty for forming a wrong contact could be negligible (even less than a k_BT), the free energy barrier involved for correcting it could be prohibitively large. The theoretical description based on energy landscape paradigm has mostly concentrated on the equilibrium aspect of folding. Although much progress has been made, we still need to understand the long-range contacts—the often used Gaussian chain model might not be the appropriate one in the collapsed state. In this scenario, there is need to perform further theoretical and simulation studies to understand contact pair dynamics, with model proteins.

Theoretical and computer simulation studies suggest that when folding of an isolated and extended protein is induced by sudden change of physical conditions (such as temperature or urea concentration), the initial stage of folding is the collapse of the chain and the secondary structure formation (particularly, helices).[14] The collapse is driven by hydrophobic interactions and this stage is known as hydrophobic collapse.[5] The time scale involved in this phase is believed to be much smaller than the total time involved and may be, for small proteins (less than 100 amino acid long) of the order of few tens of nanoseconds.[15] This time would certainly depend on external conditions because the slope of the landscape should depend strongly on such conditions (for example, hydrophobic interactions are temperature dependent). Contrary to the fast initial collapse and secondary structure formation, the total time required to arrive at the final native state is *much longer*, even for small proteins, and is often more than few microseconds. It is believed that this much longer time is used in the later stage of protein folding in forming the distant (along the chain contour) native contacts. As already mentioned, the late stage of protein folding is slowed down by the entropic bottle-neck which arises due to the necessity of cooperative motion of many groups together.

Experimentally one finds that the folding of small single domain proteins often follow a first-order kinetic law

$$\frac{dN_U}{dt} = -k_{\text{fold}} N_U , \tag{1}$$

where $N_U(t)$ is the number of unfolded protein at time t. k_{fold} is the rate of folding. As pointed out by others,[16,17] the first order kinetics, although can be justified within a two state model, is rather surprising, given the complexity of the folding process. As already mentioned, the rate limiting process is expected to be determined by the rate of contact formation among distant pairs. Evidence for this is provided by the relation of the rate k_{fold} with the contact order as[16,18]

$$\ln k_{\text{fold}} = a + bO_c , \tag{2}$$

where a and b are ''universal constants,'' and O_c denotes the average sequence distance between the nonlocal hydrophobic contacts and is defined by[19]

$$O_c = \frac{\Sigma_{i,j}(s_j - s_i)}{LN_c}, \tag{3}$$

where (i, j) are the specific hydrophobic pair contacts, N_c is the number of contacts while L is total number of hydrophobic amino acids present in the protein. s_j and s_i are the sequence number of the amino acids along the contour of the chain. Therefore, it may appear that the initial stage of rapid hydrophobic collapse and the secondary structure formation play no explicit role in determining the final rate of protein folding. One could, however, in principle define a rate k_{fast} to denote the rate of the initial process. In spite of the large difference between k_{fast} and k_{fold}, although not discussed explicitly, these two rates can still be related, because a high k_{fast} may lead to the formation of wrong contacts. A protein seems to know how to optimize k_{fast} to arrive at the ''best'' structure.

In addition to the relative contact order (O_c), the total number of topological contacts should also be an important ingredient. In their original analysis of protein folding, both Dill and co-workers[5] and Bryngelson and Wolynes[12] included the later as one of the two order parameters—the other being the size of the polymer chain. The native state free energy minimum is characterized by small size and large number of topological contacts.

An attractive way to explore relative dynamics of a hydrophobic pair is via the technique of fluorescence resonance energy transfer (FRET). In FRET, one measures the time dependence of energy transfer from a chosen donor fluorophore to a chosen acceptor. The rate of transfer may be due to dipolar interactions and the rate of transfer is given by the well-known Förster expression[20]

$$k_f = k_{\text{rad}} \left(\frac{R_F}{R} \right)^6 , \tag{4}$$

where k_{rad} is the radiative rate and R_F is the Förster radius. By suitably choosing donor–acceptor pair, R_F can be varied over a wide range. This allows the study of the dynamics of pair separation, essential to understand protein folding.[21] k_{rad} is typically less than (but of the order of) 10^9 s^{-1}. Thus Förster transfer provides us with a sufficiently fast camera to take snapshots of the dynamics of contact pair formation.

In this study, we have investigated dynamics of pair contact formation by Brownian dynamics simulations of two small model proteins. The first one we study is chicken villin head piece, popularly known as HP-36, which is one of the smallest protein (36 amino acid residues) that folds autonomously to a stable compact ordered structure, with a large helix content.[22] HP-36 is a subdomain of chicken villin which is implicated in the formation of microvilli in the absorptive epithelium of the gut and the proximal tube of the kidney.[23] All atom simulation study on this HP-36 have revealed at least two pathways of folding.[24] Earlier, several studies of HP-36 were presented using Monte Carlo[25] and Brownian dynamics simulations.[26] Some of the earlier off-lattice and Langevin dynamics simulation studies of folding used simple model potential,[5,27,28] such as bead-spring model with two kinds (hydrophobic and hydrophilic) of amino acid beads. These simulations identified a compact ordered lowest energy structure as the native state. Our minimalistic model is considerably more complex. Not only the backbone and each amino acid are accounted for explicitly, a general hydropathy scale has been used to model the effective interaction between the amino acids, which are also of different sizes. Helix propensity scale has been used to model the hydrogen

bonding, which are present between i and $i+4$ amide groups in the helices of real proteins. However, the model is simpler than what used in the well-known work of Scheraga and co-workers, who used more detailed and accurate potential.[29] The success of the Scheraga group in predicting protein structure accurately is well known.[30] However, the simplicity of our model allowed us to simulate many (about 600) folding trajectories.

While Hp-36 folds to a globular state, the small protein βA-40 has a nonglobular, stringlike structure in its biologically active state. βA-40 has been implicated in the Alzheimer disease. It is believed that βA-40 aggregates to form filaments and deposit as plaques to block the neural network. Several experimental[31–36] and theoretical[37,38] studies have been performed on the thermodynamics and kinetics of plaque formation in βA-40. The aggregation of beta amyloid is believed to happen via its unfolded monomeric state, while the folded state is not prone to aggregation.[39] The main difference between the folded and nonfolded monomer is the presence of a β bend in the former. It is the opening of this βA-40 bend which is believed to make βA prone to association. It is thus essential to understand the relative stability of the folded state with respect to its unfolded state, the kinetics of folding, and energy landscape of this protein. Such understanding can then be used to understand the relative probability and rates of its association.

Our reason for studying these two proteins together is that they represent two very different classes—HP-36 being globular while βA-40 is nonglobular. We have compared not only the folding of each with the respective known structures, but also among the two. There are not only interesting differences but also similarities, as discussed below.

The present simulations reveal several interesting results. We find that the folding trajectories of both the proteins exhibit multistage kinetics. The initial fast hydrophobic collapse is followed by a long plateau regions during which several contacts form and break. Most of the reduction in size and also a large number of hydrophobic contacts, especially the neighboring ones, have already been formed before reaching the plateau region. This plateau region is typically microsecond long. The breaking and formation of hydrophobic contacts are reflected in small variations in size and energy of the plateau. The final approach to the native state is found to be accompanied by rather sudden rise in O_c and a small but noticeable lowering in energy.

The present simulations suggest that the length of the plateau is different for different proteins. If the first order kinetics is to be strictly valid, the duration of the plateau must obey a Poissonian statistics.

There are interesting differences between the folding patterns of HP-36 and βA-40. First, the native state of βA-40 is characterized by less reduction in size from its extended state. While radius of HP-36 decreases on folding by about 50%, the same for βA-40 is only about 25%. A similar difference is also found for the total energy change. That is, the folded state of βA is not only much less compact (compared to HP-36), but the relative stability (compared to its own extended state) is less than the same for HP-36. These are of course direct consequences of the nonglobular nature of the native state of βA-40. Not only these, but also the relative arrangement of the hydrophobic and hydrophilic groups in this protein makes it amenable to association, as discussed later.

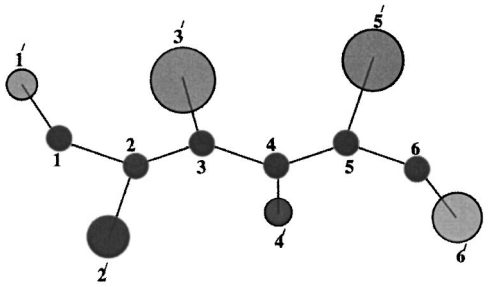

FIG. 1. The basic construction of the model of HP-36 protein is shown. Atoms marked as 1,2,3, etc. indicate backbone C_α atom, whereas, those marked with $1',2',3'$ denote the whole side chain residue attached to that particular C_α atom.

As pointed out by Snow *et al.*,[40] one needs to simulate a very large number of trajectories because only a small fraction is expected to reach the folded state within the available simulation time. In the present Brownian dynamics simulations, we have simulated about 600 trajectories and observed a similar fact—typically 3%–5% arrive at a low RMSD configuration. All the successful trajectories have been analyzed in terms of topology and related quantities.

The organization of the rest of the paper is as follows: In the next section, we discuss the construction of the interaction potential, which, as already stated, is rather new. In Sec. III, we present the simulation details. Section IV contains results and discussion of the folding of HP-36. Section V contains results on contact pair dynamics. Section VI contains the folding dynamics of βA-40. We conclude the paper in Sec. VII with a brief discussion on the implications of the results and the possible future directions of research.

II. CONSTRUCTION OF ENERGY FUNCTIONS OF THE MODEL PROTEINS

Figure 1 shows the basic construction of the model protein. The backbone atoms are numbered as i's where $i=1,2,3,...$,etc., whereas the amino acid residues are numbered as i''s, where $i'=1',2',\ldots$, etc. While the backbone atoms are all of equal size, each side chain atom bears the characteristics of the real amino acid in a given protein. Construction of the model protein has been described in detail elsewhere.[41] Similar types of model (with more rigorous force field) have recently been introduced by Scheraga *et al.*[29,30] Please note here, the construction of the model protein is modified a bit in case of the folding studies of βA-40, where the amino acid glycine is mimicked by only one atom (C_α). Every other simulation detail is same for both the model proteins discussed here.

The total potential energy function of the model protein V_{Total} is written as

$$V_{\text{Total}}=V_B+V_\theta+V_T+V_{\text{LJ}}+V_{\text{helix}}, \tag{5}$$

where V_B and V_θ are the potential contributions due to vibration of bonds and bending motions of the bond angles.

Standard harmonic potential is assumed for the above two potentials. The spring constant for the bonds between two backbone atoms is 43.0 kJ mole^{-1} Å^{-2} and that between a side residue and the adjacent backbone atom is 8.6 kJ mole^{-1} Å^{-2}. In case of the bending potential, spring constant is taken to be 10.0 kJ mol^{-1} rad^{-2}. $V_T(=\epsilon_T \Sigma_\phi(\frac{1}{2})[1+\cos(3\phi)])$ is taken as the torsional potential for the rotations of the bonds. $\epsilon_T = 1$ kJ mol^{-1}.

The nonbonding potential V_{LJ} is the sum of Lennard-Jones pair interaction between the atoms as given by

$$V_{\mathrm{LJ}} = 4\sum_{i,j} \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{6}$$

where r_{ij} and ϵ_{ij} are the separation and interaction strength of the "*i-j*" pair. σ_{ii} is the diameter of the ith atom. $\sigma_{ij} = \frac{1}{2}(\sigma_{ii}+\sigma_{jj})$ and $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$. Sizes and interactions are taken to be the same (1.8 Å and 0.05 kJ mol^{-1}, respectively) for all the backbone atoms as they represent the C_α atoms in case of real proteins. Side residues, on the other hand, carry the characteristics of a particular amino acid. Different sizes of the side residues are taken from the values given by Levitt.[42] Interactions of the side residues are obtained from the hydrophobicities of the amino acids. We construct the effective potential guided by the well-known statistical mechanical relation between potential of mean force and the radial distribution function, $V_{ij}^{\mathrm{eff}} = -k_B T \ln g_{ij}(r)$.[43] Strong correlation among the hydrophobic groups (absent among the hydrophilic amino acids) implies that the hydrophobic amino acids should have stronger effective interaction than the hydrophilic groups. So the interaction parameters of the side residues are mapped from the hydropathy scale[44] by using a linear equation of the form

$$\epsilon_{ii} = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) * \left(\frac{H_{ii} - H_{\min}}{H_{\max} - H_{\min}}\right), \tag{7}$$

where, ϵ_{ii} is the interaction parameter of the ith amino acid with itself. $\epsilon_{\min}(=0.2$ kJ mol$^{-1})$ and $\epsilon_{\max}(=11.0$ kJ mol$^{-1})$ are the minimum and maximum values of the interaction strength chosen for the most hydrophilic (arginine) and most hydrophobic(isoleucine) amino acids, respectively. H_{ii} is the hydropathy index of ith amino acid given by Kyte and Doolittle[44] and $H_{\min}(=-4.5)$ and $H_{\max}(=4.5)$ are the minimum and maximum hydropathy index among all the amino acids. Further details are available in Ref. 41.

An important part of secondary structure of the real protein is the α-helix. In the absence of hydrogen bonding, we introduce the following effective potential among the backbone atoms to mimic the helix formation along the chain of residues

$$V_{\mathrm{helix}} = \sum_{i=3}^{N-3} [\tfrac{1}{2}K_i^{1-3}(r_{i,i+2} - r_h)^2 + \tfrac{1}{2}K_i^{1-4}(r_{i,i+3} - r_h)^2], \tag{8}$$

where $r_{i,i+2}$ and $r_{i,i+3}$ are the distances of ith atom with $i+2$ and $i+3$th atoms, respectively. r_h is the equilibrium distance and is taken as 5.5 Å, motivated by the observation that the distances of r_i with r_{i+2} and r_{i+3} are nearly constant at 5.5 Å in an α helix. The summation excludes the first and last three amino acids as the helix formation is much less observed at the ends of a protein chain.[45] The force constant for the above harmonic potential is mapped from the helix propensities Hp_i obtained from Scholtz *et al.*,[46] $\mathcal{K}_i = \mathcal{K}_{\mathrm{alanine}} - Hp_i \times (\mathcal{K}_{\mathrm{alanine}} - \mathcal{K}_{\mathrm{glycine}})$. $\mathcal{K}_{\mathrm{alanine}}$ and $\mathcal{K}_{\mathrm{glycine}}$ are the force constants for alanine and glycine, 17.2 and 0.0 kJ mol^{-1}, respectively. Next, the influence of the neighboring amino acids for the formation of helix has been considered by taking an average of the spring constants, $K_i^{1-3} = \frac{1}{3}[\mathcal{K}_i + \mathcal{K}_{i+1} + \mathcal{K}_{i+2}]$ and $K_i^{1-4} = \frac{1}{4}[\mathcal{K}_i + \mathcal{K}_{i+1} + \mathcal{K}_{i+2} + \mathcal{K}_{i+3}]$, with the condition that K_i^{1-3}, $K_i^{1-4} \geqslant 0$ as the force constant must remain positive. The above formulation of helix potential is motivated by the work of Chou and Fasman about the prediction of helix formation that *the neighbors of a particular amino acid* should also be considered along with its own helix propensity.[47] That is, *the model effective helix potential must be nonlocal*.

III. SIMULATION DETAILS

The initial configurations of the model proteins were generated by configurational bias Monte Carlo (CBMC) technique.[48] Atoms attached to a single branch point were generated simultaneously. Then the initial configurations were subjected to Brownian dynamics simulation to study the dynamics of folding. Time evolution of the model protein was carried out according to the motion of each atom as below

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \frac{D_i}{k_B T}\mathbf{F}_i(t)\Delta t + \Delta \mathbf{r}_i^G, \tag{9}$$

where each component of $\Delta \mathbf{r}_i^G$ is taken from a Gaussian distribution with zero mean and variance $\langle (\Delta r_{i\alpha}^G)^2 \rangle = 2D\Delta t$.[43,49] $\mathbf{r}_i(t)$ is the position of the ith atom at time t and the systematic force on ith atom at time t is $\mathbf{F}_i(t)$. D_i is the diffusion coefficient of the ith atom calculated from the relation $D_i = C/R_i$, where C is a constant, and R_i is the radius of the ith atom. The unit of length is σ(3.41 Å) and the unit of time $\tau = \sigma^2/D_0$. D_0 is the diffusion coefficient obtained by using σ as the diameter in the above equation. τ *is* $\sim$1.2 ns *in the real unit for the reduced viscosity* $\eta = 10$ which is kept fixed. The time step Δt is taken equal to 0.001τ. The simulation has been carried out for $\mathcal{N}$ number (where $\mathcal{N} \simeq 600$) of folding trajectories for both the model proteins—HP-36 and βA-40.

IV. FOLDING CHARACTERISTICS OF HP-36

It is important for any minimalistic model to first establish and satisfy several basic criteria. In case of off-lattice model representation of the proteins, the success lies in reproducing the basic structural similarity of the model protein with the real one. Our model potential seems to be fairly successful in this regard. We first start with the amino acid sequence of HP-36, which is shown in one-letter code in Fig. 2. The solid circles indicate the hydrophobic amino acids, whereas the open circles denote the hydrophilic amino acids. Figure 3 depicts the close resemblance of the structures among the real native structure of chicken villin with one of the folded states of the model protein. Among several folded

FIG. 2. The amino acid sequence of HP-36 is shown through one letter code. The solid circles denote the hydrophobic amino acids, whereas the hydrophilic amino acids are indicated by the open circles.

states obtained from the Brownian dynamics simulation, Fig. 3(a) shows the one with the lowest root-mean-square deviation (4.5 Å), calculated among the backbone atoms, from the native real protein shown in Fig. 3(b). Thus, the minimalistic model, with the use of hydropathy scale and helix propensity, successfully reproduces the helices and bends at the appropriate positions with respect to the real protein, despite the lack of detailed interaction such as the hydrogen bonding among the amide groups, electrostatic interactions and the explicit water. Characteristic features of the model protein have been described in detail elsewhere.[41] Note that the lowest RMSD structures are not the lowest energy structure. This is a well-known price paid by the minimalistic models for not including water explicitly and has been discussed in the literature.[25]

Minimalistic models along with the use of Brownian dynamics simulation allow one to explore the landscape of folding in greater detail than a more microscopic treatment, for example, the full atomistic simulation, like the one carried out.[24] An ensemble ($\mathcal{N}$) of initial high temperature configurations was subjected to temperature quench and the subsequent folding was monitored, until each trajectory led to a final folded state. Analysis of high temperature and the low temperature equilibrated configurations reveals change in the probability distributions of various interesting quantities, such as the total energy, the radius of gyration of the model protein, relative contact order, and also the hydrophobic topological contact number at two different temperatures. These are plotted in Fig. 4. The probability distributions of energy at high temperature equilibrated configurations and low temperature folded configurations are shown in Fig. 4(a). The distributions are well separated, which signify the presence of well defined folded states. Note that the energy probability distribution is considerably narrower at the low temperature, signifying a funnel-like energy landscape.

Figure 4(b) shows the probability distribution of the number of hydrophobic topological contacts at high and low temperatures, both of which show, interestingly, Gaussian distribution. The distribution shifts toward higher value of topological contact, which characterizes the hydrophobic core formation in the folded state. Similar behavior can be observed in case of the probability distribution of radius of gyration plotted in Fig. 4(c). Here the high temperature distribution is Gaussian like, showing different possible structural arrangements, due to large conformational entropy. At low temperature, however, the distribution is peaked around 9.6 Å. Note that this is *precisely the experimental value of the radius of gyration of real native HP-36*.

The relative contact order (O_c), introduced by Baker *et al.*,[18,19] denotes the average sequence separation for the hydrophobic residues. O_c is thus a good measure of the range of contact formation. Figure 4(d) shows the probability distribution of O_c for both high and low temperatures. Note the *predominant formation of long-range contacts at the low temperature* folded configurations, which is absent at higher temperature.

In order to probe the distribution of amino acid residues

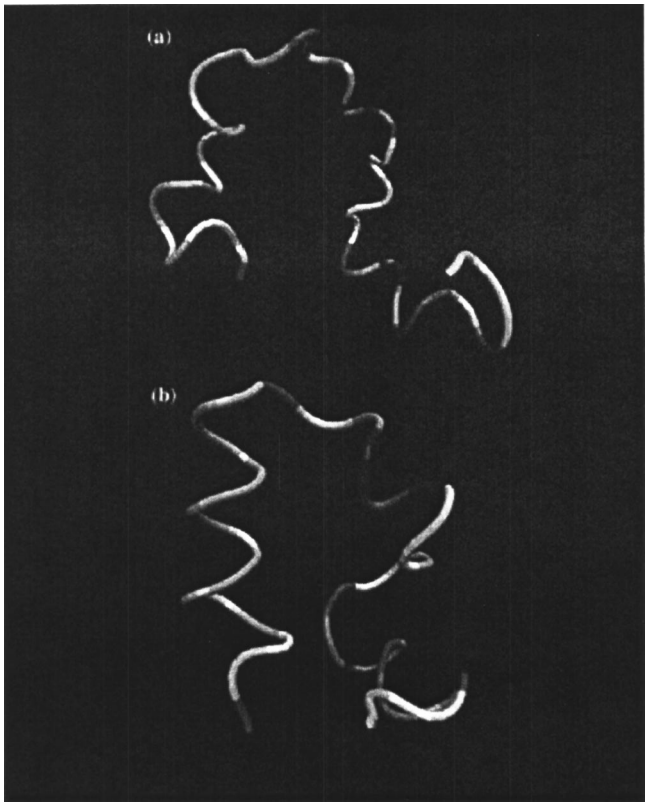

FIG. 3. (a) The backbone structure of the model protein with the lowest RMSD (4.5 Å). (b) The backbone structure of the native state of real HP-36.

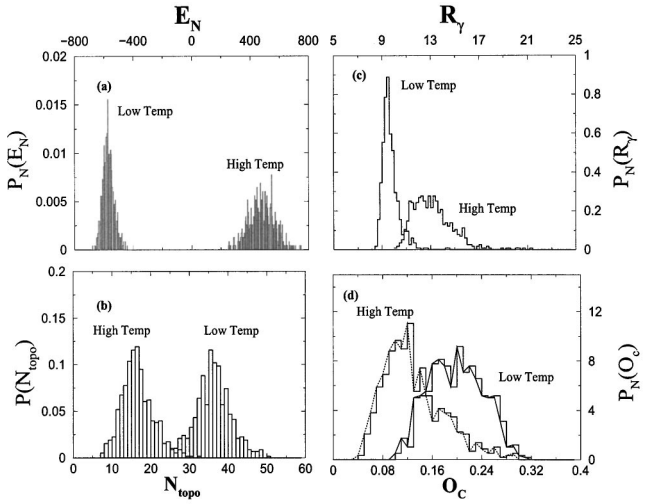

FIG. 4. Normalized probability distributions are plotted for high temperature unfolded states and low temperature folded states (as marked in the figures) for different parameters: (a) Total energy, (b) hydrophobic topological contact, (c) radius of gyration, (d) relative contact order parameter; in this figure, the lines trace the distribution to guide the eye.

in the folded and extended states, we have calculated the Forster efficiency distribution. The well-known Forster efficiency is defined by the following expression:

$$\Phi_F = \left\langle \frac{k_F}{k_F + k_{\rm rad}} \right\rangle = \langle [1 + (R/R_F)^6]^{-1} \rangle, \tag{10}$$

where $\langle \cdots \rangle$ means an equilibrium average over $\mathcal{N}$ configurations. We calculated the average Forster efficiency[50] as defined above using the equilibrium configurations both at the low and high temperatures. While the high temperature histogram shows peak at very low Forster efficiency, as expected, the distribution for low temperature folded states shows a peak at high Forster efficiency *with a tail extending upto very low efficiency*, signifying many trapped states with larger separation between donor and acceptor even at a low temperature. Presence of this prolonged tail toward low efficiency even at low temperature is probably an artifact of the present scheme as one does not expect many trapped states in real systems.[51]

V. CONTACT PAIR DYNAMICS: HP-36

A protein possesses complicated and highly correlated network of interactions, and the late stage of folding can be thought of as the interparticle diffusion over a rugged free energy landscape, which is clearly a consequence of strong intermolecular correlations at high density. During the fast hydrophobic collapse (where free energy surface is largely smooth with a steep downward slope), several contacts among the amino acids could be formed which may not be the ''best'' or the native contacts. Therefore, a subsequent cooperative rearrangement is required to replace the wrong (meaning energetically unfavorable) contacts and replace them with the correct, native contacts. Since most contacts have already been formed, the correction of the few wrong ones is a process which is entropically highly unfavorable, even the energy of activation can be significant, because many contacts have to be opened up so that the correct contacts can form. This is a slow and gradual process and is reflected in the wide separation of time scales observed in the time evolution in most dynamical quantities. This in turn may give rise to multistage dynamics in the macroscopic quantities such as total energy, radius of gyration, relative contact order, etc. The consequence of this multistage dynamics is reflected in related dynamical quantities such as survival probability for FRET. Folding can be probed microscopically by monitoring the dynamics of separation between different amino acid pairs.

The existence of the widely different time scales of movement of all the different pairs together gives rise to an overall picture of the dynamics of folding which is reflected in the macroscopic quantities. The effective dynamics of pair separation can be described by introducing a new pair correlation function defined below[26]

$$C_P^{ij}(t) = \frac{d^{ij}(t) - d^{ij}(\infty)}{d^{ij}(0) - d^{ij}(\infty)}, \tag{11}$$

where, $d^{ij}(t) = |\mathbf{r}_i(t) - \mathbf{r}_j(t)|$. $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of the ith and jth atom, respectively.

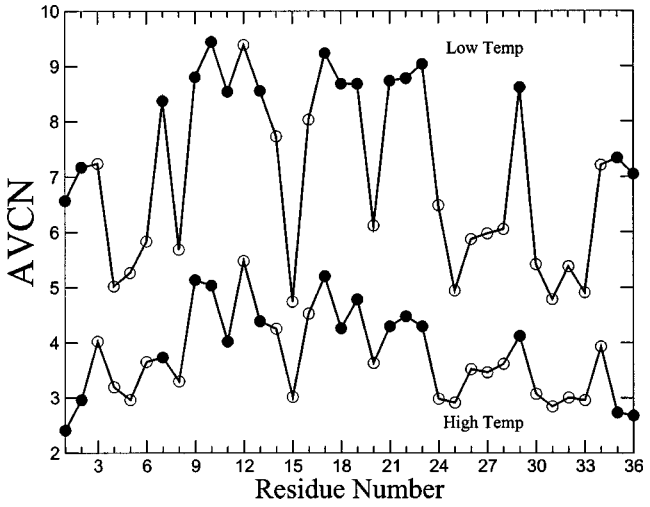

FIG. 5. Average number of contacts (AVCN) (obtained from $\mathcal{N}$ folding trajectories) formed by a particular residue is plotted against residue number, for both high and low temperatures. Filled and open circles denote the hydrophobic and hydrophilic residues, respectively. Solid lines are just for the guidance to the eye. Note that average number of contacts formed by hydrophobic residues is much larger than that of the hydrophilic residues.

To explore the dynamics of pair separation, it is essential to look for a proper nucleation center, around which the pair contact separation evolves significantly at a later stage of folding. Figure 5 shows the average number of contacts (obtained from $\mathcal{N}$ different folded states) formed by a particular residue for both high temperature unfolded and low temperature folded states. The hydrophobic beads are denoted by solid circles, whereas the open circles signify the hydrophilic ones. It is interesting to note that at high temperature, the average contact formation is random due to high conformational entropy. At low temperature, however, the average contact formation is *much higher for the hydrophobic beads, especially those which are in the middle of the sequence*.

The dynamics of contact formation is calculated for the 9th side residue (that is, $i=9$) with the other hydrophobic residues. So j in Eq. (11) is varied. Figure 6 shows the time

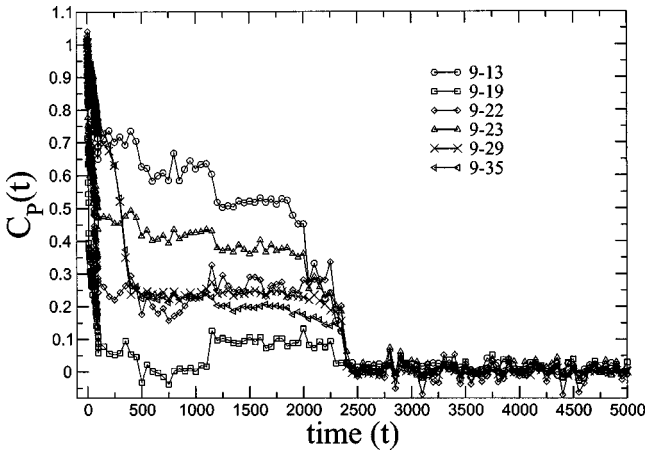

FIG. 6. Dynamics of contact formation of different side residues with the 9th side residue is shown. The multistage relaxation process in the dynamical quantities originates from the diverse dynamics of the contact pairs. The time is scaled by τ and $\tau = 1.2$ ns for an aqueous solution.

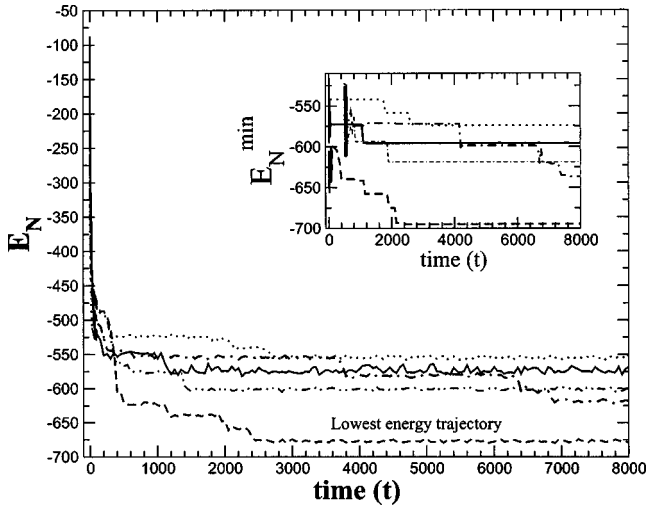

FIG. 7. Time evolution of energy of five different trajectories is shown. Inset shows the change in the energy of the local minima corresponding to the respective time evolved configurations (time evolution of energy and minimum energy for a particular trajectory is shown by a particular line type).

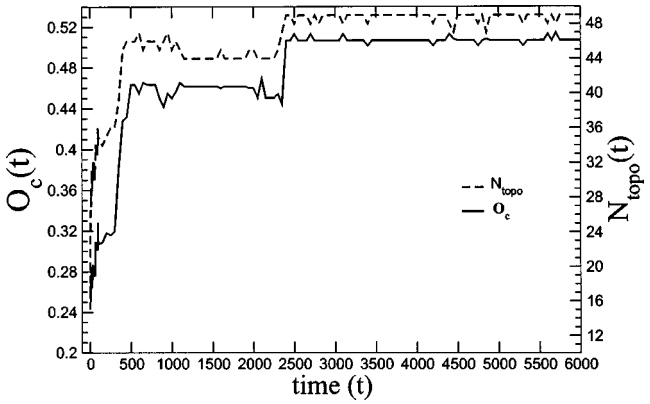

FIG. 8. Time evolution of relative contact order $[O_c(t)]$ and hydrophobic topological contacts (N_{topo}) along the folding trajectory, which leads to the lowest energy state among total $\mathcal{N}$ simulations.

dependence of contact formation subsequent to quenching, that is $C_P^{9j}(t)$ along the folding trajectory that leads to the lowest energy state. Multistage dynamics is also observed for $C_P^{9j}(t)$. The side residues, which are close to the tagged one, collapse fast, some even show an initial shoulder. Several show the plateau in the long time decay *that correlates with the similar plateau observed in case of other dynamical quantities*.[41]

Survival probability in fluorescence resonance energy transfer (FRET) also shows the signature of multistage dynamics. Survival probability was calculated using the Förster energy transfer rate k_f given in Eq. (4). During the time evolution of the folding trajectory with a discrete time step of Δt, a random number, uniformly distributed between 0 and 1, is generated at each time step. Förster transfer was assumed to take place if the drawn random number was less than or equal to $k_f \Delta t$. The survival probability S_P remains unity until the Förster transfer occurs, after which it becomes zero. This process was repeated for N_s number of times (where $N_s = 10^5$). So the survival probability is the average of many Heaviside step functions. The equation of survival probability is given by

$$S_P(t) = \frac{1}{N_s} \sum_{i=1}^{N_s} H(t - t_i), \tag{12}$$

where, $H(x)$ is a Heaviside step function, which is 1 for its positive argument, and 0 otherwise. t_i is the time of Förster transfer at the ith observation. Time dependence of FRET shows the signature of multistage decay of $S_P(t)$,[50] similar to the dynamics of other dynamical properties such as energy, relative contact order, etc. discussed later. Although the survival probability could capture the dynamics of folding for appropriate R_F, it is found to be relatively insensitive to k_{rad}.[50]

Figure 7 shows the multistage time evolution of energy of model HP-36 for five representative folding trajectories among the total $\mathcal{N}$ simulations. Each of the five trajectories corresponds to a complete Brownian dynamics simulation for a particular CBMC generated initial configuration. All the trajectories show the first initial collapse within $500t$ ($\sim$600 ns). The next stage of folding is long and takes about 1–10 μs, varying for different trajectories. The particular trajectory leading to the lowest energy state is shown. Note that, all the trajectories do not lead to the same lowest energy state, or they do not follow the same path to go to the folded state. Many of them are trapped at an early local minima, where they stay infinitely. This is reflected in the long tail of the Förster efficiency distribution at the low temperature.[50] Energy of the local minima corresponding to a particular time evolved configuration was calculated with the help of conjugate gradient technique at some time interval for all the trajectories. Inset shows the change in the local minima for all the folding trajectories. During the initial collapse, each trajectory goes over many different local minima. After the collapse, the long plateau in all the trajectories correspond to a particular local minima. Any change in the energy is accompanied by the change in the local minima as shown in the inset. This signifies a large conformational barrier faced at the later stage of folding when folding probably is driven by the need to form correct native contacts. Wrong contacts lead to a wrong path and misfolding, and probably entanglement.

The dynamics of relative contact order and hydrophobic topological contact plotted in gives further insight into the dynamics of folding in the model proteins. We have plotted the time evolution of the above quantities in Fig. 8 for the trajectory leading to the lowest energy state. The multistage folding dynamics observed in the previous figure for the time evolution of energy seems to sustain in the time variation of both O_c and N_{topo}. There is an initial rapid formation of the topological contact, and relative contact order, followed by a slower late stage dynamics. The change in energy, and the *corresponding change* in the energy of the local minima is accompanied by the change in contact formation, evident from the increase in N_{topo}, and the contacts are long ranged, as denoted by the increase in O_c. Note that, the O_c is the average sequence separation. So increase in O_c signifies the formation of long-range contacts. The contact formation and

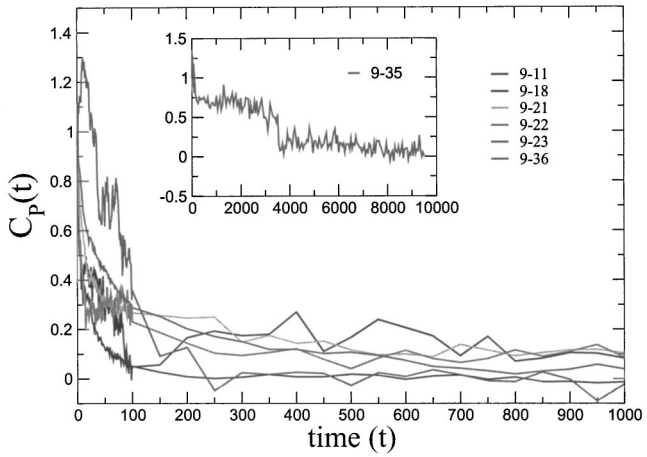

FIG. 9. Dynamics of contact formation of 9th side residues with other hydrophobic side residues. The result is an average over $\mathcal{N}$ configurations. The multistage dynamics is reflected even in this *average-microscopic* quantity. Inset shows the long relaxation dynamics of 9–35 pair contact.

breaking seem to continue even after the protein has reached the folded state. Similar picture was provided by Sali *et al.*[11]

All the individual simulation shows multistage dynamics for the dynamical variables such as energy, radius of gyration, topological contact, relative contact order etc. However, the average dynamics looks very different. In order to get an idea about the average kinetics of protein folding, we have averaged the energies at a particular time ($\langle E_N \rangle$) after the temperature quench at $t=0$, for all the $\mathcal{N}$ trajectories. Time variation of this average energy shows an overall nonexponential dynamics, with very low stretching coefficient. The signature of fast initial collapse and slow late stage dynamics is reflected also in this highly nonexponential average dynamical behavior.

Figure 9 shows the average of the contact pair dynamics obtained from $\mathcal{N}$ different contact pair trajectories. Note that, even after the average over $\mathcal{N}$ trajectories, the dynamics of contact pair still shows widely different time scales for different pairs.

VI. FOLDING OF βA-40 AND MULTISTAGE DYNAMICS

The multistage dynamics discussed in the previous sections is observed for other model proteins as well. In order to illustrate the scope of the off-lattice model of proteins discussed here, we have implemented the same model and Brownian dynamics simulations to study βA-40 protein and its different fragments, all of which are of enormous importance due to their implications in the misfolding related diseases. The misfolded βA-40 are prone to associate and they form large aggregates, polymerize to insoluble fibrils and deposit as plaque. One of the major cause of Alzheimer disease is this plaque depositions. A number of structural models of amyloid fibrils have recently been proposed that argues for both parallel[52,53] and anti-parallel β-sheet[54] organization. Multiple quantum solid-state nuclear magnetic resonance (NMR) experiment indicated an in-register parallel arrangement of Aβ.[55]

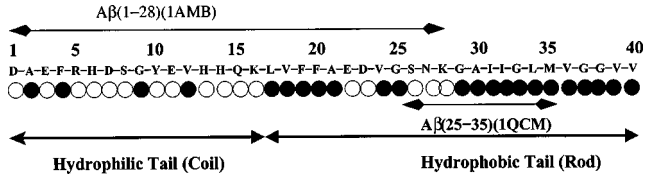

FIG. 10. Sequence of the βA-40 and its different fragments are shown. The hydrophobicity of the amino acids is shown by the open and filled circles for the hydrophilic and hydrophobic amino acids, respectively. The gray circles denote the neutral glycine residues. The coarse grained hydrophobicity is also shown.

Aβ is derived from the COOH-terminus of a 695 residue precursor protein, known as βAPP$_{695}$.[56] The sequence of Aβ includes the first 28 residues of the extracellular domain and the rest 11–14 residues of the transmembrane region. Figure 10 shows the sequence of the Aβ and its different fragments. The hydrophobic amino acids are shown by filled circles, whereas the open circles denote the hydrophilic amino acids. Glycine is almost neutral, so it is indicated by gray (shaded).

Aβ is a small protein without large stability for the native state. It has no unique native structure. The interatomic distances, obtained from two-dimensional (2D) NMR, and the constrained minimization result in 20 different structures obtained from protein data bank (PDB).[57] Individually all the NMR structures show a sequence of secondary structures: random coil $\rightarrow$ helix $\rightarrow$ bend $\rightarrow$ random coil $\rightarrow$ helix $\rightarrow$ bend. Interestingly, when plotted together, all the native structures, obtained from NMR, show the pronounced β bend in the middle (around residue 24–28) for all the folded forms.

Brownian dynamics simulations of the model βA-40 protein yield low temperature states, some of which (with low RMSD values compared to one of the NMR structure of the real protein[58]) are shown in Fig. 11(a). Here also note the same pronounced β bend. The two structures are statistically quite similar. Figure 11(b) shows the conformations of the model βA-40 equilibrated at a high temperature. The high temperature configurations assume a bent-rod structure.[59]

Figure 12 shows the multistage decay in the time evolution of the energy of β-amyloid for the representative folding trajectory, in this case the particular one leading to the lowest RMSD structure. This shows the similar features as observed in case of the folding trajectories of model HP-36. There is an initial ultrafast hydrophobic collapse, which is followed by a long plateau. This indicates the slower rearrangement of the side residues necessary to form the rate determining long- range contacts. Inset of Fig. 12 shows the change in local energy minima corresponding to the configuration at a particular time along the path of folding. *The overall monotonic decease in the local minima is a direct signature of folding*. Note that, the model showing multistage dynamics continues to exist even in case of the nonglobular proteins. So it can be regarded as a general phenomena observed in case of minimalistic model, which can infer some insight into the dynamics of folding.

Since βA-40 is a nonglobular protein, the stability of the native state or the decrease of the radius of gyration is expected to be much smaller than that of the globular HP-36 protein. Figure 13 shows the decrease in radius of gyration

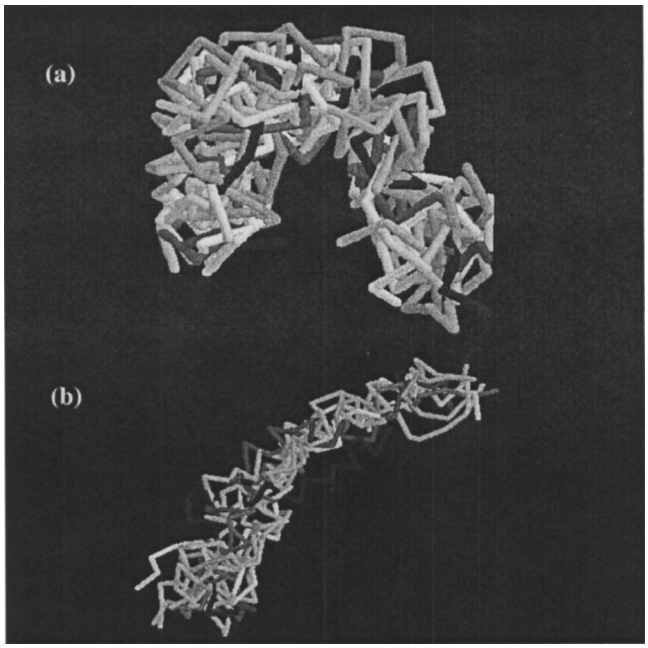

FIG. 11. (a) The 10 different folded (quenched) backbone structures (plotted with Rasmol software) of the model Aβ obtained from Brownian dynamics simulations with the minimalistic model. The β turn is clearly visible. (b) The configurations at high-temperature show an extended bent-rodlike structure.

per amino acids for globular HP-36 and nonglobular βA-40 and its different fragments. The fragments of βA-40 are also important in misfolding related diseases and they are well studied. The peptide containing the first 28 amino acids [β(1–28)] of βA-40 has been enlisted in the PDB[57] with pdb-id 1amb,[60] whereas the one with 25th to 35th residues [β(25–35)] of βA-40 exists in the PDB with the pdb-id 1qcm.[61] Both the fragments show string like structure. The decrease is much more for the globular protein, while the fragments having string like structures show least decrease in radius of gyration per amino acid.

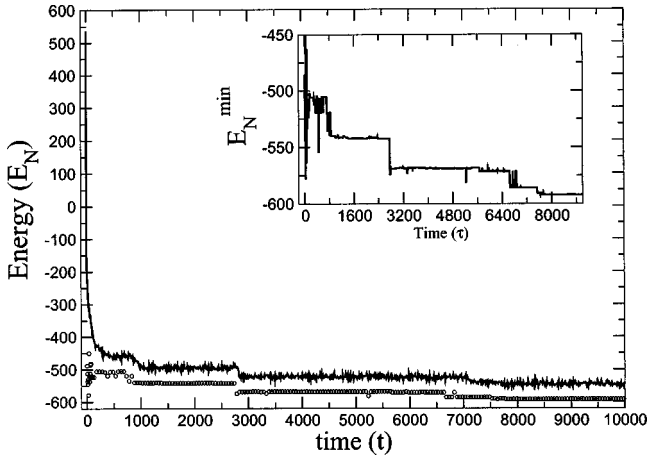

FIG. 12. Multistage dynamics is shown for the time evolution of energy of βA-40. Inset shows the monotonically decreasing energy of the local minima corresponding to the time evolved configuration at a particular time. Time is scaled by τ, where *tau* is ~1.2 ns for aqueous solution.

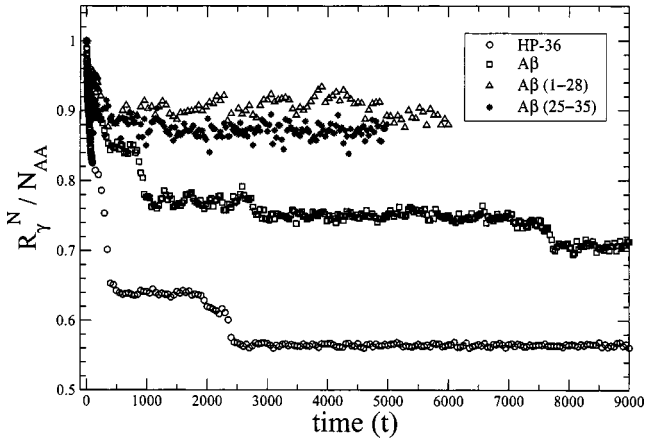

FIG. 13. Radius of gyration per amino acid (R_γ^N/N_{AA}) for different proteins—HP36, βA-40, β(1-28), and β(25-35). Note the relatively smaller decrease in R_γ^N/N_{AA} for nonglobular proteins.

Based on the above results, we propose the following model of aggregation and subsequent deposition of βA-40 protein. Basic physical concepts that determine the aggregation of Aβ are the following. (i) Interactions (with nucleation sites) among the hydrophobic residues decrease the energy by association, (ii) the entopic contribution comes from the randomization of the hydrophilic tail. These two factors can lead to a rich and complex free energy landscape. The dimer itself can have two alternating arrangements[59] as shown in Fig. 14. In these two arrangements, π-stacking interaction among phenylalanine groups and other such strong hydrophobic interactions would favor parallel arrangement, but at the cost of entropy loss at the hydrophilic chain end. The hydrophilic tails would favor anti-parallel arrangement due to the larger accessible degrees of freedom. However, in the anti-parallel arrangement, the hydrophobicity remains frus-

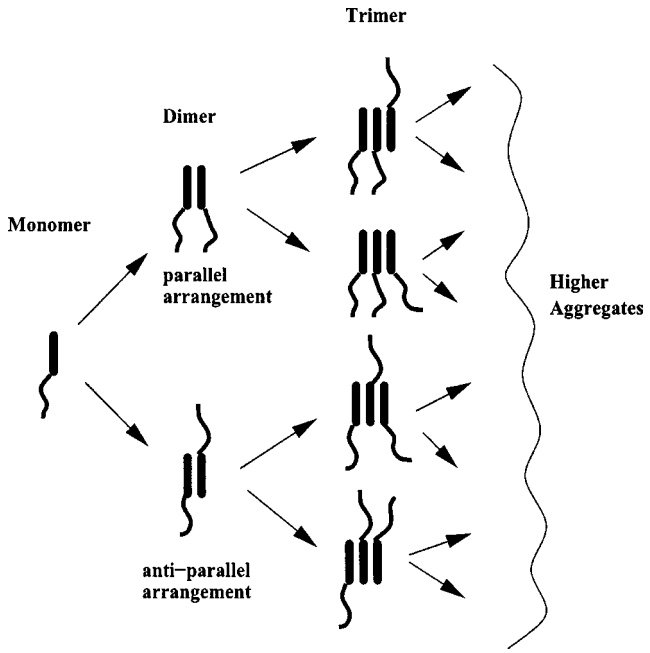

FIG. 14. Schematic representation of the proposed aggregation model. The possibility of both parallel and anti-parallel arrangements is shown.

trated. *Such arrangements have indeed been predicted and observed in the rod-coil diblock copolymers*,[62] which lend support to our argument.

VII. CONCLUDING REMARKS

In this work we have used the newly developed minimalistic model to study several dynamic aspects of folding in two small but important proteins—HP-36 and βA-40. The minimalistic model is sufficiently simple to allow BD simulation of large number of folding trajectories, yet it retains a few of the complexities of the real protein. Contact pair dynamics and the time evolution of energy, radius of gyration, relative contact order formation, etc. bring out the rich and diverse dynamics of protein folding. The initial ultrafast hydrophobic collapse signifies that the upper part of the funnel is steep—followed by a change in slope. The *rate determining step, however, arises from the final stage of folding on a flat and rugged energy landscape marked by large conformational entropy barrier, but with little energy change*.[8] This entopic bottleneck arises from the necessity to form long-range hydrophobic contacts, as envisaged earlier by Dill and Wolynes.[5,12] This separation of time scale was earlier suggested by Sali *et al.*[11] The dynamics in the flat rugged energy landscape is rather similar in nature to the slow dynamics observed in computer simulations on relaxation and transport in glassy liquids.

The present work also brought out several detailed features. The atoms modeling the whole side chain of the real protein play important role in structural and dynamical aspects of folding. Moreover, the contact pair correlation function introduced can probe the folding events in minute detail. The dynamics of contact pair formation can be explored by FRET. It is interesting to note that our minimalistic model can reproduce several features observed experimentally, not only the structure but also (somewhat surprisingly) the known time scale of folding for HP-36.

The study of βamyloid reveals several interesting features. Our model correctly reproduces the β bend in the folded state and predicts that the high-temperature form of βA is a bent-rod extended structure. Depending on this structure of the unfolded state, we have proposed a model of association of β amyloids. This model is based on the analogy of extended βA-40 with rod-coil diblock copolymers. However, a lot more work is required to develop this model further.

How to develop a theory to describe the dynamics of the late stage of protein folding where contacts are broken and formed? As already emphasized, this is a strongly correlated process. Recently *Makarov et al.* have presented an elegant analysis of the rate of contact formation assuming Gaussian distribution of the probability of intersegment distance. This analysis finds the following expression for the rate k_{fold} as given below[16]

$$k_{\mathrm{fold}} = N k_d \exp[-F_0/k_B T] \exp[-N\Delta F/k_B T], \tag{13}$$

where ΔF is the mean free energy gained when forming a contact and k_d is the mean rate constant. Notably, those authors could derive the experimentally observed logarithmic dependence of rate on contact order. They have also observed that faster scale dynamics of the individual contact formation is beyond the time scale of observation employed in the current experiment.

We finally address the issue of the observed first-order kinetics in the folding of small proteins. As observed in present simulations, the waiting time for the contact formation is different for different trajectories. A simple, straightforward but microscopic explanation of the exponential kinetics emerges if this waiting time distribution for contact pair formation obeys a Poissonian distribution as given by

$$P(\tau) = \frac{1}{\tau} e^{(-t/\tau)}. \tag{14}$$

The reason for the Poissonian distribution for the waiting time could be a severe entopic bottleneck, whose ''width'' is nonfluctuating, and the waiting time itself has a narrow distribution. These are of course the classic conditions for the observation of an exponential kinetics.[63]

The origin of the (on the average) microsecond long waiting time is due to combination of two factors. One is of course the narrow entopic bottleneck. The second one is that in the collapsed state the effective diffusion coefficient of amino acid monomer is small because of the high density. This concept can be quantified in the following way. When a particle escapes from a cavity through a narrow window, it's survival probability decays as a single exponential, in spite of the fact that no energy barrier is involved. For example, the escape rate of a particle of diffusion coefficient D through a round hole of radius a from a cavity of volume V is given by[64]

$$k = 4Da/V. \tag{15}$$

In the collapsed state, not only the equivalent of a/V is small, but the effective diffusion constant D is also small. The treatment of Makarov *et al.* addresses the free energy barrier part but not the kinetics. Notable advances in addressing the dynamics has recently been made by Wolynes and co-workers.[9]

ACKNOWLEDGMENTS

The authors thank CSIR, DAE, and DST, India for financial support. We thank Professor P. Balaram and Professor P. G. Wolynes for useful comments. A.M. thanks Ashwin Sampangiri, Kausik Chakrabory, Prasanth P. Jose, and Dwaipayan Chakrabarti for technical discussions.

[1] C. B. Anfinsen, Science **181**, 223 (1973).
[2] M. Levitt and A. Warshel, Nature (London) **253**, 694 (1975).
[3] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[4] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[5] K. A. Dill, D. O. V. Alonso, and K. Hutchinson, Biochemistry **28**, 5439 (1989); K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[6] P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **89**, 8721 (1992).
[7] R. Zwanzig, A. Szabo, and B. Bagchi, Proc. Natl. Acad. Sci. U.S.A. **89**, 20 (1992).
[8] R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **92**, 9801 (1995).
[9] J. J. Portman, S. Takada, and P. G. Wolynes, Phys. Rev. Lett. **81**, 5237 (1998); J. Chem. Phys. **114**, 5069 (2001); **114**, 5082 (2001).

[10] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[11] A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).
[12] J. D. Bryngelson, M. Montal, J. N. Socci, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **21**, 167 (1995); J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).
[13] F. A. Tezcan, W. M. Findley, B. R. Crane, S. A. Ross, J. G. Lyubovitsky, H. B. Gray, and J. R. Winkler, Proc. Natl. Acad. Sci. U.S.A. **99**, 8626 (2002).
[14] W. A. Eaton, V. Munoz, V. , P. A. Thompson, E. R. Henry, and J. Hofrichter, Acc. Chem. Res. **31**, 745 (1998).
[15] V. Munoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature (London) **390**, 196 (1997).
[16] D. E. Makarov, C. A. Keller, K. W. Plaxco, and H. Metiu, Proc. Natl. Acad. Sci. U.S.A. **99**, 3535 (2002).
[17] L. J. Lapidus, W. A. Eaton, and J. Hofrichter, Proc. Natl. Acad. Sci. U.S.A. **97**, 7220 (2000).
[18] V. Grantcharova, E. J. Alm, D. Baker, and A. L. Horwich, Curr. Opin. Struct. Biol. **11**, 70 (2001).
[19] K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).
[20] Th. Förster, Ann. Phys. (Leipzig) **2**, 55 (1948).
[21] J. R. Telford, P. Wittung-Stafshede, H. B. Gray, and J. R. Winkler, Acc. Chem. Res. **31**, 755 (1998); T. Pascher, J. P. Chesick, J. R. Winkler, and H. B. Gray, Science **271**, 1558 (1996).
[22] C. J. McKnight, D. S. Doering, P. T. Matsudaria, and P. S. Kim, J. Mol. Biol. **260**, 126 (1996).
[23] A. Bretscher and K. Weber, Proc. Natl. Acad. Sci. U.S.A. **76**, 2321 (1979).
[24] Y. Duan and P. A. Kollman, Science **282**, 740 (1998).
[25] U. H. E. Hansmann and L. T. Wille, Phys. Rev. Lett. **88**, 068105 (2002).
[26] G. Srinivas and B. Bagchi, J. Chem. Phys. **116**, 8579 (2002); **119**, 8 (2003).
[27] J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **87**, 3526 (1990); Biopolymers **32**, 695 (1992).
[28] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **97**, 7254 (2000).
[29] A. Liwo *et al.*, J. Comput. Phys. **18**, 850 (1997).
[30] J. Pillardy, C. Czaplewski, A. Liwo *et al.*, Proc. Natl. Acad. Sci. U.S.A. **98**, 2329 (2001); J. Lee, A. Liwo, and H. A. Scheraga, *ibid.* **96**, 2025 (1999).
[31] W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghilardi, J. P. Lee, P. W. Mantyh, and J. E. Maggio, Biochemistry **39**, 6288 (2000).
[32] J. P. Lee, E. R. Stimson, J. R. Ghilardi *et al.*, Biochemistry **34**, 5191 (1995).
[33] Y. Kusumoto, A. Lomakin, D. B. Teplow, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **95**, 12277 (1998).
[34] D. B. Teplow, Int. J. Exp. Clin. Invest. **5**, 121 (1998).
[35] B. P. Tseng, W. P. Esler, C. B. Clish, E. R. Stimson, J. R. Ghilardi, H. V. Vinters, P. W. Mantyh, J. P. Lee, and J. E. Maggio, Biochemistry **38**, 10424 (1999).
[36] W. P. Esler, A. M. Felix, E. R. Stimson, J. Struct. Biol. **130**, 174 (2000).
[37] J. E. Straub, J. Guevara, S. Huo, and J. P. Lee, Acc. Chem. Res. **35**, 473 (2002).
[38] F. Massi, J. W. Peng, J. P. Lee, and J. E. Straub, Biophys. J. **80**, 31 (2001).
[39] R. J. Ellis and T. J. T. Pinheiro, Nature (London) **416**, 483 (2002).
[40] C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[41] A. Mukherjee and B. Bagchi, J. Chem. Phys. **118**, 4733 (2003).
[42] M. Levitt, J. Mol. Biol. **104**, 59 (1976).
[43] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic, New York, 1986).
[44] J. Kyte and R. F. Doolittle, J. Mol. Biol. **157**, 105 (1982).
[45] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[46] C. N. Pace and J. M. Scholtz, Biophys. J. **75**, 422 (1998).
[47] P. Y. Chou and G. Fasman, Biochemistry **13**, 211 (1974).
[48] G. C. A. M. Mooij, D. Frenkel, and B. Smit, J. Phys.: Condens. Matter **4**, L255 (1992); D. Frenkel and B. Smit, Mol. Phys. **75**, 983 (1992).
[49] D. L. Ermak and J. A. McCammon, J. Chem. Phys. **69**, 1352 (1978).
[50] A. Mukherjee and B. Bagchi, Curr. Sci. **85**, 68 (2003).
[51] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss, Proc. Natl. Acad. Sci. U.S.A. **97**, 5179 (2000).
[52] P. T. Lansbury, P. R. Costa, J. M. Griffiths *et al.*, Nat. Struct. Biol. **2**, 990 (1995).
[53] L. Li, T. A. Darden, L. Bartolotti, D. Kominos, and L. G. Pedersen, Biophys. J. **76**, 2871 (1999).
[54] T. L. S. Benzinger, D. M. Gregory, T. S. Burkoth, H. Miller-Auer, D. G. Lynn, R. E. Botto, and S. C. Meredith, Proc. Natl. Acad. Sci. U.S.A. **95**, 13407 (1998).
[55] O. N. Antzutkin, J. J. Balbach, R. D. Leapman, N. W. Rizzo, J. Reed, and R. Tycko, Proc. Natl. Acad. Sci. U.S.A. **97**, 13045 (2000).
[56] J. Kang, H. G. Lemaire, A. Unterbeck, J. M. Salbaum, C. L. Masters, K. H. Grzeschik, G. Multhaup, K. Beyreuther, and B. Muller-Hill, Nature (London) **325**, 733 (1987).
[57] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, J. Mol. Biol. **112**, 535 (1977).
[58] H. Sticht, P. Bayer, D. Willbold, S. Dames, C. Hilbich, K. Beyreuther, R. W. Frank, and P. Rosch, Eur. J. Biochem. **233**, 293 (1995).
[59] A. Mukherjee and B. Bagchi, Proc. Indian Acad. Sci. (Chem. Sci.) **115**, 621 (2003).
[60] J. Talafous, K. J. Marcinowski, G. Klopman, and M. G. Zagorski, Biochemistry **33**, 7788 (1994).
[61] T. Kohno, K. Kobayashi, T. Maeda, K. Sato, and A. Takashima, Biochemistry **35**, 16094 (1996).
[62] D. R. M. Williams and G. H. Fredrickson, Macromolecules **25**, 3561 (1992).
[63] S. Yang and J. Cao, J. Phys. Chem. B **105**, 6536 (2001).
[64] I. V. Grigoriev, Y. A. Makhnovskii, A. M. Berezhkovskii, and V. Y. Zitserman, J. Chem. Phys. **116**, 9574 (2002).