

## Probing folding free energy landscape of small proteins through minimalistic models: Folding of HP-36 and **b**-amyloid<sup>†</sup>

ARNAB MUKHERJEE and BIMAN BAGCHI\*

Solid State and Structural Chemistry Unit, Indian Institute of Science,  
Bangalore 560 012, India  
e-mail: bbagchi@sscu.iisc.ernet.in

**Abstract.** Folding dynamics and energy landscape picture of protein conformations of HP-36 and **b**-amyloid (**A<sub>b</sub>**) are investigated by extensive Brownian dynamics simulations, where the inter amino acid interactions are given by a minimalistic model (MM) we recently introduced [*J. Chem. Phys.* **118** 4733 (2003)]. In this model, a protein is constructed by taking two atoms for each amino acid. One atom represents the backbone C<sub>α</sub> atom, while the other mimics the whole side chain residue. Sizes and interactions of the side residues are all different and specific to a particular amino acid. The effect of water-mediated folding is mapped into the MM by suitable choice of interaction parameters of the side residues obtained from the amino acid hydrophobicity scale. A new non-local helix potential is incorporated to generate helices at the appropriate positions in a protein. Simulations have been done by equilibrating the protein at high temperature followed by a sudden quench. The subsequent folding is monitored to observe the dynamics of topological contacts ( $N_{topo}$ ), relative contact order parameter (RCO), and the root mean square deviation (RMSD) from the real-protein native structure. The folded structures of different model proteins (HP-36 and **A<sub>b</sub>**) resemble their respective real native state rather well. The dynamics of folding shows *multistage decay*, with an initial hydrophobic collapse followed by a long plateau. Analysis of  $N_{topo}$  and RCO correlates the late stage folding with rearrangement of the side chain residues, particularly those far apart in the sequence. The long plateau also signifies large entropic free energy barrier near the native state, as predicted from theories of protein folding.

**Keywords.** Protein folding; free energy landscape, HP-36; **b**-amyloid.

### 1. Introduction

Protein folding is a problem of immense current interest. The structure of a protein is intimately connected to its functionality. Elucidation of structures from the sequence of amino acids has remained virtually unsolved, in spite of a colossal number of studies in this area.<sup>1</sup> Dynamics of folding equally poses a challenge to unravel the unresolved issue of Levinthal paradox, which says that a protein should take astronomically large time to find its native state.<sup>2,3</sup> In reality, it folds in the range of microseconds to seconds.

Several studies on protein structure and dynamics have established few basic aspects of folding. A protein could be thought of as a heteropolymer, consisting of mainly hydrophobic and hydrophilic amino acid residues. The water repellent hydrophobic residues are the ones that drive the initial collapse. Many theoretical studies have been carried out

<sup>†</sup>Dedicated to Professor C N R Rao on his 70th birthday

\*For correspondence

on the models of small single domain proteins. The early statistical mechanical theories of Dill and co-workers<sup>4,5</sup> and of Bryngelson and Wolynes<sup>6</sup> were based on heteropolymer collapse and re-ordering among hydrophilic and hydrophobic residues, which follows the collapse, leading the protein to its native state.

Protein folding is a collective self-organisation process, which could occur by a multiplicity of routes down a folding free energy funnel.<sup>6-8</sup> The free energy surface may be constructed by using several different order parameters such as topological contact, radius of gyration etc.<sup>9</sup> More recently, Wolynes and co-workers have presented a detailed variational theory of protein folding.<sup>10,11</sup>

The all atom simulation of protein folding is computationally expensive. So simple lattice and off-lattice models of proteins have been used to study the statistical and dynamical aspects of folding.<sup>12,13</sup> Levitt pioneered in the computational studies of protein folding using off-lattice protein models.<sup>14</sup> A recent off-lattice model study of HP-36 based on hydrophobicity tried to correlate the folding with many important equilibrium properties.<sup>15,16</sup> Recently, united atom model of Scheraga *et al*<sup>17</sup> extended the scope of off-lattice model by successfully reproducing the native state of a 254-residue protein. All atom molecular dynamics<sup>18</sup> and Monte Carlo simulations<sup>19</sup> have also been carried out to get the intricate detail of the structure and dynamics of protein folding.

Here, we have used a simple off-lattice model to study the structure, energy landscape and topology of the model protein HP-36 and **A**myloid. HP-36 is the 36 residue protein obtained from the headpiece subdomain of villin. A large number of studies have been devoted to this particular protein because it is the smallest protein which can autonomously fold to its native state. **A**myloid is a 40 residue protein derived from the COOH-terminus of a 695 residue precursor protein, known as **A**PP<sub>695</sub>.<sup>20</sup> It is also an important protein in the aspect of misfolding related diseases. **A**myloid, being a non-globular protein, is prone to associate and deposit in the form of plaque. This is believed to be the major cause of well-known Alzheimer disease. Naturally, amyloid plaque formation is now a subject of intense study.

Construction of our model protein is based on a linear mapping of hydrophobicity of the amino acids into the interaction potential of model proteins following the statistical mechanical concept of 'potential of mean force'. In addition, the helix propensities of the amino acids are also used to effectively generate helices in the model protein. Extensive Brownian dynamics simulations are carried out for several initial configurations, where folding is initiated by sudden temperature quench. Many interesting features that come out of this study are as follows. (i) The folded structures of the model protein resemble well with the structure of the corresponding real native protein. (ii) Funnel-type energy landscape is obtained from the ensemble of folded states. (iii) Topological contact and relative contact order show Gaussian distribution. (iv) Multistage dynamics is observed with initial hydrophobic collapse followed by a long time tail. We have also discussed about the possible mechanisms of association of **A**

## 2. Construction and energy functions of the model proteins

Each amino acid of the model protein is represented by two atoms. The smaller atom represents the backbone  $C_a$  atom of real protein, while the bigger atom mimics the whole side chain residue. Figure 1 shows the basic construction of the model protein. The backbone atoms are numbered as  $i$ 's, where  $i = 1, 2, 3, \dots$ , etc. whereas the side residues are numbered as  $i'$ 's, where  $i' = 1', 2', \dots$ , etc. Each  $i$  and  $i'$  together represent one amino

acid. Construction of the model protein has been described in detail elsewhere.<sup>21</sup> Similar types of model (with more rigorous force field) have recently been introduced by Scheraga *et al.*<sup>22</sup> The total potential energy function of the model protein  $V_{\text{Total}}$  is written as,

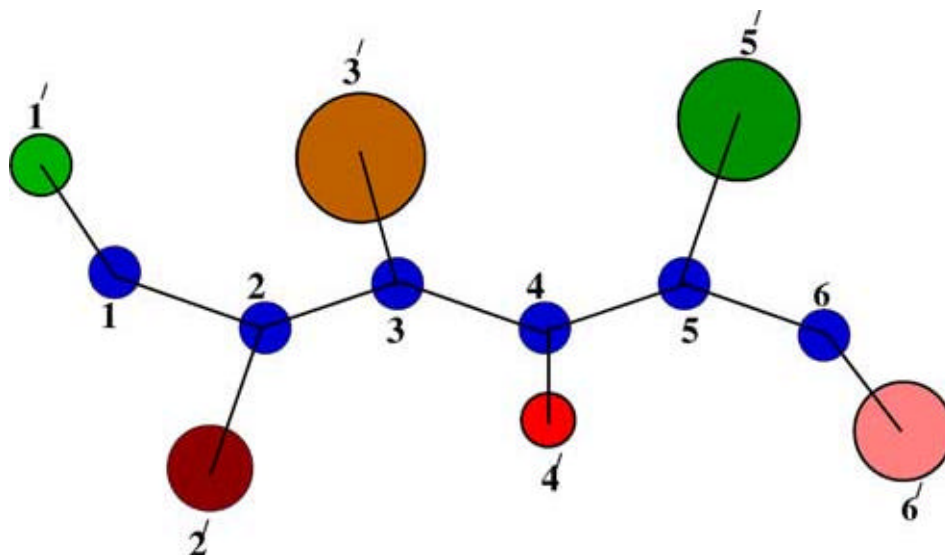
$$V_{\text{Total}} = V_B + V_{\mathbf{q}} + V_T + V_{LJ} + V_{\text{helix}} \quad (1)$$

where  $V_B$  and  $V_{\mathbf{q}}$  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  $43.0 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$  and that between a side residue and the adjacent backbone atom is  $8.6 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ . In case of the bending potential, spring constant is taken to be  $10.0 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ .  $V_T (= \epsilon \sum_{\mathbf{f}} (1/2)[1 + \cos(3\mathbf{f})])$  is taken as the torsional potential for the rotations of the bonds.  $\epsilon = 1 \text{ kJ mol}^{-1}$ .

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

$$V_{LJ} = 4 \sum_{i,j} \epsilon_{ij} \left[ \left( \frac{\mathbf{s}_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\mathbf{s}_{ij}}{r_{ij}} \right)^6 \right], \quad (2)$$

where  $r_{ij}$  and  $\epsilon_{ij}$  are the separation and interaction strength of the ‘ $i$ – $j$ ’ pair,  $\mathbf{s}_{ij} = \frac{1}{2} (\mathbf{s}_i + \mathbf{s}_j)$  and  $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$ . Sizes and interactions are taken to be the same ( $1.8 \text{ \AA}$ , and  $0.05 \text{ kJ mol}^{-1}$  respectively) for all the backbone atoms as they represent the  $C_{\alpha}$  atoms in case of real proteins. Side residues, on the other hand, carry the characteristics of a



**Figure 1.** Basic construction of the model protein is shown.  $C_{\alpha}$  atoms are numbered as 1, 2, 3 etc. whereas the side residues are shown by 1', 2', 3' etc. Note the varying size of the side residues.

particular amino acid. Different sizes of the side residues are taken from the values given by Levitt.<sup>13</sup> 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,  $V_{ij}^{eff} = -k_B T \ln g_{ij}(r)$ .<sup>23</sup> 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 hydrophathy scale<sup>24</sup> by using a linear equation of the form:

$$\mathbf{e}_i = \mathbf{e}_{\min} + (\mathbf{e}_{\max} - \mathbf{e}_{\min}) * \left( \frac{H_{ii} - H_{\min}}{H_{\max} - H_{\min}} \right), \quad (3)$$

where,  $\mathbf{e}_i$  is the interaction parameter of the  $i$ th amino acid with itself.  $\mathbf{e}_{\min}$  ( $= 0.2 \text{ kJ mol}^{-1}$ ) and  $\mathbf{e}_{\max}$  ( $= 11.0 \text{ kJ mol}^{-1}$ ) 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 hydrophathy index of the  $i$ th amino acid given by Kyte and Doolittle<sup>24</sup> and  $H_{\min}$  ( $= -4.5$ ) and  $H_{\max}$  ( $= 4.5$ ) are the minimum and maximum hydrophathy index among all the amino acids. Table 1 shows the hydrophathy index and the interaction parameters for all the 20 different amino acids. Further details are available in elsewhere.<sup>21</sup>

An important part of secondary structure of the real protein is the formation of  $\alpha$  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,

**Table 1.** Sizes and equilibrium bond angle values for all the different amino acids. The Kyte–Doolittle hydrophathy scale and its translation to LJ interaction parameter.

Amino acid	Size ( $\text{\AA}$ )	Bond angle ( $^\circ$ )	$H_{ii}$	$\mathbf{e}_i$ ( $\text{kJ mol}^{-1}$ )
ala	4.60	121.90	1.8	7.76
val	5.80	121.70	4.2	10.64
leu	6.30	118.10	3.8	10.16
ile	6.20	118.90	4.5	11.00
cys	5.00	113.70	2.5	8.60
met	6.20	113.10	1.9	7.88
pro	5.60	81.90	1.6	7.52
phe	6.80	118.20	2.8	8.96
tyr	6.90	110.00	-1.3	4.04
trp	7.20	118.40	-0.9	4.52
asp	5.60	121.20	-3.5	1.40
asn	5.70	117.90	-3.5	1.40
gln	6.10	118.00	-3.5	1.40
his	6.20	118.20	-3.2	1.76
glu	6.10	118.20	-3.5	1.40
ser	4.80	117.90	-0.8	4.64
thr	5.60	117.10	-0.7	4.76
arg	6.80	121.40	-4.5	0.20
lys	6.30	122.00	-3.9	0.92
gly	3.80	109.50	-0.4	5.12

$$V_{\text{helix}} = \sum_{i=3}^{N-3} \left[ \frac{1}{2} K_i^{1-3} (r_{i,i+2} - r_h)^2 + \frac{1}{2} K_i^{1-4} (r_{i,i+3} - r_h)^2 \right], \quad (4)$$

where  $r_{i,i+2}$  and  $r_{i,i+3}$  are the distances of the  $i$ th 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  $\alpha$  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.<sup>25</sup> The force constant for the above harmonic potential is mapped from the helix propensities  $Hp_i$  obtained from Scholtz *et al.*,<sup>26</sup>  $\mathcal{K}_i = \mathcal{K}_{\text{alanine}} - Hp_i \times (\mathcal{K}_{\text{alanine}} - \mathcal{K}_{\text{glycine}})$ .  $\mathcal{K}_{\text{alanine}}$  and  $\mathcal{K}_{\text{glycine}}$  are the force constants for alanine and glycine, 17.2 and 0.0 kJ mol<sup>-1</sup> respectively. The values of  $\mathcal{K}_i$  are given in table 2. Next, the influence of the neighbouring 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  $K_i^{1-3}, K_i^{1-4} \geq 0$  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 neighbours of a particular amino acid* should also be considered along with its own helix propensity.<sup>27</sup>

### 3. Simulation detail

The initial configurations of the model proteins were generated by configurational bias Monte Carlo technique.<sup>28</sup> 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,

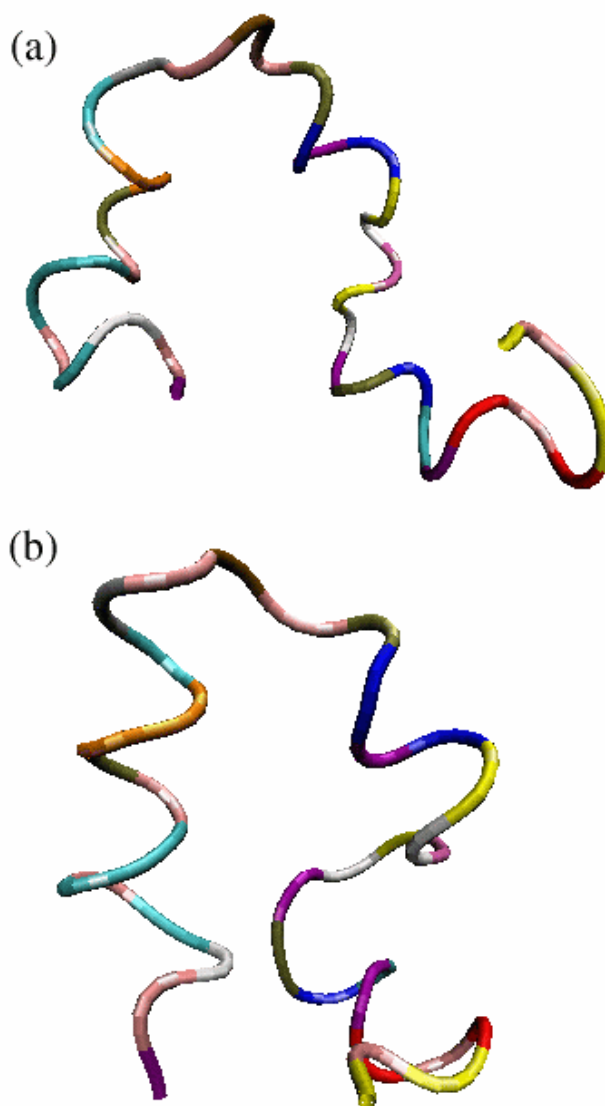
$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + (D_i / k_B T) \mathbf{F}_i(t) \Delta t + \Delta \mathbf{r}_i^G, \quad (5)$$

where each component of  $\Delta \mathbf{r}_i^G$  is taken from a Gaussian distribution with zero mean and variance  $\langle (\Delta r_{i\mathbf{a}}^G)^2 \rangle = 2D\Delta t$ .<sup>23,29</sup>  $\mathbf{r}_i(t)$  is the position of the  $i$ th atom at time  $t$  and the

**Table 2.** Basic spring constant values of the individual amino acids used in  $V_{\text{helix}}$  potential. Values obtained from a linear mapping from the helix propensities.

Amino acid	$\mathcal{K}_i$ (kJ mol <sup>-1</sup> Å <sup>-2</sup> )	Amino acid	$\mathcal{K}_i$ (kJ mol <sup>-1</sup> Å <sup>-2</sup> )
ala	17.20	trp	8.77
glu	14.45	tyr	8.08
leu	13.59	phe	7.91
met	13.07	val	6.71
arg	13.59	thr	5.85
lys	12.73	his	7.57
gln	10.49	cys	5.50
ile	10.15	asn	6.02
asp	9.80	gly	0.00
ser	8.60	pro	-37.15

systematic force on the  $i$ th atom at time  $t$  is  $\mathbf{F}_i(t)$ .  $D_i$  is the diffusion coefficient of the  $i$ th atom calculated from the Stokes–Einstein (SE) relation  $D_i = k_B T / 6\eta R_i$ .  $R_i$  is the radius of the  $i$ th atom and  $\eta$  is the viscosity of the solvent.  $k_B$  and  $T$  are the Boltzmann constant and absolute temperature, respectively. The unit of length is  $s$  (3.41 Å) and the unit of time  $t = s^2 / D_0$ .  $D_0$  is the diffusion coefficient obtained by using  $s$  as the diameter in the SE equation.  $t$  is approximately 1.2 ns in the real unit for the reduced viscosity  $\eta = 10$ . The time step  $\Delta t$  is taken equal to 0.001  $t$ .



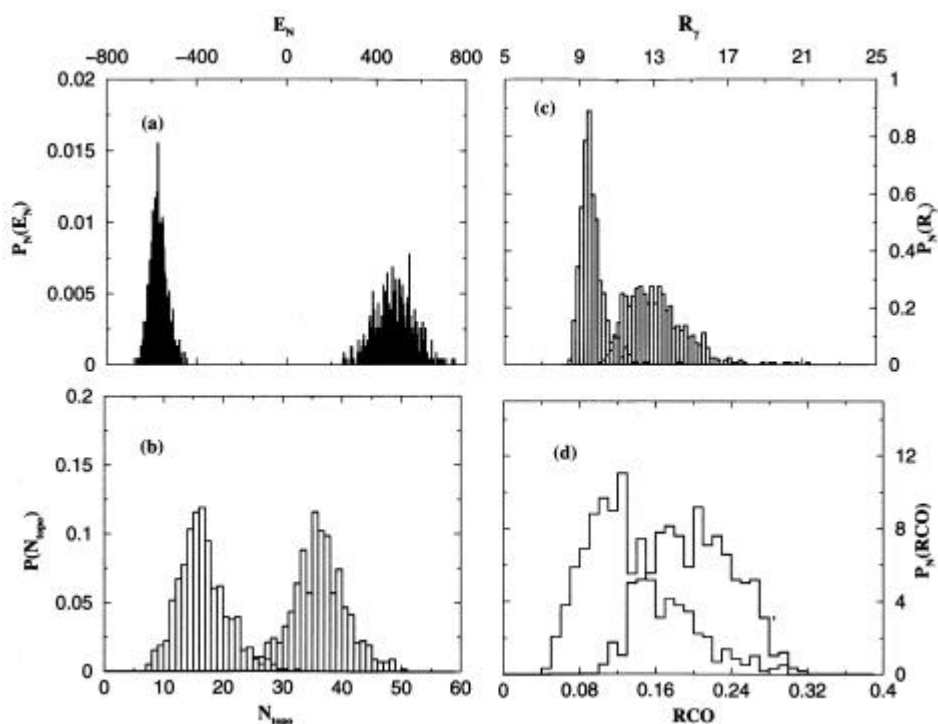
**Figure 2.** (a) Backbone structure of the model protein with lowest RMSD (4.5 Å) amongst all the 77 folded states. (b) Backbone structure of real HP-36 in the native state as obtained from its protein data bank.

#### 4. Results of the globular protein HP-36

The minimalistic model presented here shows good structural agreement with the real protein. Figure 2a shows the best folded structure obtained from our simulation with an RMSD of 4.5 Å, with the native NMR structure of the real HP36<sup>30</sup> shown in figure 2b. Note that the helices and bends of the model protein have occurred at the appropriate positions with respect to the structure of the real protein.

##### 4.1 Statistical distribution

Statistical probability distributions are obtained by performing Brownian dynamics simulations for  $\mathcal{N}$  initial configurations, where  $\mathcal{N} = 565$ . The high temperature equilibrated configurations and low temperature folded states are analysed in figure 3 to show the changes in the distributions at two different temperatures. The probability distributions of energy at high temperature equilibrated configurations and low temperature folded configurations are plotted in figure 3a. Distributions are well separated, which signify the presence of a well defined folded state. Note that the probability distribution narrows down at low temperature signifying a funnel energy landscape. Figure 3b shows the probability distribution of the number of hydrophobic topological contacts at high and low temperatures, both of which show Gaussian distribution. The distribution shifts

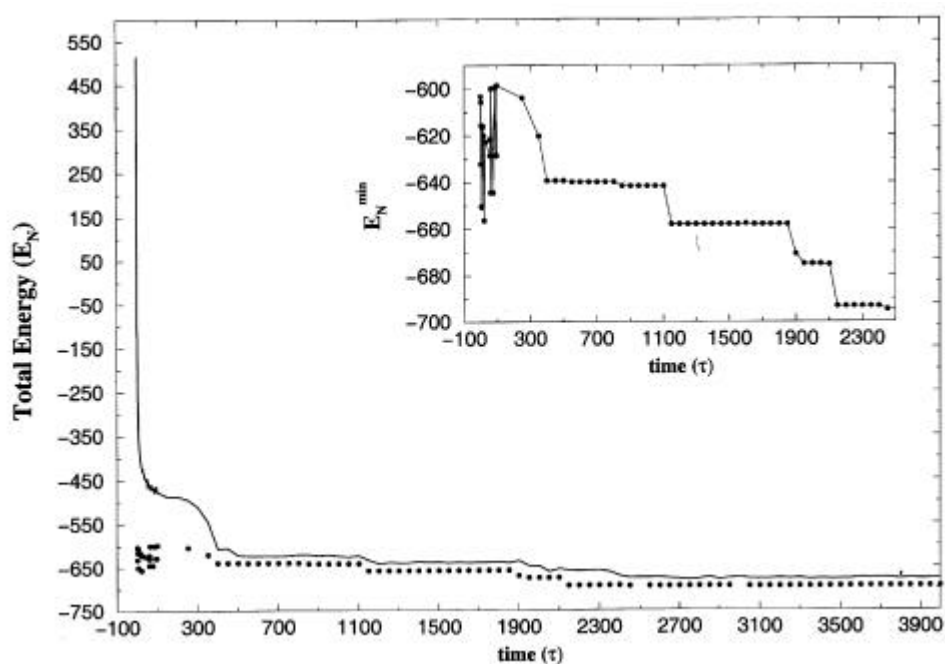


**Figure 3.** Normalised probability distribution is plotted for high temperature unfolded states (red) and low temperature folded states (blue) for different parameters: (a) total energy, (b) hydrophobic topological contact, (c) radius of gyration, (d) relative contact order parameter.

toward higher value of topological contact at low temperature. This characterises the hydrophobic core formation in the folded state. Similar behaviour can be observed in case of the probability distribution of radius of gyration plotted in figure 3c. 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  $9.6 \text{ \AA}$ , which is the experimental value of the radius of gyration of real native HP-36. Relative contact order (RCO), introduced by Baker *et al*<sup>31,32</sup>, denotes the average sequence separation for the hydrophobic residues. RCO is thus the perfect measure of the range of contact formation. Figure 3d shows the probability distribution of RCO for both high and low temperatures. Note the predominant formation of long range contacts for the low temperature folded configurations, which is absent at higher temperature.

#### 4.2 Multistage folding dynamics

A protein possesses a very complicated and highly correlated network of interaction. The late stage of folding can be thought of as the interparticle diffusion over a rugged landscape. Naturally, a wide separation of time scales may be observed in the time evolution of pair separation of the side residues, which in turn may give rise to the multistage dynamics in the macroscopic quantities such as total energy, radius of gyration, relative contact order etc. Figure 4 shows the multistage temporal decay of energy



**Figure 4.** The solid line shows the variation of energy ( $E_N$ ) with time. The circles show the minimised energies ( $E_N^{\text{min}}$ ) corresponding to a particular energy value at a certain time. Inset shows the magnified plot of the change in the minimised energy during folding.



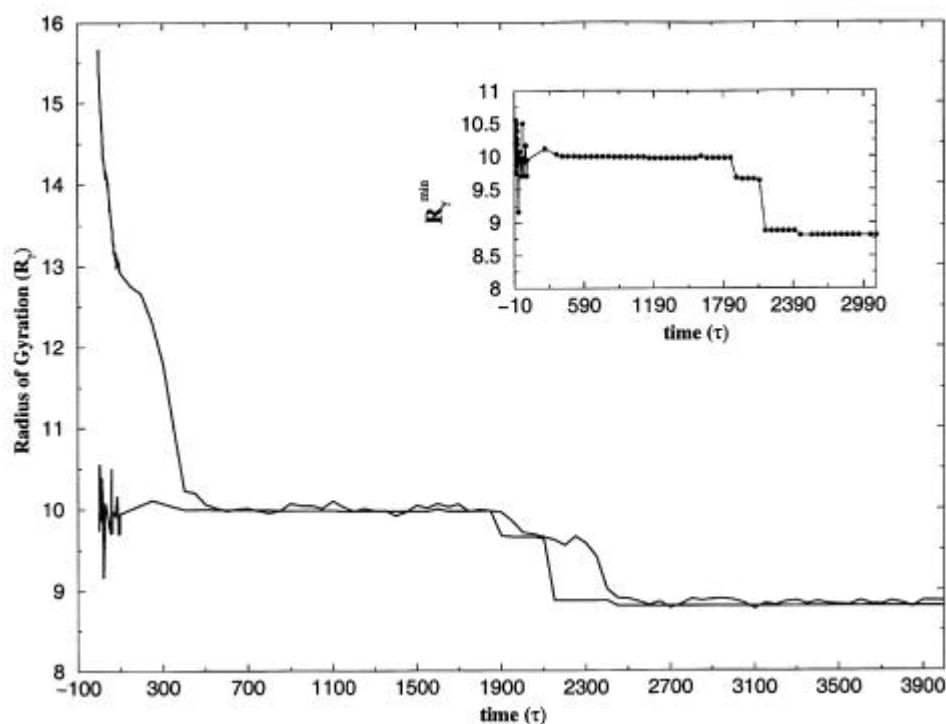
for a single *representative* folding trajectory of model HP-36, which leads to the lowest energy state. Evidence of multistage dynamics observed in our model protein strengthens from the time evolution of the energy of local minima configurations plotted in the inset of figure 4. These configurations are obtained by minimising different time evolved configurations along the path of folding. The monotonically decreasing local minima are a direct signature of multistage decay.

Figure 5 shows the similar multistage time evolution of radius of gyration. After the initial hydrophobic collapse, a long plateau is observed. This accounts for the large conformational entropy near the fully folded states, which acts as an entropic barrier near the native state.

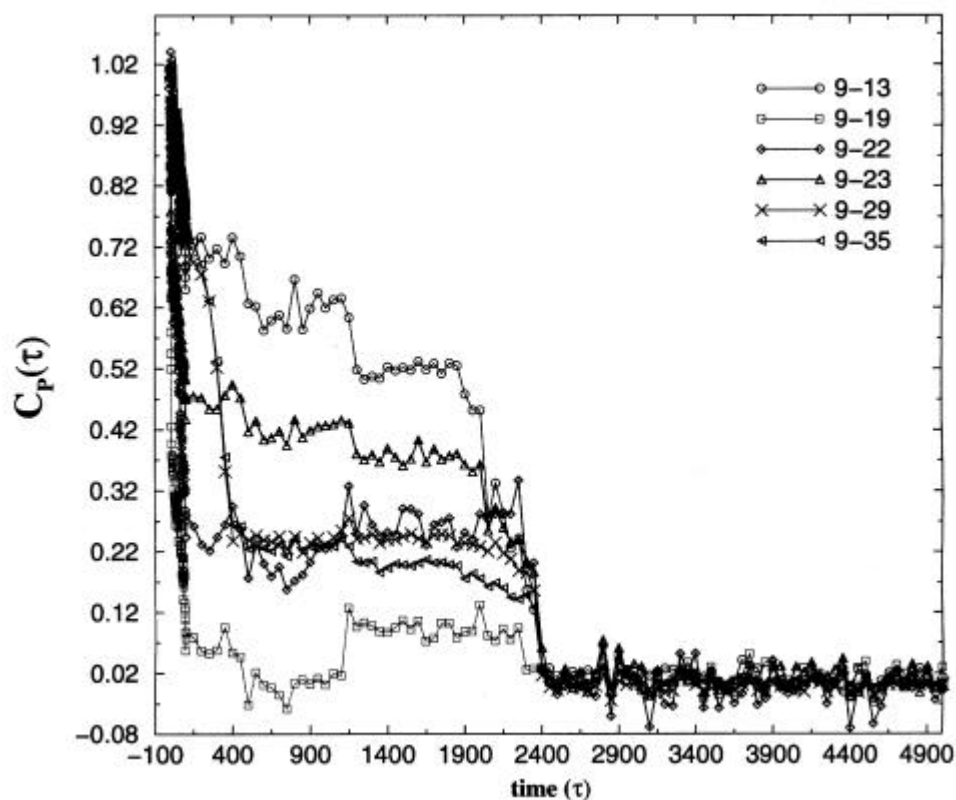
In order to understand the microscopic origin of the multistage dynamics, we have introduced a new contact pair correlation function (CPCF)  $C_p^{ij}(t)$  defined as,<sup>21</sup>

$$C_p^{ij}(t) = \frac{d^{ij}(t) - d^{ij}(\infty)}{d^{ij}(0) - d^{ij}(\infty)}, \quad (6)$$

where,  $d^{ij}(t) = r_i(t) - r_j(t)$ .  $r_i$  and  $r_j$  are the positions of the  $i$ th and  $j$ th atom, respectively.  $i$  and  $j$  can be the indices of either backbone or side residue atom. CPCF is normalised between 1 at  $t = 0$  and 0 at  $t \rightarrow \infty$ . Figure 6 shows the  $C_p^{ij}(t)$  of the 9th side residue with



**Figure 5.** The solid line shows the decrease in radius of gyration ( $R_g$ ) with time. The circles with dashed line show the  $R_g^{\min}$  for the corresponding minimised configurations.



**Figure 6.** Dynamics of contact formation of different side residues with the 9-th side residue is shown. The multistage relaxation process in the dynamical quantities originates from the diverse dynamics of the contact pairs.

many other hydrophobic side residues for the particular folding trajectory leading to the lowest energy state. The three stages of the folding process are reflected by mainly three different dynamical behaviour seen amongst the side residues. Side residues closed to the tagged one collapse very fast. Some show an initial shoulder and only a few show the plateau in the long time *that correlates with the similar plateau observed in case of the other dynamical quantities*. The vastly different time scales of pair separation are reflected in the multistage dynamics of macroscopic quantities plotted in figures 4 and figure 5.

### 5. Folding and association of **A<sub>b</sub>**amyloids

The structural resemblances and multistage dynamics are observed also in case of other proteins such as **A<sub>b</sub>** and its different fragments. Here we discuss the aspects of folding and association of **A<sub>b</sub>** which have become the cynosure of research in the area of misfolding related diseases.

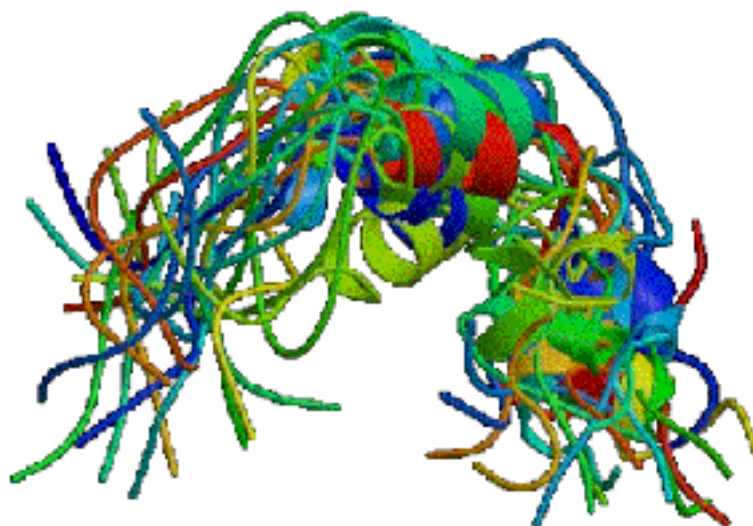
**A<sub>b</sub>** is a small protein without large stability for the native state. It has no unique native structure. The interatomic distances, obtained from 2-D NMR, and the constrained mini-

randomisation result in 20 different structures<sup>33</sup> obtained from protein data bank<sup>34</sup> as shown in figure 7. Individually all the native states show a sequence of secondary structure: 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 **b**bend in the middle (around residue 24–28) for all the folded forms.

Brownian dynamics simulations of the model **A**b**** protein yield low temperature states, some of which are shown in figure 8a. Here also note the same pronounced **b**bend. The two structures are statistically quite similar. Figure 8b shows the conformations of the model **A**b**** equilibrated at a high temperature. The high temperature configurations assume a bent-rod structure.<sup>35</sup>

We propose a microscopic model of aggregation and subsequent deposition of **b** amyloid protein. Basic physical concepts that determine the aggregation of **A**b**** are: (i) hydrophobic interactions (with nucleation sites) which decrease the energy and (ii) the entropic contribution, which can be increased by randomisation of the hydrophilic tail. These two factors can lead to a rich and complex free energy landscape.<sup>6</sup> The dimer itself can have two alternating arrangements as shown in figure 9.<sup>35</sup> In these two arrangements,  **$\pi$** stacking interaction among phenylalanine groups and other such strong hydrophobic interactions would favour parallel arrangement, but at the cost of entropy loss at the hydrophilic chain end. The hydrophilic tails would favour anti-parallel arrangement due to the larger accessible degrees of freedom. However, in the anti-parallel arrangement, the hydrophobicity remains frustrated. *Such arrangements have indeed been predicted and observed in the rod-coil diblock copolymers,*<sup>36</sup> which lend support to our argument.

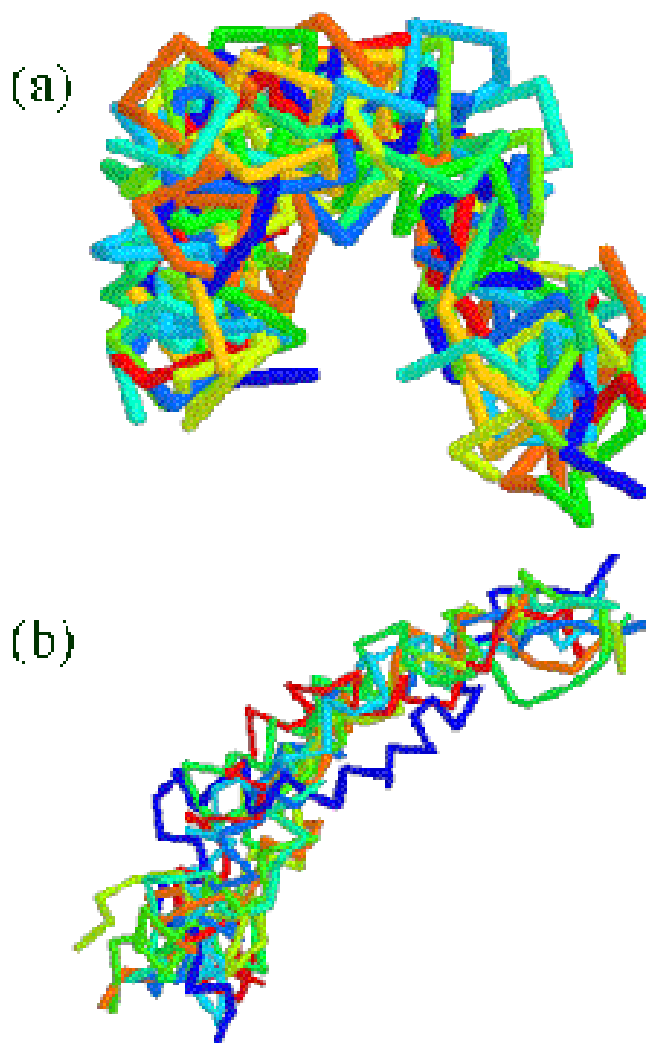
Growth of two different dimers can take up three identifiable isomeric forms of trimer. A tetramer can have four distinct forms. One should also consider the free energy of solvation of these species. By using its analogy with the diblock copolymers and using



**Figure 7.** Collection of reported folded structures of **A**b**** (PDB code 1AML) obtained from NMR. Note the **b** turn around 24–28 amino acid residues.

simple expression of entropy of a random flight polymer inside a tube,<sup>37</sup> we can write the approximate free energy of an  $n$ th-mer aggregate in parallel and anti-parallel arrangement as given below,<sup>35</sup>

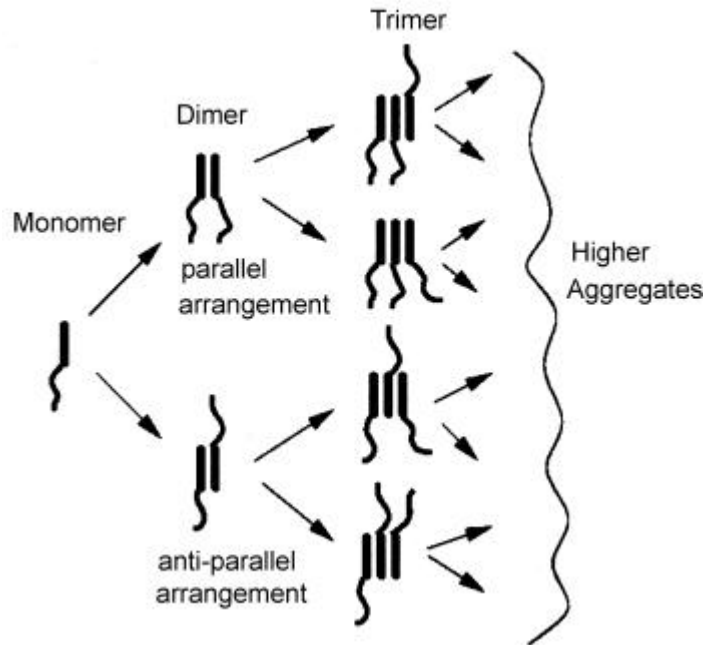
$$\frac{G_1}{k_B T} \simeq -\frac{S_L(0)}{k_B} + \frac{\Delta H_{sol}^{(1)}}{k_B T},$$



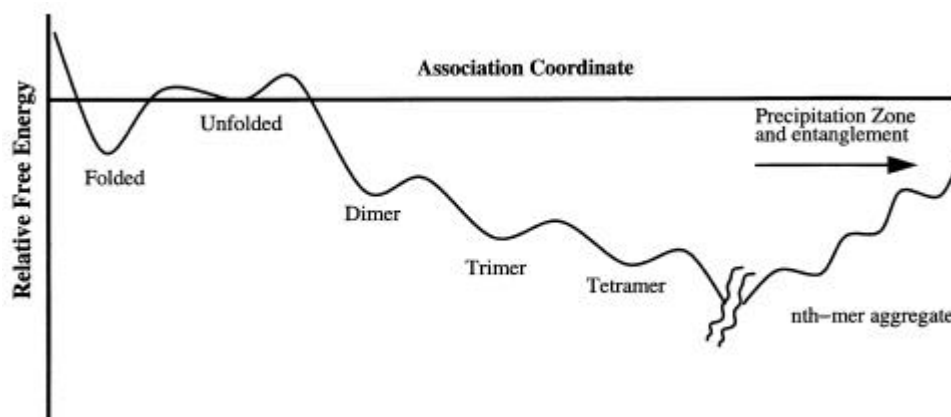
**Figure 8.** (a) The 10 different folded (quenched) backbone structures (plotted with Rasmol software) of the model **Ab** obtained from Brownian dynamics simulations with the minimalistic model. The **b** turn is clearly visible. (b) The configurations at high temperature show an extended bent-rod like structure.

$$\begin{aligned}
 \frac{G_n^{\uparrow\downarrow}}{k_B T} &\simeq -(n-1)L_r \mathbf{e}^{\uparrow\downarrow}(T) - \frac{nS_{L_c}(0)}{k_B} \left[ \frac{(n-4)}{k_B T} \right] \times \\
 &H(n-4) \left[ \frac{3}{2} \left( \frac{L_c}{L_{0c}} \right)^2 + \mathbf{a}_0 \left( \frac{L_{0c}}{4D} \right) \right] + \frac{\Delta H_{sol}^{(n)}}{k_B T}, \\
 \frac{G_n^{\uparrow\uparrow}}{k_B T} &\simeq -(n-1)L_r \mathbf{e}^{\uparrow\uparrow}(T) - \frac{nS_{L_c}(0)}{k_B} \left[ \frac{(n-2)}{k_B T} \right] \times \\
 &H(n-2) \left[ \frac{3}{2} \left( \frac{L_c}{L_{0c}} \right)^2 + \mathbf{a}_0 \left( \frac{L_{0c}}{2D} \right) \right] + \frac{\Delta H_{sol}^{(n)}}{k_B T}, \tag{7}
 \end{aligned}$$

where  $G_1$  is the free energy of an unfolded monomer,  $G_n$  is the free energy of  $n$ th-mer aggregate,  $L_r$  is the length of the hydrophobic rod and  $L_c$  is the length of the hydrophilic coil ( $L = L_r + L_c$ ),  $S_L$  and  $S_{L_c}$  denote the entropy in free space for the whole monomer and the hydrophobic tail respectively,  $H(x)$  is the Heaviside step function which is unity for its positive argument and zero otherwise,  $L_{0c}$  is the Gaussian length of the hydrophilic tail,  $D$  is the distance between two hydrophilic chains in parallel arrangement,  $\mathbf{a}_0$  is a



**Figure 9.** Schematic representation of the proposed aggregation model. The possibility of both parallel and anti-parallel arrangement is shown (see (7) and (8)).

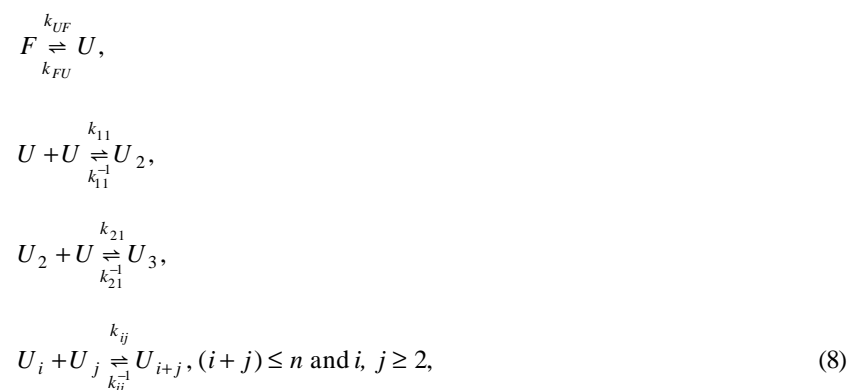


**Figure 10.** Schematic representation of the potential energy diagram of  $A\beta$  amyloid and its aggregates.

certain numerical factor which depends on the shape of the cross-section of the tube,  $\epsilon_s$  are the temperature dependent hydrophobic interactions per unit length of the rod and  $\Delta H_{sol}^{(n)}$  is the interaction enthalpy for the  $n$ -th-mer aggregate with the surrounding solvent. We find from the hydrophathy scale that  $\Delta H_{sol}^{(1)}$  is about  $2-3 \text{ kcal mol}^{-1}$ .  $G_1$  can be negative due to entropy contributions.

Different combinations of parallel and anti-parallel arrangements of an  $n$ th-mer aggregate will lie in between the values of free energies of complete parallel and anti-parallel arrangements given above. According to the above expressions of free energy, higher aggregates can be stabilised by favourable enthalpy contributions from the rod part. However, as the aggregate grows, the *surface tension due to frustrated hydrophobicity and entanglement among the hydrophilic tails can reduce the free energy of an  $n$ th-mer aggregate, leading to the precipitation of the aggregate.* The free energy diagram for association is shown schematically in figure 10.

The reaction scheme for reversible association is written below,



where,  $F$  and  $U$  are the concentrations of folded and unfolded monomer respectively.  $U_i$  is the concentration of the  $i$ th-mer aggregate and  $k$ 's denote the rates of reaction. A

similar kind of scheme for irreversible aggregation has recently been given by Robert,<sup>38</sup> in the general context of protein association.

## 6. Conclusion and future problems

In the present work, we have carried out Brownian dynamics simulations and theoretical analysis of folding of small proteins using minimalistic models introduced recently by us.<sup>21</sup> With the use of hydropathy scale and the helix propensity of the amino acids, the folded structures of the model proteins show good agreement with the structure of the native state of the real protein. Time evolution of the dynamical variables such as energy, radius of gyration, relative contact order etc. shows multistage decay. We have also discussed the possible mechanisms of folding and association of non globular protein **A<sub>1</sub>B<sub>1</sub>**. The model could be easily employed to study many other proteins such as **L**-repressor, **b** barrel, three-helix bundle protein etc. The model can be directly used to study the kinetics of amyloid association. Research in this direction is in progress.

## Acknowledgements

We thank Department of Science and Technology, Govt. of India and the Council of Scientific and Industrial Research, Department of Atomic Energy, New Delhi for financial support.

## References

1. Anfinsen C B 1973 *Science* **181** 223; Sela M, White F H and Anfinsen C B 1957 *Science* **125** 691
2. Levinthal C 1969 in *Mossbauer spectroscopy in biological systems* (eds) P Debrunner, JCM Tsibris and E Munck (Urbana: University of Illinois Press) p. 22
3. Zwanzig R, Szabo A and Bagchi B 1992 *Proc. Natl. Acad. Sci. USA* **89** 20
4. Dill K A, Alonso D O V and Hutchinson K 1989 *Biochemistry* **28** 5439
5. Yue K and Dill K A 1993 *Phys. Rev. E* **48** 2267; Chan H S and Dill K A 1990 *Proc. Natl. Acad. Sci. USA* **87** 6368
6. Bryngelson J D and Wolynes P G 1987 *Proc. Natl. Acad. Sci. USA* **84** 7524; *J. Phys. Chem.* **93** 6902
7. Leopold P E, Montal M and Onuchic J N 1992 *Proc. Natl. Acad. Sci. USA* **89** 8721
8. Bryngelson J D, Onuchic J N and Succi N D and Wolynes P G 1995 *Proteins: Structure, function and genetics* **21** 167
9. Nymeyer H, Succi N D and Onuchic J N 2000 *Proc. Natl. Acad. Sci. USA* **97** 634
10. Portman J J, Takada S and Wolynes P G 1998 *Phys. Rev. Lett.* **81** 5237
11. Portman J J, Takada S and Wolynes P G 2001 *J. Chem. Phys.* **114** 5069, 5082
12. Dill K A 1990 *Biochemistry* **29** 7133
13. Levitt M and Warshel A 1975 *Nature (London)* **253** 694
14. Levitt M 1976 *J. Mol. Biol.* **104** 59
15. Srinivas G and Bagchi B 2002 *J. Chem. Phys.* **116** 8579
16. Srinivas G and Bagchi B 2003 *Theor. Chem. Acts.* **109** 8
17. Pillardy J, Czaplowski C, Liwo A, Lee J, Ripoll D R, Kazmierkiewicz R, Oldziej S, Wedemeyer W. K, Gibson K D, Armutova Y A, Saunders J, Ye Y-J and Scheraga H A 2001 *Proc. Natl. Acad. Sci. USA* **98** 2329; Lee J, Liwo A and Scheraga H A 1999 *Proc. Natl. Acad. Sci. USA* **96** 2025
18. Duan Y and Kollman P A 1998 *Science* **282** 740
19. Hansmann U H E and Wille L T 2002 *Phys. Rev. Lett.* **88** 68105
20. Kang J, Lemaire H G, Unterbeck A, Salbaum J M, Masters C L, Grzeschik K H, Multhaup G, Beyreuther K and Muller-Hill B 1987 *Nature (London)* **325** 733

21. Mukherjee A and Bagchi B 2003 *J. Chem. Phys.* **118** 4733
22. Liwo A, Oldziej S, Pincus M R, Wawak R J, Rackovsky S and Scheraga H A 1997 *J. Comput. Chem.* **18** 850
23. Hansen J P and McDonald I R 1986 *Theory of simple liquids* (London: Academic Press)
24. Kyte J and Doolittle R F 1982 *J. Mol. Biol.* **157** 105
25. Zimm B H and Bragg J K 1959 *J. Chem. Phys.* **31** 526
26. Pace C N and Scholtz J M 1998 *Biophys. J.* **75** 422
27. Chou P Y and Fasman G 1974 *Biochemistry* **13** 211
28. Mooij G C A M, Frenkel D and Smit B 1992 *J. Phys. Condens Matter* **4**L255; Frenkel D and Smit B 1992 *Mol. Phys.* **75** 983
29. Ermak D L and McCammon J A 1978 *J. Chem. Phys.* **69** 1352
30. McKnight C J, Doering D S, Matsudaria P T and Kim P S 1996 *J. Mol. Biol.* **260** 126
31. Plaxco K W, Simons K T and Baker D 1998 *J. Mol. Biol.* **277** 985
32. Grantcharova V, Alm E J, Baker D and Horwich A L 2001 *Curr. Opin. Struct. Biol.* **11** 70
33. Sticht H, Bayer P, Willbold D, Dames S, Hilbich C, Beyreuther K, Frank R W and Rosch P 1995 *Eur. J. Biochem.* **233** 293
34. Bernstein F C, Koetzle T F, Williams G J B, Meyer E F, Brice M D, Rodgers J R, Kennard O, Shimanouchi T and Tasumi M 1977 *J. Mol. Biol.* **112** 535
35. Mukherjee A and Bagchi B *J. Chem. Phys.* (accepted)
36. Williams D R M and Fredrickson G H 1992 *Macromolecules* **25** 3561
37. de Gennes P G 1979 in *Scaling concepts in polymer physics* (Ithaca, NY: Cornell University Press); Doi M and Edwards S F 1986 in *Theory of polymer dynamics* (Oxford: Clarendon Press)
38. Roberts C J 2003 *J. Phys. Chem.* **B107** 1194