

Conformation of a coarse-grained protein chain (an aspartic acid protease) model in effective solvent by a bond-fluctuating Monte Carlo simulation

R. B. Pandey¹ and B. L. Farmer²

¹*Department of Physics and Astronomy, University of Southern Mississippi, Hattiesburg, Mississippi 39406-5046, USA*

²*Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio 45433, USA*

(Received 12 November 2007; published 5 March 2008)

In a coarse-grained description of a protein chain, all of the 20 amino acid residues can be broadly divided into three groups: Hydrophobic (H), polar (P), and electrostatic (E). A protein can be described by nodes tethered in a chain with a node representing an amino acid group. Aspartic acid protease consists of 99 residues in a well-defined sequence of H , P , and E nodes tethered together by fluctuating bonds. The protein chain is placed on a cubic lattice where empty lattice sites constitute an effective solvent medium. The amino groups (nodes) interact with the solvent (S) sites with appropriate attractive (PS) and repulsive (HS) interactions with the solvent and execute their stochastic movement with the Metropolis algorithm. Variations of the root mean square displacements of the center of mass and that of its center node of the protease chain and its gyration radius with the time steps are examined for different solvent strength. The structure of the protease swells on increasing the solvent interaction strength which tends to enhance the relaxation time to reach the diffusive behavior of the chain. Equilibrium radius of gyration increases linearly on increasing the solvent strength: A slow rate of increase in weak solvent regime is followed by a faster swelling in stronger solvent. Variation of the gyration radius with the time steps suggests that the protein chain moves via contraction and expansion in a somewhat quasiperiodic pattern particularly in strong solvent.

DOI: [10.1103/PhysRevE.77.031902](https://doi.org/10.1103/PhysRevE.77.031902)

PACS number(s): 87.14.E-, 05.10.Ln, 87.15.A-, 87.15.H-

I. INTRODUCTION

The coarse-graining [1–18] appears almost unavoidable in modeling the global structure and dynamics of proteins no matter which descriptions, all atomic [7–11], minimalist [12–17], or a combination thereof, one may choose to use. The past couple of decades has witnessed an upsurge of interest in studying the universal structures and complex response properties including relaxation pathways to folding of proteins. The variety of proteins and their universal and unique characteristics are so rich that probing even the general characteristics of such stimuli driven macromolecular systems are far from complete. For example, a protease plays a crucial role for a human immunodeficiency virus (HIV) to integrate into the host cell and exploit the intricate molecular pathways to multiply [19]. Search for protease inhibitors has led to a considerable success in assembling a cocktail of drugs to alleviate the pain and sufferings for many AIDS patients [20]. Therefore, it would be interesting to probe how a protease moves, conforms, and responds to such stimuli as the quality of solvent.

Neri *et al.* [18] have recently studied the structure of two proteins of the aspartic acid protease class near their native state using a coarse-grained model involving some of the atomistic details with a combination of molecular dynamics simulation and *ab initio* calculations. Such attempts are useful in examining part of the localized structures [21,22] by different computer simulation models [23–25] to predict such characteristics as binding of a protein (protease) to specific target site, folding and dimerization. In order to investigate the large scale structural pathways and global dynamics, alternative approaches are, however, highly desirable for such a class of complex protein. In this article, we would like to follow the well-established route of investigating the ho-

mopolymer chains [26–29] and present a computer simulation study of an aspartic acid protease by a coarse-grained bond-fluctuation model [30]. Unlike a common polymer chain, the protein chains are much more complex. Each node of a coarse-grained protein chain can be one of the 20 amino acid groups each with its unique specificity. Twenty amino acid groups are grouped into three categories: Hydrophobic (H), polar (P), and electrostatic (E) for further simplifications. Interaction of nodes in solvent and their sequences are important in orchestrating their local segmental dynamics and structural relaxation with pushing and pulling of nodes as the intrachain constraints and its interaction with the surrounding solvent compete.

A number of models have been proposed to understand the mechanism of protein relaxation. As in studies of polymer, most of the models for protein chains involve discrete lattice [12–17] and off-lattice [7–11]. Despite the limited degrees of freedom with the constant bond model on a discrete lattice, Dill and co-workers [13] have successfully described the core assembly and protein folding into native structure via funnel pathways with an *HP* protein model (a minimalist approach). Off-lattice approaches (including all atomic methods) [7–11] are often used to probe such specific configurations as alpha helices and beta sheets especially in folding of small proteins [31–33]. Large degrees of freedom are very important in probing the structural fluctuations around known stable structures. The multiscale dynamics and related relaxation pathways from one structure to another is however severely limited in proteins with hundreds of amino acids due to long relaxation time with atomistic details. In order to reproduce known characteristics (information based approach), some constraints are usually invoked, e.g., peptide group dipole interaction [34] in addition to standard interactions and constraints.

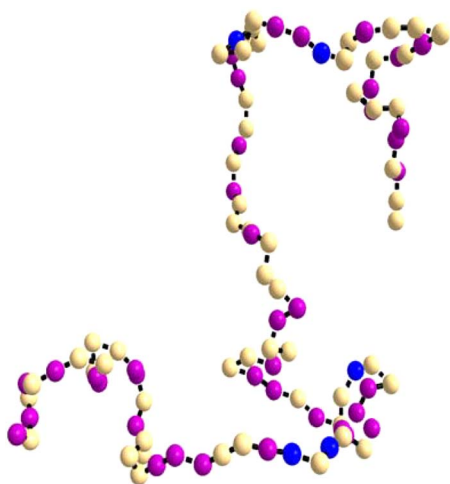


FIG. 1. (Color online) Coarse grained description of aspartic acid protease (1DIFA) with H (pink), P (gold), and E (blue).

Constant bond model on discrete lattices are successfully used to probe stability of secondary and tertiary structures of coarse-grained models of proteins with somewhat unphysical constraints on bond angles and interactions among the nodes (the approximate representation of amino-acid groups) [15]. In general, the degrees of freedom are considerably enhanced in going from constant bond to bond-fluctuating (BF) models for a polymer chain [29,30]. The bond-fluctuation model has been recently used to study the conformational relaxation into native structure [17] of a general HP protein chain and even a specific protein, sensory rhodopsin I [16] without severe constraints. The advantages of this approach over the methods used in probing the folding dynamics are computational efficiency, covering large scales, and avoid unphysical constraints; the disadvantage is the lack of atomistic details. We would like to employ BF description here to examine the conformation of aspartic acid protease in a range of solvent media. The model is presented in next section followed by result and discussion with a conclusion at the end.

II. MODEL

The aspartic acid sequence consists of 99 residues in a specific sequence [35,36]. With a coarse-grained grouping of 20 amino acid groups into three categories [36] of H , P , and E , a protease configuration can be represented as in Fig. 1. Note that there are 39 hydrophobic groups, 55 polar nodes, and 5 electrostatic nodes. We consider a cubic lattice of size L^3 with $L=30-100$. The protein chain is placed on the lattice with each node occupying a unit cube and the distance between the consecutive nodes restricted to 2, $\sqrt{5}$, $\sqrt{6}$, 3, and $\sqrt{10}$ as in bond-fluctuation model for a chain [29,30] to impose the excluded volume constraints.

The empty lattice sites constitute the effective solvent medium. Nodes interact with the solvent (S) sites with interaction energy, $\varepsilon_{HS} = -\varepsilon_{PS} = \varepsilon$, where the interaction strength ε can be varied to control the quality of the solvent; $\varepsilon=0-4$ are used in this study with $\varepsilon_{ES}=0$. The range of interaction up to $r^2=6$ with the same interaction strength is used. Unlike

the constrained models for the protein chain [15] with selective node-node interactions, there is no node-node interaction except the excluded volume effects. Each node executes its stochastic movement to one of its 26 neighboring sites (adjacent positions) [29,30,37], with the Metropolis algorithm, i.e., with probability $\exp(-\Delta E/k_B T)$ where ΔE is the change in energy between new and old configuration of the node, k_B is the Boltzmann constant, and T is the temperature; $k_B T=1$ is used in this study. Note that only an isolated particle (node) has 26 neighbors to choose from; a node in the chain has lower number of possible adjacent positions to move successfully due to presence of the other segment of the chain. Attempts to move each node once defines the unit Monte Carlo step (MCS) time [37].

During the simulation we keep track of the mean square displacements of the center node and that of the center of mass of the chain, radius of gyration, and end-to-end distance. Simulation is performed for a sufficiently long time (10^7 MCS) either to reach equilibrium conformation (where ever feasible) or to identify its relaxation trend. It is difficult to reach equilibrium in stronger solvent. Further, a number of independent runs are used to estimate the average of the physical quantities. Obviously, this is a simplified representation of the protease and many specific details would not be meaningful at this stage. For example, identifying the binding affinity of specific segment of the protease with a particular reactant (or agent) would be hard to assess without exploring the general characteristics first and identifying the range of parameters (e.g., interaction strength) for its appropriate structural response.

III. RESULTS AND DISCUSSION

Most of the simulations are performed on 30^3 and 60^3 lattices with a number (50–750) of independent samples. Some data are also generated on a 100^3 lattice and the finite size effects are discussed accordingly. In general, data with larger samples have lower fluctuations. Note that the statistical fluctuations (generally larger with smaller samples) in data points with the MC simulations do not necessarily imply the finite size effect unless they affect the qualitative behavior of the physical system. If the size of the lattice becomes smaller than the characteristic correlation length, i.e., the radius of gyration of the chain here, then the finite size effects become severe. In that case a comprehensive finite size analysis [37] must be performed as appropriate. This is not the case in our study and therefore we resort to 60^3 samples for most of our data sets unless stated otherwise. One should however check for the finite size effects regardless.

A. General observation

As mentioned above, the aspartic acid protease (Fig. 1) contains 39 hydrophobic groups, 55 polar groups, and only five nodes with electrostatic characteristics. Can the conformation of the protein with a specific sequence of 99 nodes differ from a chain of the same size but with nodes in different sequence or order? In order to address this question along with exploring the effects of solvent on the general charac-

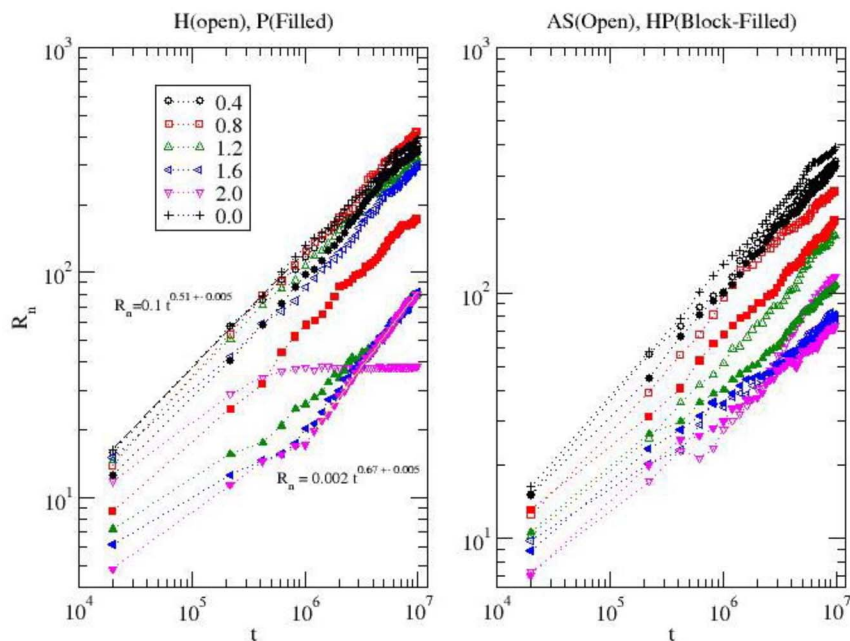


FIG. 2. (Color online) Variation of the rms displacement of the center node with the time step t for hydrophobic (H) and polar (P) chains (left) and HP block-copolymer and aspartic acid (AS) chain (right) for various solvent media characterized by the interaction strengths ($\varepsilon=0.0-2.0$). Lattice size 60^3 with 50 independent samples is used for each chain.

teristics of the aspartic acid protease, we consider four types of chains. Since the five electrostatic nodes appear to have rather weak interactions [36], we consider only their excluded volume effects and leave them at the same sequence positions (Fig. 1) in all simulations presented here. Types of chains considered are (i) H chain with 94 hydrophobic nodes, (ii) P chain with 94 polar nodes, (iii) HP chain with a diblock copolymer consisting of 39 H nodes and 55 P nodes, and (iv) aspartic acid protease with a specific sequence of H and P nodes (as in Fig. 1). If we consider only excluded volume interaction of each node (i.e., $\varepsilon=0$), the sequence of nodes become irrelevant and the chain behaves as a self-avoiding walk. The interaction strength is varied with $\varepsilon=0.0-4.0$ the effects of which should be reflected in the segmental dynamics and conformation of the protease.

Visual inspections of the snapshots reveal that the structure of hydrophobic chain (i) converges to a stable globular conformation, a nearly spherical shape in the asymptotic time limit especially at higher values of the interaction energy, i.e., with $\varepsilon=2.0$. The hydrophilic chain (ii) on the other hand opens up and remains highly extended. The di-block copolymer (iii) exhibits an interesting asymptotic structure with a part of chain (with H groups) conforms to a compact structure anchoring the remaining segment (with P groups) in extended configurations. The aspartic acid protease (iv) conforms to a range of structures which depend on the quality of the solvent, i.e., the magnitude of the interaction strength. The bond-fluctuation model nevertheless captures the general characteristics of its constitutive segments based on their sequence (iv) of tethering different from that of the homo-polymers consisting of H and P chains [(i), (ii), and the HP diblock copolymer (iii)]. The quantitative analysis of the conformation and dynamics of the protease chain is presented in the following sections.

Using the bond-fluctuation model, Khokhlov and Khaltur [38] have already attested to somewhat similar observation on the distinct physical properties of a specific sequence

of H and P nodes, different from that of the HP random copolymer. They have constructed a stable globular conformation formed by an *ad hoc* attractive interaction between the nodes of a homopolymer. They identified the nodes on the surface and those in the interior and ascribed them hydrophilic and hydrophobic to design a chain with a specific sequence of H and P groups. They find that coil to globule transition as a function of temperature of this sequenced chain is different from that of a HP random copolymer. This work deals with only attractive interaction among hydrophobic groups apart from excluded volume constraints without the solvent. The specific sequence is generated from identifying the surface and interior nodes in a globular configuration generated by an *ad hoc* attractive interaction between each node of a homopolymer—a reverse problem of designing the chain from a known end result, i.e., a globular conformation. We consider a protein chain with specific sequence, the bio-chemical properties of which are well known [36]. Including the interaction of each node (H and P) with the solvent apart from the excluded volume constraint in our coarse description of this protein chain enhances the specificity of the conformation and dynamics of its sequence as described below.

B. Dynamics

As mentioned above, we keep track of the root mean square (rms) displacement of the center of mass of the chains (R_c) and that of its center node (R_n) and the radius of gyration as a function of time steps. Figure 2 shows, the variation of the rms displacement of the center node of the four chain types [(i)–(iv)]. How fast do these chains move? We can estimate the dynamics by analyzing the corresponding data with appropriate power laws, i.e., if the variation of the rms displacement with time step t is described by

$$R_{n(c)} = At^\nu,$$

where A is a constant, then the magnitude of the exponent ν characterizes the dynamics. The hydrophobic and polar

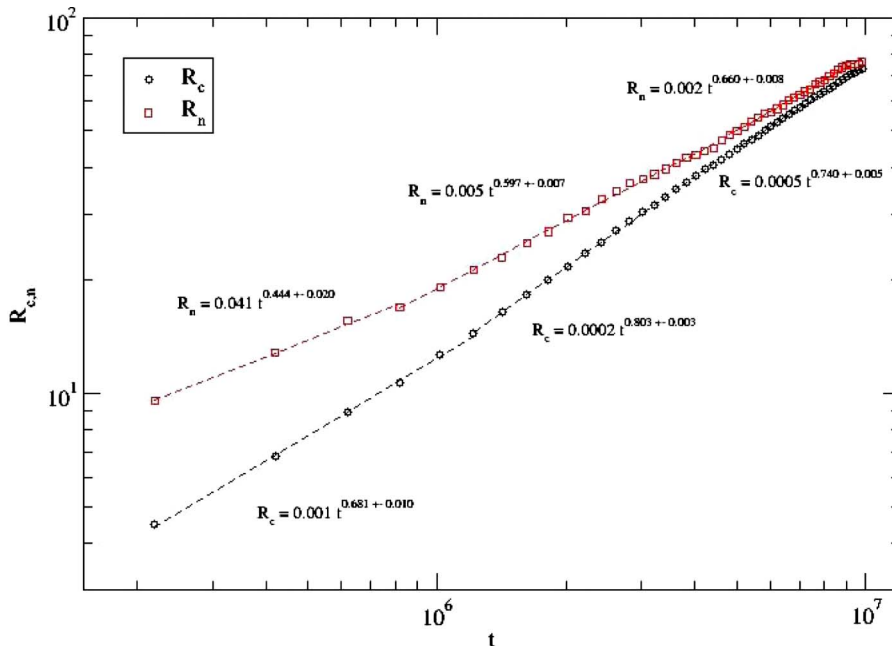


FIG. 3. (Color online) Variation of the rms displacement of the center node (R_n) and center of mass (R_c) of the polar chain on a 100^3 lattice at $\varepsilon=2.0$ with 200 independent samples.

chains exhibit diffusive behavior, i.e., $\nu=\frac{1}{2}$ with relatively weak solvent interactions ($\varepsilon=0.0-1.6$). Increasing the solvent interaction strength to $\varepsilon=2.0$, freezes the hydrophobic chain but accelerates the chain with polar groups to a superdiffusive behavior with $\nu=2/3$. The block-copolymer chains and the aspartic acid chain continue to move with similar speed at higher interaction strengths.

There are three type of asymptotic motion for the center node: (1) No-motion ($\nu=0$), (2) diffusive [with the power-law exponent ($\nu=\frac{1}{2}$)], and (3) superdiffusive ($\nu>\frac{1}{2}$) motion depending on the type of chain, its interaction with the solvent, and the time regime. For the homopolymer polar chain, the center node clearly shows this asymptotic exponent $\nu \approx 2/3$, as further evidenced by the following data with different sample sizes. Variations of R_n and R_c with t for $\varepsilon=2.0$ on a larger 100^3 and a smaller 30^3 lattice are presented

in Figs. 3 and 4, respectively. The data on the larger sample (Fig. 3) show a systematic approach of the power-law exponent to $\nu \approx 2/3$ in the long time regime for both the center node as well as the center of mass of the chain. Note how R_n and R_c are approaching the same global dynamics in the asymptotic time regime, a general feature expected for a chain.

Data for small sample (Fig. 4) show that a superdiffusive ($\nu \approx 2/3$) power-law dynamics of the node is followed by a somewhat slower dynamics presumably approaching to a diffusive behavior in the long time regime. Why do dynamics on the small sample (Fig. 4) differ from that with the larger sample (Fig. 3) in the long time? Let us digress for the moment to dig deeper into the finite size effect. Perhaps our long time is too long for this sample (Fig. 4). Since the minimum distance between the consecutive nodes is two in

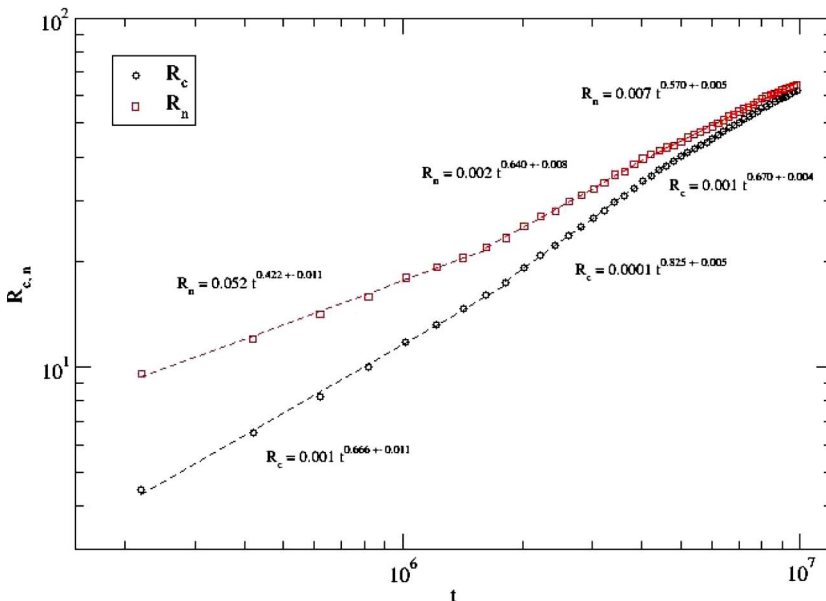


FIG. 4. (Color online) Variation of the rms displacement of the center node (R_n) and center of mass (R_c) of the polar chain on a 30^3 lattice at $\varepsilon=2.0$ with 500 independent samples.

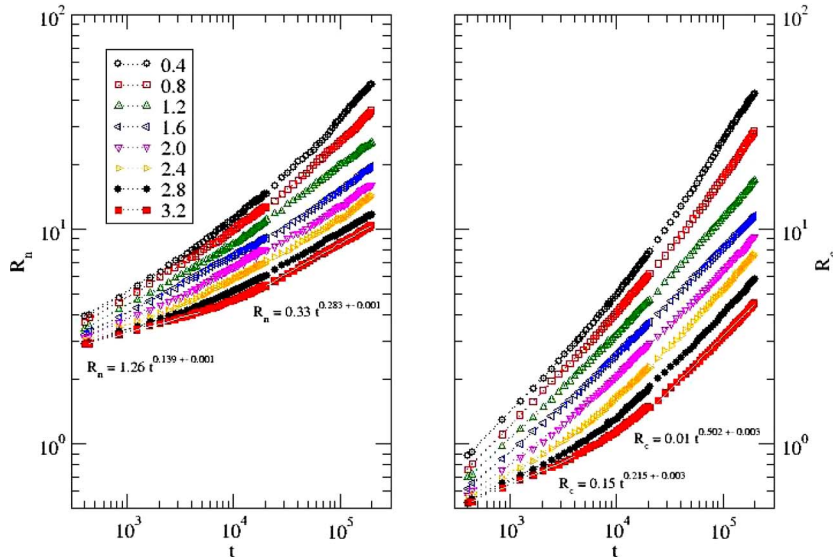


FIG. 5. (Color online) Variation of the rms displacement of the center node (R_n , left) of aspartic acid chain and that of its center of mass (R_c , right) with the time step t for various solvent media characterized by the interaction strengths ($\varepsilon=0.0-3.2$). Lattice size 60^3 with 50–2000 independent samples is used in each solvent medium.

unit of the lattice constant, a chain with 99 nodes can occupy at least 198 consecutive lattice sites which are relatively large on a 30^3 lattice. Further, the distance traveled by chain should be of the order of the lattice size in order to avoid finite size effects.

A close examination of Fig. 4 reveals that, for the distance of the order of lattice size, the segmental motion of the center node is indeed superdiffusive ($\nu \approx 2/3$). In longer times, chain travels such a long distance that it traverses across the sample about two times. If the simulations were continued further for longer times such that the chain would have traversed the lattice many times, i.e., a random walk back and forth across the sample, the chain would behave like a point particle. Thus the random walk motion of the chain in the long time becomes diffusive ($\nu \approx 1/2$). The approach to such long time diffusion is faster for aspartic acid protease on a 30^3 lattice (see below). The effect of the medium and the detail structure of the chain (i.e., the local sequence of H and P) become a part of the pre-factor (A) of the power-law dynamics. The asymptotic long time diffusive behavior of the P chain is due to finite size effect. How long is long is dictated by the size of the sample and the chain among other factors such as correlation among constituents. However, for samples of size 60^3 and larger the asymptotic power-law behavior of R_n and R_c over the span of our observation time steps (10^7) is independent of the sample size.

Despite fluctuations in the data points (Fig. 2), it is possible to distinguish the segmental dynamics of four different types of chains particularly at higher interaction strength. Data with smaller (Fig. 4) and larger (Fig. 3) samples illustrates the finite size effects and that the power-law dependence of the rms displacement of the node presented in Fig. 2 is independent of the sample size.

Let us examine the dynamics of the aspartic acid chain more closely from the variation of the RMS displacement of the center node of the chain (Fig. 2) and that of its center of mass as presented in Fig. 5. A systematic change in dynamics from short time to long time regime is clearly seen on increasing the strength of the solvent interaction. First, there is a crossover from short time slow dynamics to long time fast

dynamics. At a relatively high solvent interaction strength ($\varepsilon=3.2$), the center node exhibits a very slow segmental dynamics with exponent $\nu \approx 0.14$ in the short time regime followed by a subdiffusive motion with $\nu \approx 0.28$. Corresponding global motion of the center of mass shows a crossover from a subdiffusive ($\nu \approx 0.22$) dynamics in short time to diffusion ($\nu = \frac{1}{2}$) before accelerating to superdiffusion in the long time regime (see below). Thus the quality of solvent plays a significant role in orchestrating the segmental dynamics of the chain particularly at high interaction strength in short to intermediate time regimes.

How do these dynamics and crossover appear in the long time and what is the effect of the lattice size? We present the variation of R_n and R_c with the time step t for $\varepsilon=3.2$ on a larger 100^3 and a smaller 30^3 lattice samples in Figs. 6 and 7, respectively. The asymptotic long time power-law dependence of both the center node and the center of mass of the aspartic acid protease seem to approach the same superdiffusive behavior with $\nu \approx 2/3$ (Fig. 6). Trends for the intermediate crossovers from the short time (Fig. 5) seem to continue (Fig. 6) as the slow moving segments (nodes) catch up with the faster moving center of mass of the chain. The finite size effects are clearly visible with the 30^3 sample (Fig. 7) especially in the asymptotic long time regime when the chain has crossed the sample back and forth reaching its diffusive behavior. As discussed above (Fig. 4), for an straight chain extending about 200 consecutive lattice sites, 30^3 lattice becomes too small for the chain to cross it many times in the long time. The intermediate segmental and global power-law dynamics seem however comparable (compare Figs. 6 and 7).

C. Conformation

Variation of the radius of gyration (R_g) of the hydrophobic, polar, block-copolymer, and aspartic acid chains with time steps is presented in Fig. 8. The magnitude of R_g reaches its equilibrium value for all chains in different solvent media considered here, some with more fluctuations than others. As noticed before, the hydrophobic chains con-

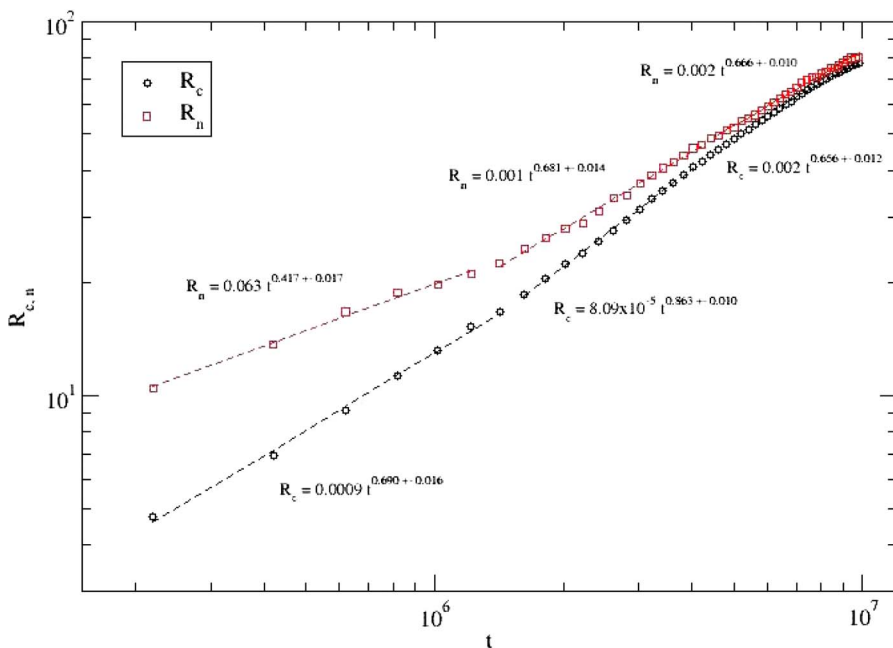


FIG. 6. (Color online) Variation of the rms displacement of the center node (R_n) and center of mass (R_c) of the aspartic acid chain at $\epsilon=3.2$ on a 100^3 lattice with 200 independent samples.

tract and the polar chains expand. Increasing the interaction strength enhances the change in R_g .

The *HP* block-copolymer chains remain compact with relatively less variation in their radius of gyration on changing the quality of solvent. Opposite is the pattern for the aspartic acid chain which responds more strongly on increasing the interaction strength between the nodes (the amino groups) and the solvent. Figure 8 shows clearly that the protease chain swells more than the polar chain. This is a unique feature of *H* and *P* nodes in a specific sequence. The presence of hydrophobic group seems to provide local anchoring which may translate into intrachain (local) stiffness. As a result the chain becomes more extended than the corresponding (a relatively more flexible) polar homopolymer. Although the data are more fluctuating, a systematic pattern seems to

appear in variation of R_g with the time step (as follows).

The average value of the radius of gyration R_{gs} in equilibrium can however be evaluated along with statistical errors. Figure 9 shows the variation of the equilibrium value of R_g with the solvent interaction strength. The aspartic acid chain expands or swells, i.e., the radius of gyration increases on increasing the solvent interaction strength. Increase of R_{gs} implies expansion or swelling of the chain. Clearly the expansion is linear with the interaction strength in relatively weaker solvent interaction regime ($\epsilon \approx 0.0-1.2$) and nonlinear at higher interaction strengths ($\epsilon \approx 1.2-2.0$). Such a wide variation in R_g with the interaction strength implies that secretion of chemical ingredients for controlling the quality of solvent is very important in orchestrating the structure of the aspartic acid chain.

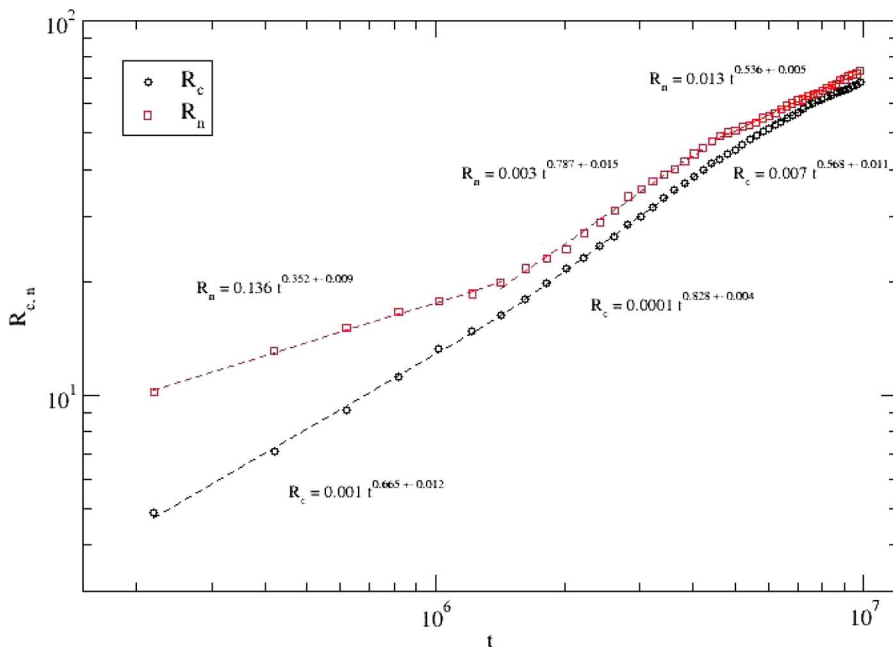


FIG. 7. (Color online) Variation of the rms displacement of the center node (R_n) and center of mass (R_c) of the aspartic acid chain on a 30^3 lattice at $\epsilon=3.2$ with 200 independent samples.

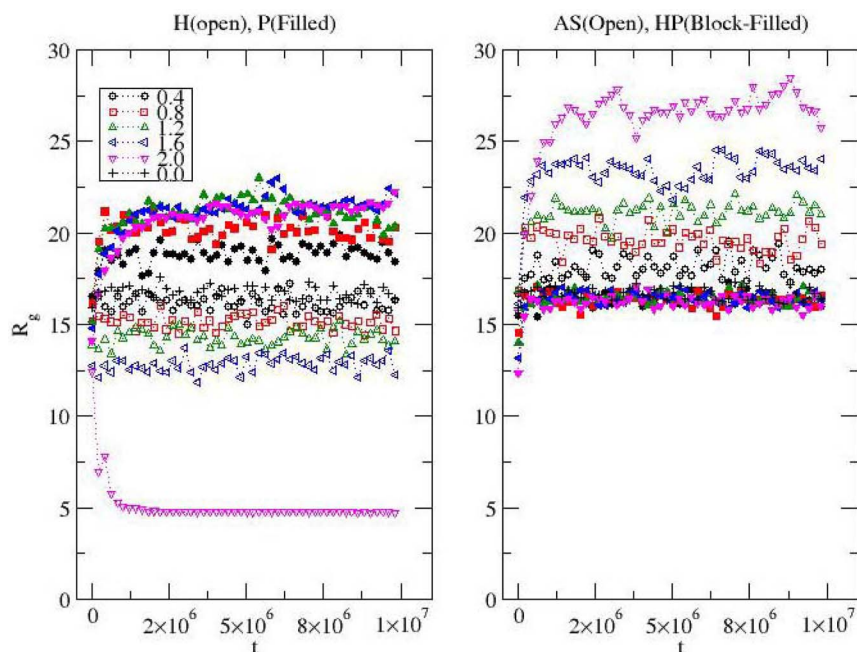


FIG. 8. (Color online) Variation of the radius of gyration (R_g) with the time step t for hydrophobic (H) and polar (P) chains (left) and HP block-copolymer and aspartic acid chain (right) for various solvent media characterized by the interaction strengths ($\epsilon = 0.0-2.0$). Lattice of size 60^3 with 50 independent samples is used for each chain.

The structure of protein chain is crucial in controlling its functions. How the structure changes as the protease executes its stochastic moves is also important in understanding its dynamic response to external stimuli. Insets in Fig. 9 show the variation of the radius of gyration (R_g) of the chain and its end-to-end distance (R_e) in the long time regime in a relatively strong solvent medium. The oscillatory behavior of R_g and R_e seems to suggest that chains contract and expand over the time scales that depend on the quality of the solvent. Further, the temporal variations of the radius suggest that contraction and expansion resemble an earthworm motion than a slithering snake reptation [28–30].

One may argue that the random oscillation of the radius of gyration may be just the random fluctuation around its average or steady-state value (R_{gs}) over the time steps in equilibrium.

However, the intermediate steps through which the chain relaxes, provides insight into the structural changes during the course. Figures 10 and 11 show the direct variation of R_g with the time step without statistical averaging. As seen the fluctuations in R_g though random have some persistence involving cooperative and collective segmental response to structural relaxation. We believe that the “earthworm” movement with contraction (minima) and expansion (maxima) of R_g is due to cooperative motion of its different sequenced (block) regions. For example, we know that the radius of gyration (R_g) of the polar chain is larger than that of the hydrophobic chain. However, R_g of the aspartic acid protease (a specific sequence of H , P , and E) is still larger than that of the polar chain and it depends on the quality of the solvent (see Fig. 8). The polar groups are pulled by the

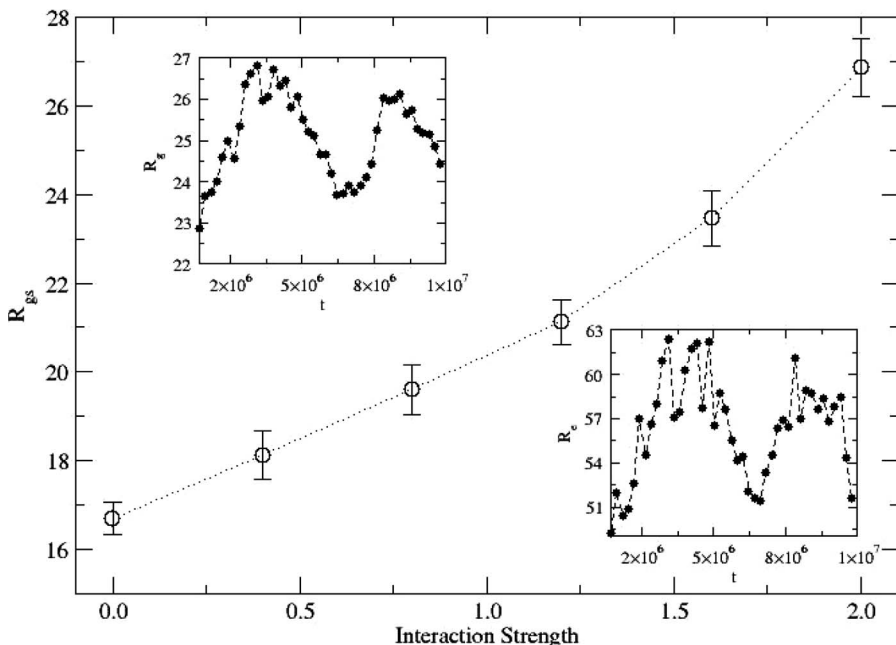


FIG. 9. Equilibrium radius of gyration (R_{gs}) vs interaction strength. Inset shows the variation of the radius of gyration (R_g) of the aspartic acid chain and its end-to-end distance (R_e) in the long time regime for the solvent interaction ($\epsilon=3.2$). Statistics is the same as Figs. 2–4.

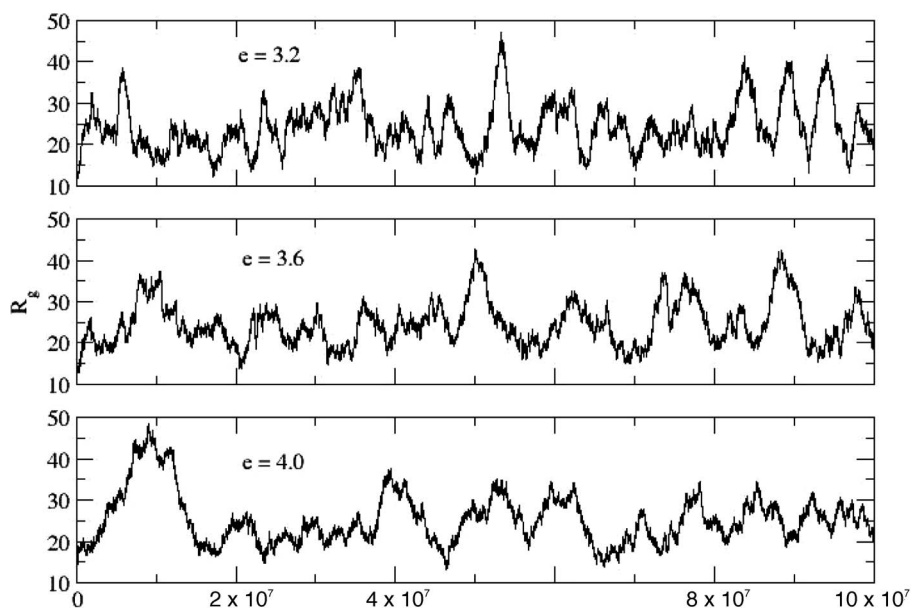


FIG. 10. Variation of the radius of gyration (R_g) with the time step (t) on a 100^3 sample in a single run with time interval of 10^2 between consecutive data points at different $\epsilon = e$.

solvent while hydrophobic groups act as local anchors which enhance the local stiffness of the chain. The dynamic changes in segmental configurations lead to global contraction and expansion—the “earthworm” motion, a cooperative response due to competition between H and P sequences and their dynamic environment.

IV. SUMMARY AND CONCLUSIONS

Conformation and dynamics of a chain consisting of 99 nodes in a specific sequence of hydrophobic, polar, and electrostatic characteristics, a coarse-grained representation of an aspartic acid protease are studied using Monte Carlo simulations. The bond-fluctuation model is used for the coarse-grained description of the chain on a cubic lattice to capture the efficiency with appropriate degrees of freedom. The Metropolis algorithm is used to perform stochastic movement of

tethered nodes as in computer simulation of polymer chains instead of imposing brute-force constraints [15,34] on interaction and local configuration on nodes to describe such specific structures as secondary and tertiary structures. This study is focused on the general characteristics of structural evolution and segmental dynamics of the chain in a range of effective solvent environment, different from the studies on protein folding [1–15,31–34]. In order to verify the effectiveness of capturing the specificity of the sequence, independent simulations are also carried out with the (i) hydrophobic, (ii) polar, and (iii) di-block-copolymer chains, and results are compared with that of the aspartic acid protease (iv) at each solvent environment i.e., interaction strength. The aspartic acid protease chain with a specific sequence of H , P , and E nodes shows distinct conformational and dynamical properties particularly at higher solvent interaction strength, a feature generally not tested in protein modeling [1–15,31–34].

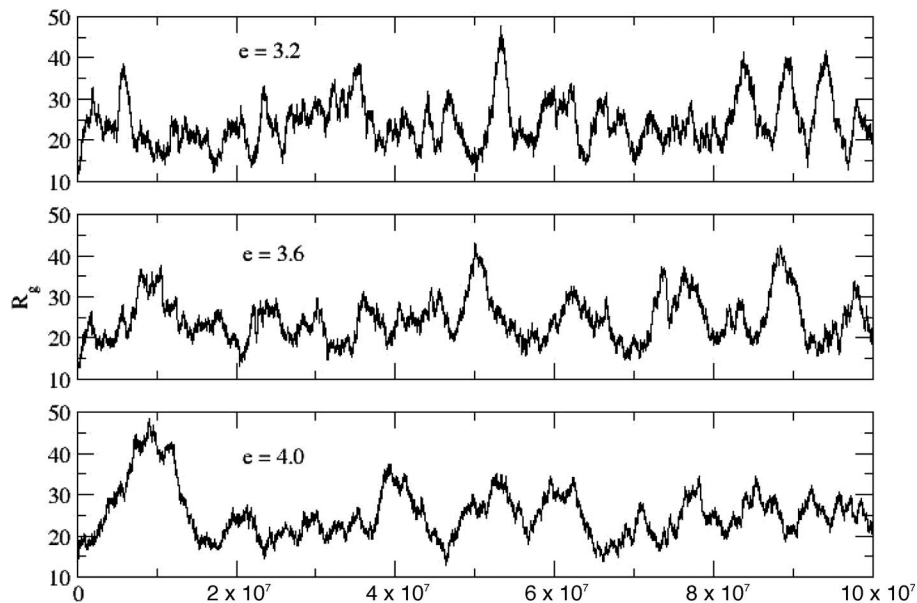


FIG. 11. Variation of the radius of gyration (R_g) with the time step (t) on a 100^3 sample in a single run with time interval of 10 between consecutive data points at different $\epsilon = e$.

Because of the repulsive (attractive) interaction between the hydrophobic (polar) nodes and solvent, the segments with overwhelmingly hydrophobic nodes are less mobile and tend to anchor the local structure. The polar groups are relatively more mobile and provide segmental dynamics and structure. Specific sequence of these nodes is thus capable of capturing the specific function of a protein in appropriate solvent environment. We have observed a relatively slow segmental dynamics in the short time and diffusive motion in the asymptotic long time regime. The crossover relaxation depends on the strength of the solvent interaction. The difference in segmental dynamics is amplified at higher values of the solvent interaction. For example, the center node of the aspartic acid chain exhibits a very slow subdiffusive motion characterized by the exponent $\nu \approx 0.14$ in the short time followed by a somewhat faster but still subdiffusive dynamics $\nu \approx 0.28$ with the solvent interaction $\varepsilon = 3.2$. Obviously, it takes very long time for the chain to reach its superdiffusive dynamics in such a strong solvent. The radius of gyration, however, reaches its equilibrium values in the asymptotic regime in all solvent conditions we have studied.

As described above, the aspartic acid chain expands, i.e., the radius of gyration increases on increasing the solvent interaction strength. The rate of expansion or swelling is linear in weaker solvent regime and accelerates nonlinearly in stronger solvent. Chains contract and expand over the time scales (longer at higher interactions) that depend on the quality of the solvent. The temporal variations of radius of gyration suggest that contraction and expansion resemble an earthworm motion than a slithering snake reptation. The segmental dynamics and structure resulting from the cooperative motion of H , P , and E nodes appear more general than the specific structures encountered in all atomic models for small proteins [31–33]. Nevertheless, the cooperative dynamics and structure observed is shown to be a unique characteristic of the specific sequence of H , P , and E groups.

ACKNOWLEDGMENTS

Support from the Materials and Manufacturing Directorate of the Air Force Research Laboratory, the Air Force Office of Scientific Research (AFOSR) is gratefully acknowledged.

-
- [1] Z. Bagci, A. Kloczkowski, R. L. Jernigan, and I. Bahar, *Proteins* **53**, 56 (2003).
- [2] H. Behringer, A. Degenhard, and F. Schmid, *Phys. Rev. Lett.* **97**, 128101 (2006).
- [3] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, *Science* **267**, 1619 (1995).
- [4] M. Skepo, P. Linse, and T. Arnebrant, *J. Phys. Chem. B* **110**, 12141 (2006).
- [5] B. Gillespie and K. W. Plaxco, *Annu. Rev. Biochem.* **73**, 837 (2004).
- [6] J. R. Banavar and A. Maritan, *Rev. Mod. Phys.* **75**, 23 (2003).
- [7] U. H. E. Hansmann, *Comput. Sci. Eng.* **5**, 64 (2003).
- [8] Y. Zhou and M. Karplus, *Nature (London)* **401**, 400 (1999).
- [9] E. J. Sorin and V. S. Pande, *Biophys. J.* **88**, 2472 (2005).
- [10] M. Shen and K. F. Freed, *Proteins* **49**, 439 (2002).
- [11] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, L. Barry, B. L. Stoddard, and D. Baker, *Science* **302**, 1364 (2003).
- [12] J. Skolnick and A. Kolinsky, *J. Mol. Biol.* **221**, 499 (1991).
- [13] K. M. Fiebig and K. A. Dill, *J. Chem. Phys.* **98**, 3475 (1993).
- [14] B. Gertsman and Y. Garbourg, *J. Polym. Sci., Part B: Polym. Phys.* **36**, 2761 (1998).
- [15] P. P. Chapagain and B. Gertsman, *Biopolymers* **81**, 167 (2006).
- [16] C.-M. Chen and C.-C. Chen, *Biophys. J.* **84**, 1902 (2003).
- [17] J. Bjursell and R. B. Pandey, *Phys. Rev. E* **70**, 052904 (2004).
- [18] M. Neri, C. Anselmi, M. Cascella, A. Maritan, and P. Carloni, *Phys. Rev. Lett.* **95**, 218102 (2005).
- [19] M. Miller, J. Schneider, B. K. Sathyanarayan, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. Kent, and A. Wlodawer, *Science* **246**, 1149 (1989).
- [20] S. G. Deeks, F. M. Hecht, M. Swanson, T. Elbeik, R. Loftus, P. T. Cohen, and R. M. Grant, *AIDS* **13**, F35 (1999).
- [21] S. Piana, P. Carloni, and M. Parrinello, *J. Mol. Biol.* **319**, 567 (2002).
- [22] M. Cascella, C. Micheletti, U. Rothlisberger, and P. Carloni, *J. Am. Chem. Soc.* **127**, 3734 (2005).
- [23] M. A. Neri, M. Baaden, V. Carnevale, C. Anselmi, A. Maritan, and P. Carloni, *BioPhys. J. BioFast* September 7 (2007).
- [24] E. Jenwitheesuk and R. Samudrala, *BioMed Central, BMC Struct. Biol* (2003).
- [25] Y. Levy, A. Caffish, J. N. Onuchic, and P. G. Wolynes, *J. Mol. Biol.* **340**, 67 (2004).
- [26] P. G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
- [27] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, New York, 1986).
- [28] M. Rubinstein and R. H. Colby, *Polymer Physics* (Oxford University Press, New York, 2003).
- [29] *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, edited by K. Binder (Oxford University Press, New York, 1995).
- [30] I. Carmesin and K. Kremer, *Macromolecules* **21**, 2819 (1988).
- [31] Y. Wei, W. Nadler, and U. H. E. Hansmann, *J. Chem. Phys.* **128**, 025105 (2008).
- [32] O. Zimmermann and U. H. E. Hansmann, *Biochim. Biophys. Acta* **1784**, 252 (2008).
- [33] T. Herges and W. Wenzel, *Phys. Rev. Lett.* **94**, 018101 (2005).
- [34] A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *Protein Sci.* **2**, 1697 (1993).
- [35] Website <http://www.bmrwisc.edu/referenc/hydroph.html> (as of 02/05/2007).
- [36] Sequence of aspartic acid proteinase 1DIFA <http://molsun1.cbrc.aist.go.jp/papia/cgi/simplifiedisplay.pl?id=seq.18240> (as of 02/05/2007).
- [37] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, UK, 2000).
- [38] A. R. Khokhlov and P. G. Khalatur, *Phys. Rev. Lett.* **82**, 3456 (1999).