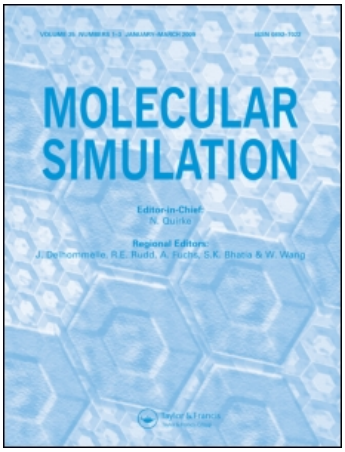

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Coarse grain molecular dynamics simulation for the prediction of tertiary conformation of lysozyme adsorbed on silica surface

Xiaoyu Wu[a]; Ganesan Narsimhan[a]
[a] Biochemical and Food Process Engineering, Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, USA

To cite this Article Wu, Xiaoyu and Narsimhan, Ganesan(2009) 'Coarse grain molecular dynamics simulation for the prediction of tertiary conformation of lysozyme adsorbed on silica surface', Molecular Simulation, 35: 10, 974 — 985
To link to this Article: DOI: 10.1080/08927020903015338
URL: http://dx.doi.org/10.1080/08927020903015338

Molecular Simulation
Vol. 35, Nos. 10–11, September 2009, 974–985

Taylor & Francis
Taylor & Francis Group

Coarse grain molecular dynamics simulation for the prediction of tertiary conformation of lysozyme adsorbed on silica surface

Xiaoyu Wu and Ganesan Narsimhan*

Biochemical and Food Process Engineering, Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907, USA

(*Received 22 December 2008; final version received 4 May 2009*)

A coarse-grain (CG) algorithm is developed for molecular dynamics (MD) simulation of proteins in which the polypeptide backbone as well as side chains were mapped into spherical interaction centres. The force field parameters for non-bonded interaction between these centres (beads) were evaluated from the interaction of all atoms corresponding to these beads using pairwise additivity. The validity of CG algorithm was demonstrated by comparing the potential energy, radius of gyration (RG), end-to-end distance and root-mean-square deviation with those obtained by all-atom (AA) simulation for a small protein molecule Trp-cage as well as for lysozyme in solution. The CG simulation for lysozyme in solution requires less than 1/50 CPU-time compared to that for the AA method. Also, the CG method converges to equilibrium potential energy much faster than AA simulation. CG simulation for lysozyme adsorbed on silica surface showed that the molecule is more unfolded with a less compact tertiary structure compared to that in the solution (with higher RG, end-to-end distance and the projected area on silica surface), this effect being more pronounced at higher temperature. A higher ionic strength resulted in a more extended structure for the lysozyme on silica surface.

Keywords: lysozyme; molecular dynamics simulation; all atom molecular dynamics simulation; coarse grain molecular dynamics simulation; silica surface; tertiary conformation

1. Introduction

How a protein reaches its folded, biologically active conformation or other states remains elusive despite years of intense research [1–4]. To map this folding reaction we need to characterise all states along the pathway [such as native state (N), intermediate state (I) and denatured state (D)], as well as the mechanism of the conversion between these states. Such characterisation cannot always be determined using experiments. All-atomic molecular dynamics (MD) simulations provide an alternative approach to acquire such information which complements the experimental data, thereby allowing one to map the dynamic reaction coordinate at the atomic level [5]. Moreover, all-atomic MD simulation can be readily applied to elucidate kinetic pathways by reconstructing accurate pathways from energy surfaces [6], and provides links between structure and dynamics by enabling the exploration of the conformational energy landscape accessible to protein molecules [7–10]. However, in most cases, all-atomic MD simulation is suitable for polypeptides or for mimicking the early stages of the protein folding process, and realistic MD simulations for folding of a large protein molecule are still not currently possible due to the limited computation efficiency [a real protein unfolding takes $\mu s \sim ms$, while a general simulation step is less than 1 fs (10^{-15} s)]. A larger protein (possessing a large number of atoms) has a large number of atomic interactions that need to be calculated for every simulation step and thus needs a much longer computation time for every simulation step, not feasible by today's computer technology. In order to simplify the calculations, three-dimensional lattice Monte Carlo simulation algorithms have been developed to investigate the behaviour of amino acids in the solution as well as at interfaces [11–13]. It is necessary to develop a coarse-grain (CG) simulation algorithm that captures the essential features of molecular interactions in order to make the MD simulation feasible for large protein molecules.

The CG methodology generally involves grouping together selected clusters of atoms into single macrosites to significantly reduce the number of interactions calculated and, hence, also the computational cost. A large number of models have been developed for different systems and with differing degrees of simplification [13–17]. In this paper, we build CG models of silica surface and lysozyme that retain an explicit connection with the chemical identity of the systems described. The principle of our CG model is based on the work of Marrink et al. [14] and Basdevant et al. [15]. On average, an approximate four-to-one mapping of heavy atoms (i.e. not including hydrogen) is used to represent a CG single interaction centre. As described in the reference, only four

*Corresponding author. Email: narsimha@purdue.edu

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020903015338
http://www.informaworld.com

main types of interaction sites are considered: polar (P), non-polar (N), apolar (C) and charged (Q). The polar sites represent neutral groups of atoms that would easily dissolve in water, apolar sites represent hydrophobic moieties and non-polar groups are used for mixed groups that are partly polar. Charged sites are reserved for ionised groups. For particles of type N and Q, four subtypes (0, d, a and da) are further distinguished. The subtypes allow fine-tuning of the interactions on the basis of the chemical nature of the atoms, which are represented by the CG groups. Subtype 0 applies to groups in which no hydrogen bonding capabilities exist, d and a to groups that could act as a hydrogen-bond donor or acceptor, respectively, and da to groups with both donor and acceptor options (as shown in Method Section).

In a MD simulation study, a representative sampling over the entire phase space is needed to obtain an accurate canonical distribution at a given temperature. For large molecules such as proteins, this sampling is usually difficult, especially at a physiological temperature, because molecules tend to be trapped in a large number of local energy minima, which slows the sampling of phase space. In protein folding simulations, it is usually difficult to obtain accurate canonical distributions at low temperatures by conventional simulation methods because simulations at low temperatures tend to get trapped in one of a large number of local minimum-energy states. One way to overcome this multiple-minima problem is to perform a simulation based on non-Boltzmann probability weight factors so that a random walk in energy space may be realised, which is called replica exchange molecular simulation (REMD) [18].

Lysozyme is an enzyme which hydrolyses the polysaccharides found in many bacterial cell walls [19]. Lysozyme is a compact globular protein with 129 residues, consisting of five α helices, 1 three-stranded antiparallel β sheet and a large amount of random coil and β turns [20–22]. Also its structure is stabilised by four disulphide bonds [23], with most of the cysteins located in the α helices. Recent reviews describe the folding and unfolding pathways of lysozyme in solution [24–26]. Some investigations have employed tryptophan fluorescence [27] to characterise the change in the tertiary conformation of proteins using the extent of blue shift of the emission spectrum. The secondary structures of proteins were monitored using Fourier transform infrared spectroscopy (FTIR) [28–30] and circular dichroism (CD) [31,32]. Hydrophobic [29] and electrostatic interactions [28] between the protein and adsorbing surface have been shown to influence the changes in the secondary structure. Adsorption onto silica nanoparticle was shown to result in a greater loss of α-helix content of lysozyme for larger surface coverage [32]. We have investigated the kinetics of tertiary conformational changes as well as the extent of secondary conformational changes of lysozyme on silica nanoparticle surfaces using tryptophan fluorescence and CD [33]. The tertiary conformational changes were shown to depend on surface packing density, pH and ionic strength and were influenced by protein–protein interactions on the surface. The rate of unfolding was found to be higher at lower surface concentrations, pH 4, higher ionic strengths, higher 2,2,2-trifluoroethanol and dithiothreitol concentrations. CD spectra showed that the α-helix content was lower for the adsorbed lysozyme compared to the bulk with a corresponding increase in β sheet and random coil. This decrease in α helix was found to be more pronounced at lower surface concentrations.

In this paper, we present a modified approach of CG algorithm which employs the REMD method. Multiple replicas are used for each temperature level and exchanges between these replicas are also tried, eliminating the synchronisation needed in the original REMD method. This multiplexed-replica exchange molecular dynamics (MREMD) method is tested with a small model protein (Trp-cage) starting from a fully unfolded state both in all-atomic and CG algorithms. After the validation, we also use lysozyme as a model to test the CG algorithm in solution by comparison of the results with all atomic simulation. Finally, the CG algorithm is employed to simulate the unfolding of lysozyme on the silica surface.

2. Methods

All the simulations were run using AMBER 10.0 program [34] where the AMBER ff99 force field [35] was employed with the exception of Φ/Ψ dihedral parameters which were refitted [36]. The initial linear structures of Trp-cage (NLYIQWLKDG GPSSGRPPPS) were built up by the LEap module included in AMBER 10.0 package. Minimisation of the structure to remove bad contact and steric hindrances was performed for 1000 steps before starting the MD calculations. At a constant temperature, the MD simulations were fully unrestrained in canonical ensemble using Sander module and separated into two stages. In the first stage, a 50 ps calculation with a 0.5 fs time-step was performed as the system was slowly heated to 325 K at 1 ps coupling constant for temperature control [37]. Following the first heating stage, a 100 ns simulation with a 2 fs time-step was run at 325 K with 1 ps coupling constant [36,38,39]. The GB/SA implicit solvent model was employed to incorporate the solvent effect [40,41]. The trajectories of the simulation were collected for every 500 ps. For REMD simulation, 16 replicas were used with temperatures of 269.5, 285.5, 300.0, 316.5, 334.0, 351.0, 371.8, 391.0, 413.9, 435.0, 460.7, 485.0, 512.9, 535.9, 570.9 and 595.2 K, respectively. Neighbour lists were utilised and updated every 2.5 fs. To maintain the temperatures at the chosen levels and the pressure at 1 bar, Berendsen weak coupling [37] was used, with coupling constants of 0.1 ps for the temperature and 1 ps for the pressure. Constraints were used for bond lengths

using the LINCS algorithm for the protein [42]. All simulations and analysis were performed using the AMBER and VMD software.

To illustrate the tertiary structure change of the protein, the following order parameters were analysed using the Modules in AMBER 10.0 package: potential energy, root-mean-square deviation (RMSD), radius of gyration (RG) and end-to-end distance. It is to be noted that RMSD is calculated with respect to the lowest energy state as the reference for the folding pathway. Most of the plots were drawn by Xmgr software except the 3D structures of Trp-cage and lysozyme, which were created by VMD software [43].

2.1 Coarse Grain Topology

For CG topology of proteins, each amino acid residue is mapped onto 1–3 CG beads depending on the size of the side chains. The backbone atoms for all 20 amino acids were represented by one bead, placed at the C_α position and modelled as a non-polar interaction site (N0). The details of the mapping of the side chain to interaction centres (beads) are shown in Table 1. The distance between backbone beads was kept at 3.8 Å for all protein residues, while the bond length between the side chain bead and the backbone bead depends on which amino acid they represent. The bond length L_{b-s} was assigned the average distance between the mass centre of the side chain and the mass centre of the backbone evaluated by the real amino acid structure. The initial distance between the backbone beads as well as the distance between the backbone beads and side chain were obtained from all-atom (AA) simulation. It is to be noted that these distances are not constant but vary according to the bonded interactions $V_{bond}(R)$ as given by Equation (3).

Table 1. The bond length L_{b-s} was assigned the average distance between the mass centre of the side chain and the mass centre of the backbone converted by the real amino acid structure.

Type of CG bead	Residue	Bond length L_{b-s} (Å)
Hydrophobic (C)	Ala	2.0
	Ile	2.7
	Leu	3.5
	Val	2.7
	Pro	2.5
Hydrophobic (N_0–N_0)	Phe	3.5–2.0
Hydrophobic (N_0)	Cys	2.7
	Met	3.8
Hydrophilic (N_{da})	Asn	2.8
	Gln	4.0
Hydrophilic (N_{da})	Ser	2.5
	Thr	2.7
Hydrophobic (C–N_{da})	Tyr	3.5–2.7
Hydrophilic (C–N_{da})	His	3.0–2.8
Hydrophobic (N_{da}–C)	Trp	3.1–2.9
Hydrophilic (Q_a)	Asp	3.0
	Glu	4.0
Hydrophilic (C–Q_{da})	Lys	2.0–3.2
	Arg	2.2–3.3

2.2 Force field

The effective interactions between the beads are accounted for by a similar potential equation as the classical MD simulation as:

$$V = V_{bond} + V_{non\text{-}bond}. \quad (1)$$

2.3 Bond interaction

The main bond interaction between the CG sites includes bonded, angle and dihedral interaction.

$$V_{bond} = V_{bonded}(R) + V_{angle}(\theta) + V_{dihedral}. \quad (2)$$

The bonded, angle and dihedral interactions are defined and calculated by the equations given below [34].

The bonded interaction between the connected CG sites is based on a weak harmonic potential.

$$V_{bonded}(R) = \sum_{bondeds} K_{bonded}(R - R_{bonded})^2, \quad (3)$$

where the Lennard-Jones (LJ) interaction is excluded between bonded particles. The force constant of the harmonic bonding potential is $K_{bonded} = 149\,kcal\,mol^{-1}\,nm^{-2}$ for backbone as well as for side chain beads, and the equilibrium distance is $R_{bonded} = 3.8$ Å for backbone. The equilibrium bond lengths for the side chains are given in Table 1.

A weak harmonic potential $V_{angle}(\theta)$ of the cosine type is employed to represent chain stiffness:

$$V_{angle}(\theta) = \sum_{angles} K_{angle}[\theta - \theta_0]^2, \quad (4)$$

where the force constant of $K_{angle} = 12.5\,kJ\,mol^{-1}\,rad^{-2}$ for both backbone as well as for side chains, and the equilibrium bond angle θ_0 has a value of 92° along the main chain, 134° for angles including a side chain bead and 180° for a proline on the centre of the angle.

The dihedral energy is defined for any four successively bonded CG beads.

$$V_{dihedral} = \sum_{dihedrals} \frac{V_d}{2}[1 + \cos(n\phi - \gamma)], \quad (5)$$

where the dihedral force constant $V_d = 2.42\,kJ\,mol^{-1}$ for both backbone as well as for side chains, $n = 1$, phase shift $\gamma = 130°$ for all non-proline backbone dihedrals, and 180° for proline.

2.4 Non-bond interaction

The non-bonded interactions between interaction beads are described by the LJ potential (van der Waals interaction) in addition to the Coulomb potential of the charged beads.

$$V_{\text{non-bond}} = \sum_{i<j}^{\text{beads}} \varepsilon_{ij} \left[\left(\frac{r_{ij}^{*}}{R_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{*}}{R_{ij}} \right)^{6} \right] + \sum_{i,j}^{\text{beads}} \frac{q_i q_j}{4\pi\varepsilon\varepsilon_0 R_{ij}}, \tag{6}$$

where the r_{ij}^{*} represents the effective minimum distance between two CG beads with an assumed value of 2.35 Å; ε represents the relative dielectric constant with a value of 20. The charge of CG beads is $+0.71|e|$ for the side chain bead of lysine and arginine. The charge of CG beads is $-0.71|e|$ for the side chain bead of aspartate and glutamate. The relative dielectric constant value of 20 for charged beads is lower than that of water because of its charge [14,16,17].

The non-bonded interaction between two CG rigid beads was separated into non-polar and electrostatic energies and expressed as sums of pairwise potential energy functions. The CG van der Waals potential was obtained by numerical integration of the non-polar mean potential between two amino acid molecules by using classical all-atom force fields as described below. The classical LJ mean potential energy between any two homologue groups of amino acid atoms were calculated as a function of centre-to-centre distance between beads that are confined in a cubic box of 30 Å length for 50 ns. All possible homologue pairs of amino acids of the 20 amino acids at fully flexible state with vanishing charge (in order to capture the purely non-electrostatic interaction) are considered in which the dihedral and bond interactions are neglected. The mean LJ potential energy is obtained by averaging over all possible configurations of amino acid atoms within a box from MD simulations using AMBER 10 program at 300 K. Although the 20 amino acids in our case were represented by 26 beads (shown in Table 1), some of them are designated as the same type of bead (for example, Cys and Met are designated as Nda type). For these cases, we assigned the parameters as the average of the interactions between homologue pairs of beads belonging to the same type. The total number of the bead type for the amino acid is 10 as shown in Table 2. The parameters for the crossed interactions between two pseudo-atoms of different types were calculated by using the mixing rule $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$ to keep the number of parameters as small as possible.

For the evaluation of the average electrostatic interaction between charged beads, the same procedure as described above was adopted in which the van der Waals, dihedral and bond interactions were neglected. Since the four charged amino acids [aspartate (−), glutamate (−), lysine (+) and arginine (+)] have been classified as the same electrostatic types (see Table 1; aspartate and glutamate are represented by Q_a; lysine and arginine are represented by Q_{da} – the difference is in their ability to form a hydrogen bond), we performed the homologue pairwise calculation for these four amino acids and averaged the parameters. The LJ constant ε_{ij} depends on the type of the possible interaction beads (Table 2), which are given by: $\varepsilon_p = 0.965$ kcal/mol, $\varepsilon_{N0} = 0.793$ kcal/mol, $\varepsilon_{Nd} = 0.889$ kcal/mol, $\varepsilon_{Na} = 0.889$ kcal/mol, $\varepsilon_{Nda} = 0.946$ kcal/mol, $\varepsilon_c = 0.545$ kcal/mol, $\varepsilon_{Q0} = 0.831$ kcal/mol, $\varepsilon_{Qd} = 0.908$ kcal/mol, $\varepsilon_{Qa} = 0.908$ kcal/mol, $\varepsilon_{Qda} = 1.022$ kcal/mol.

2.5 Silica surface

The SiO_2 structure is built up by the program SIO (FORTRAN), in which the initial positions of Si and O atoms in one unit cell are based on the literature values [44,45]. There are two different types of Si atoms and five different types of O atoms, which include 16 Si atoms and 32 O atoms in one unit cell. The unit cell dimensions are $a = 7.1364$ Å, $b = 12.3695$ Å, $c = 7.1742$ Å. The structure of the repeat units in the system can be obtained by the

Table 2. Interaction matrix.

			N					Q			
Group	Subtype	P	0	d	a	da	C	0	d	a	da
P		I	IV	III	III	II	V	I	I	I	I
N	0	IV	III	III	III	III	III	III	III	III	III
	d	III	III	II	II	II	IV	III	III	II	II
	a	III	III	II	II	II	IV	III	II	III	II
	da	II	III	II	II	I	V	III	II	II	I
C		V	III	IV	IV	V	III	V	V	V	V
Q	0	I	III	III	III	III	V	III	III	III	II
	d	I	III	III	II	II	V	III	III	II	I
	a	I	III	II	III	II	V	III	II	III	I
	da	I	III	II	II	I	V	II	I	I	I

Level of interaction I (attractive), II (semi-attractive), III (intermediate), IV (semi-repulsive) or V (repulsive). Four different groups are considered: polar (P), non-polar (N), apolar (C) and charged (Q). Both groups N and Q have four subtypes: 0 for no hydrogen bonding capabilities present, d for groups acting as hydrogen bond donor, a for group acting as hydrogen bond acceptor, and da for groups with both donor and acceptor options.

following steps: for X direction, add a value (7.1367 Å) to every X position of the atoms in the first unit to get the new X positions of the second unit, and add $2a$ value for the 3rd units and so on, until adding $(n-1)a$ value for the nth unit. Thus the total system in X direction is na (7.1364n Å). Y and Z positions do not change for molecules in the same plane. The next layer on the Y and Z direction of the system was built following the same procedure, and addition of molecules in one direction does not affect the other two directions. The structure of the silica for the simulation includes 154 units: every 11 units repeat in X direction (total of 78.5 Å in length), four units repeat in Y direction (total of 49.5 Å in length), and two units repeat in Z direction (total of 14.3 Å in length). The final structure is modified by Leap program to build up the right format for AMBER simulation. The oxygen of silica is treated as an Nd type bead and the silica atom is treated as an N0 type bead, to mimic a hydrophobic surface, and no charge is designated to the surface. The interaction between the silica surface and the beads of protein is same as listed in the Table 2.

The total dimension of the box simulation of lysozyme and silica is 150 Å × 100 Å × 140 Å, including the dimension of the silica as 78.5 Å × 49.5 Å × 14.3 Å. A 15 Å cut-off distance is applied for the non-bond interaction.

3. Results and discussion

3.1 *Comparison of all-atom and CG simulation of Trp-cage in solution*

Trp-cage (TC5b), a small protein, was employed to test the CG algorithm. The initial fully extended structures for Trp-cage (built by all atomic and CG algorithms) are shown in Figure 1(a). It was observed from Figure 1(a) that these two structures have a similar linear structure, although the CG structure lacks the atomic information. Since the initial structure built up by Xleap module may have steric clashes, a short minimisation for the starting structure was executed to remove the bad contacts before starting MD simulation.

Figure 1(b) shows the comparison of the structures of Trp-cage obtained by all-atomic and CG simulation after 900 ns: both structures possess typical features of a stable globular protein. Both structures exhibit a single big turn in residues 11–14 and a C-terminal extension along the helical axis to enwrap the hydrophobic core in the centre.

Changes in potential energy of Trp-cage by AA algorithm and CG algorithm are compared in Figure 2(a) for different temperatures. It is observed that the ΔG (between the initial and lowest energy) by these two methods are comparable, especially at lower temperatures. This is consistent with the comparison of tertiary structures at lowest energy obtained by these two methods as shown in Figure 1(b). The equilibrium RG of Trp-cage calculated by

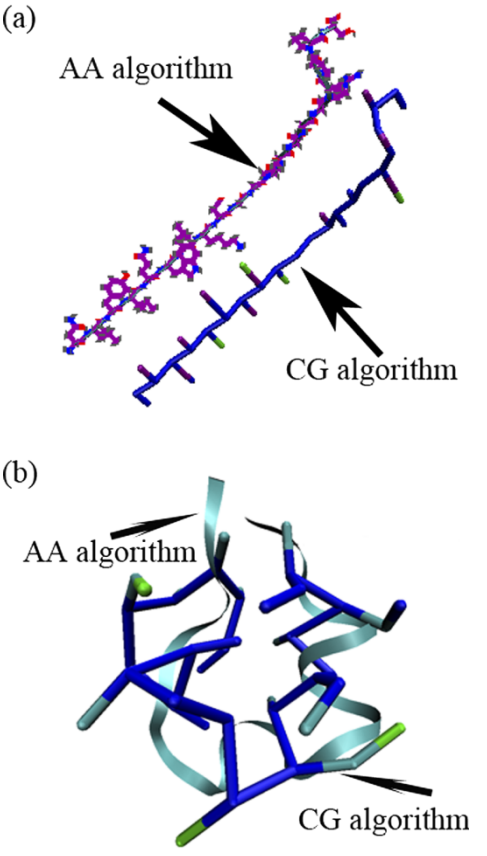

Figure 1. (a) Comparison of the initial fully extended structures of Trp-cage between the all-atom structure and CG structure. (b) Comparison of the lowest energy structures of Trp-cage for all-atom algorithm simulation and CG algorithm simulation after 900 ns. The pictures were built up by VMD software.

AA algorithm and CG algorithm for different temperatures were also comparable (shown in Figure 2(b)), thus indicating that the CG algorithm gives a compact structure for the Trp-cage similar to that obtained by the AA algorithm, especially at low temperatures. The RG obtained by the CG algorithm is less than that obtained by the AA algorithm for all the temperatures, probably due to the smaller space occupied by the side chain bead in the CG algorithm. The time-evolutions of the tertiary structural features, such as RMSD and RG, were also calculated during the simulation at 300 K as shown in Figure 2(c),(d), respectively. Here, the time-evolution of RMSD for an AA algorithm was based on comparison with experimental NMR structure (pdb code: 2vb1) of Trp-cage as a reference structure, while the time-evolution of RMSD for a CG algorithm is based on the comparison of the CG structures during the simulation to the CG structure which was built from the NMR structure. From Figure 2(c), it can be observed that the CG algorithm has a faster rate of convergence of RMSD to its equilibrium value, suggesting a much lower kinetic energy barrier for the CG algorithm during this folding simulation. Although it is found that the

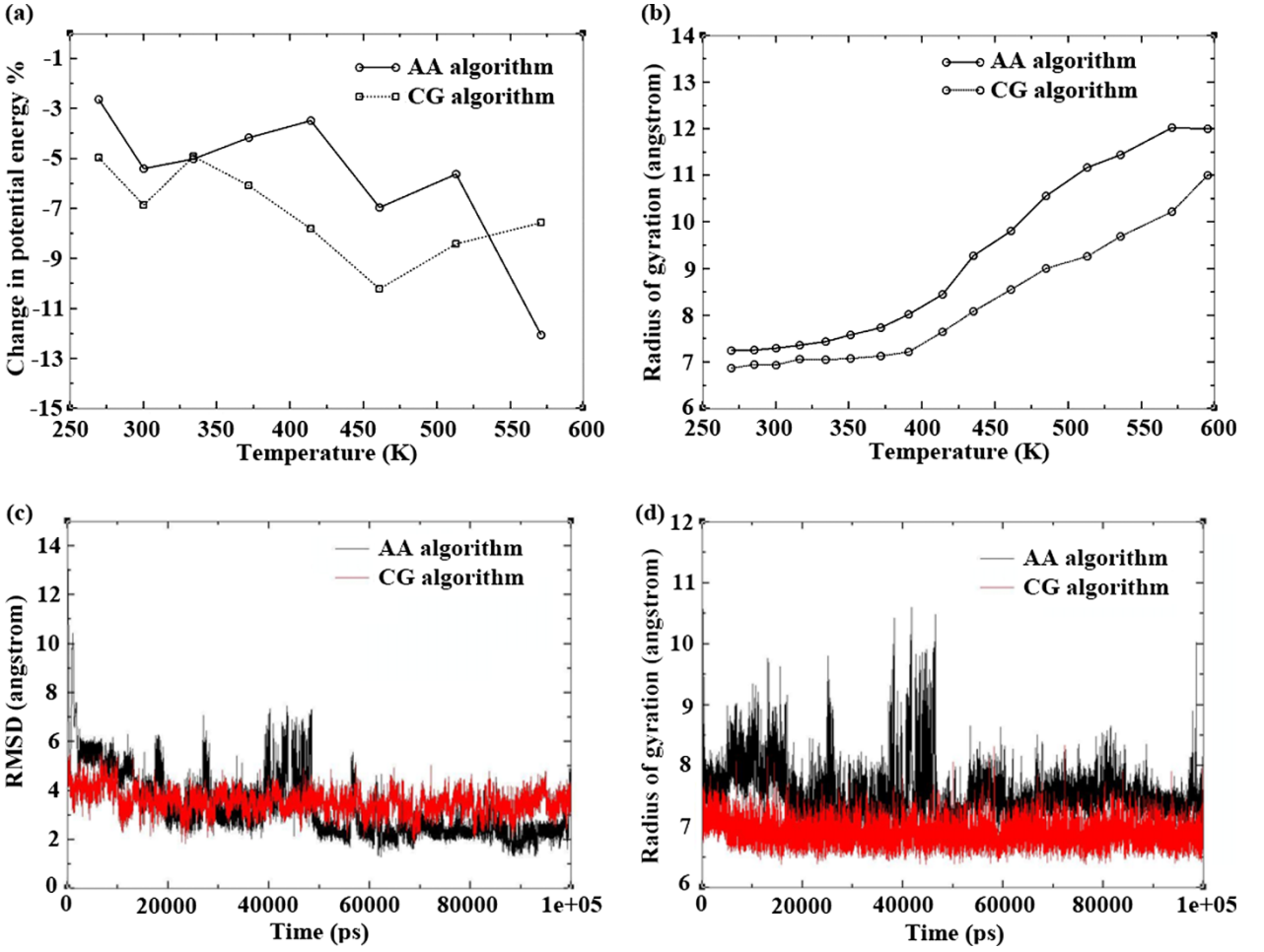

Figure 2. Comparison of CG algorithm and AA algorithm during the simulation. (a) Comparison of the change in potential energy between the lowest energy structures of Trp-cage in solution obtained by all-atom algorithm simulation and CG algorithm simulation for different temperatures. (b) Comparison of the RG obtained by all-atom algorithm simulation and CG algorithm simulation for different temperatures. (c) Comparison of the time-evolution of the RMSD obtained by all-atom algorithm simulation and CG algorithm simulation at 300 K. (d) Comparison of the time-evolution of the RG obtained by all-atom algorithm simulation and CG algorithm simulation at 300 K.

equilibrium value of the RMSD for a CG algorithm is higher than that for an AA algorithm, the values are quite close. The comparison of the time-revolution of RG for these two algorithms as shown in Figure 2(d) exhibited a similar feature as observed in the RMSD plots: a faster convergence rate with a comparable equilibrium value of the RG for a CG algorithm compared to that for an AA algorithm.

3.2 Comparison of all-atom and CG simulation for lysozyme in solution

The initial structures of lysozyme (pdb code: 2vb1) built by the all-atomic and CG algorithms are shown in Figure 3, and both show the native lysozyme structure in the solution with a compact globular structure. It is not surprising that the AA structure is also identical to the CG structure since the position of CG site is based on the AA structure (see Methods section). However, such initial

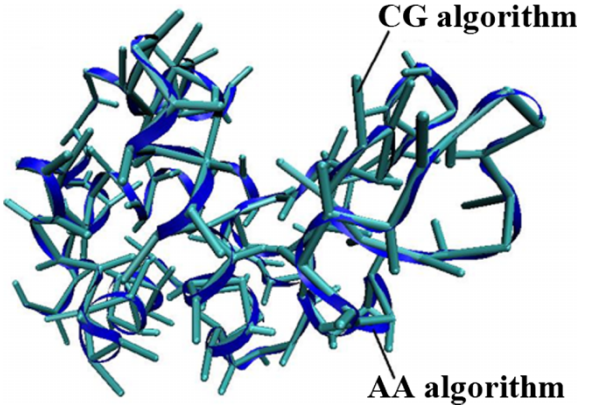

Figure 3. Comparison of the initial fully extended structures of lysozyme between the all-atom structure (based on pdb code: 2vb1) and CG structure.

structures may have some steric hindrance. Therefore, a short time minimisation of the initial structure is necessary before the start of the dynamic simulation. Since the CG method maps several atoms to a single CG bead with a large mass and different interaction properties, it cannot explicitly account for the hydrogen bond and salt bridge (they have been built into the CG force field). Also, the comparison of backbone RMSD between CG and AA structure for folding pathway is not suitable since the backbone elements for CG method (big bead obtained by combining heavy atoms from the AA structure) and AA method (real atoms) are different. Since the CG simulation algorithm represents every side chain by beads without any atomic details, it cannot provide accurate information on the secondary structure (hydrogen bond between the hydrogen and oxygen at a neighbouring peptide, the angle of the rotation of the helix, etc.). However, the tertiary structure obtained by AA or CG algorithm could be compared by comparing the values of the RG or end-to-end distance. The evolution of RMSD with respect to lowest energy conformation will also provide information on the folding pathway. Due to the different force fields employed in the CG and AA algorithms, the absolute potential energies calculated by the two algorithms are quite different, and therefore will not be compared in this paper. However, the percentage of the change of potential energy is a measure of the change in the tertiary structure. As shown in Figure 4, the change in the potential energies of lysozyme due to unfolding in solution as calculated by these two algorithms are quite comparable, especially at low temperatures. The percentage of change in the potential energy due to unfolding is in the range of 0.5–2.0% for both the algorithms. The extent of change in potential energy is found to increase with temperature thus indicating more unfolding at higher temperatures.

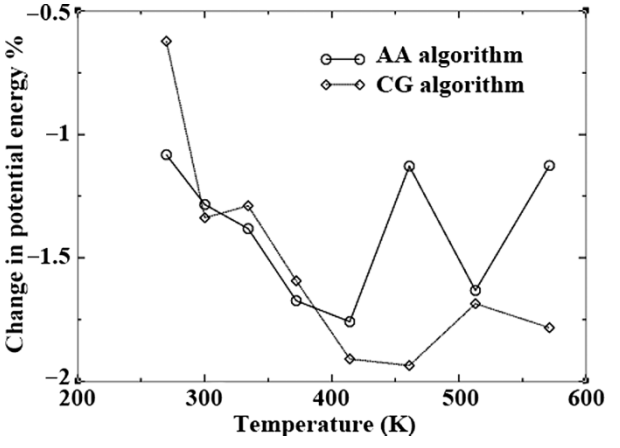

Figure 4. Comparison of the change in potential energy between the lowest energy structures of lysozyme in solution obtained by all-atom algorithm simulation and CG algorithm simulation for different temperatures.

Figure 5(a) compares the calculated values of RG of lysozyme during unfolding for both algorithms. It can be observed from the figure that the RG calculated by both the methods are quite comparable, differing only by 0.5 Å. It was also found that the RG by the AA simulation algorithm gives a larger value than that by CG simulation algorithm at all the temperatures. This is possibly due to mapping of the atoms in side chains into one or two beads in the CG algorithm thereby reducing the size of the structure. The calculated values of end-to-end distances at 900 ns for both algorithms are shown in Figure 5(b). The results indicate that the proposed CG algorithm predicts the trend of the variation of tertiary conformation of lysozyme with temperature even though the predicted end-to-end distance by CG is found to be lower than the values predicted by AA simulation by 15–20% except at 270 K. It is to be noted that the calculated values of end-to-end distance by AA algorithm increase faster with temperature compared to the calculated values by CG algorithm.

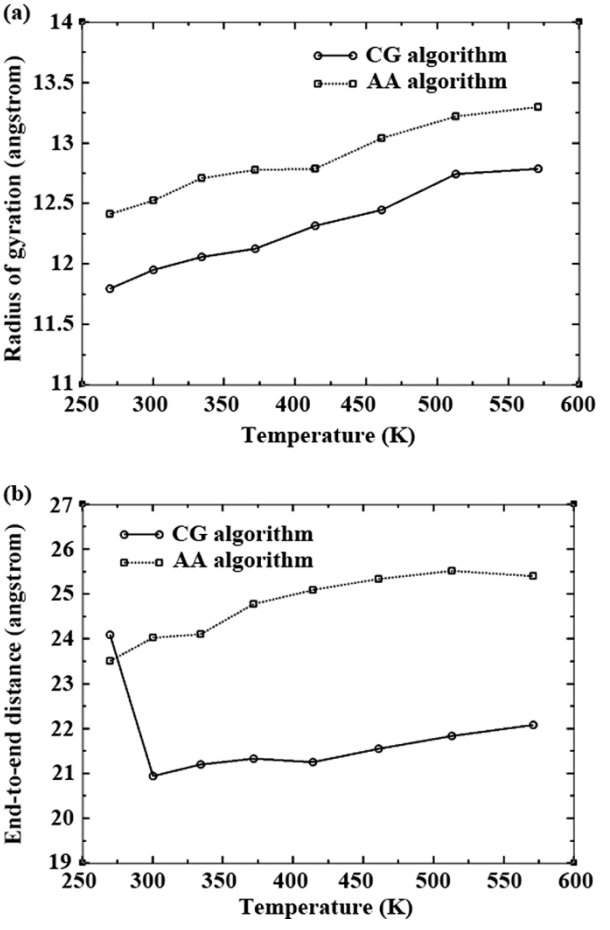

Figure 5. Comparison of the parameters for tertiary structure of lysozyme at the lowest energy state in solution between all-atom algorithm and CG algorithm at different temperatures: (a) RG (Å). (b) End-to-end distance (Å).

RMSD with respect to the lowest energy structure by simulation (reference structure) is calculated as a function of time during the evolution of the folding pathway for lysozyme obtained by CG and AA algorithms at different temperatures. In both cases, the RMSD values increased with temperature as a result of more unfolding at higher temperatures. With respect to lowest energy structure, the RMSD fluctuation during the simulation is found to be around 2 Å at the lowest temperature (278 and 300 K), thereby indicating that little unfolding occurred at these low temperatures (see Figure 6(a)). It is interesting to note that the RMSD values for the AA structure is more sensitive to the temperature with higher values compared to those obtained for CG structure especially at higher temperatures (Figure 6(b)). This is believed to be due to a larger number of interaction centres in AA simulation compared with CG. Direct comparison of RMSD values for the two algorithms could not be made because of the difference in the total number of interaction centres resulting from simplification of the structure in CG. Figure 6(c),(d) give the evolution of RMSD with respect to the intial structure for CG and AA algorithms, respectively. Unlike the earlier cases, RMSD value increases with time. The effect of temperature on RMSD follows the same trend as above.

3.3 Comparison of behaviour of lysozyme on silica surface with solution using CG simulation

Figure 7(a) shows the comparison of the calculated values of RG corresponding to the lowest energy structure for lysozyme in solution and on silica surface using CG simulation algorithm at different temperatures. The RG values for lysozyme on silica surface are found to be around 1 Å larger than those in solution, thus indicating that the tertiary structure of lysozyme on the surface is more unfolded as a result of adsorption.

Figure 7(b),(c) shows the comparison of the calculated values of end-to-end distance corresponding to the lowest energy structure and its SD for lysozyme in solution and on silica surface using CG simulation algorithm at different temperatures, respectively. It was observed that the end-to-end distance of lysoyzme structure on silica surface has a higher value compared to that in solution, thus indicating a less compact structure on the surface. This is consistent with our previous finding for RG. A higher deviation of end-to-end distance is found for lysozyme on silica surface, which suggests that the structure of lysozyme on the surface may be more unstable compared to that in solution. Figure 8 shows the evolution of RMSD for lysozyme on silica surface with respect to the

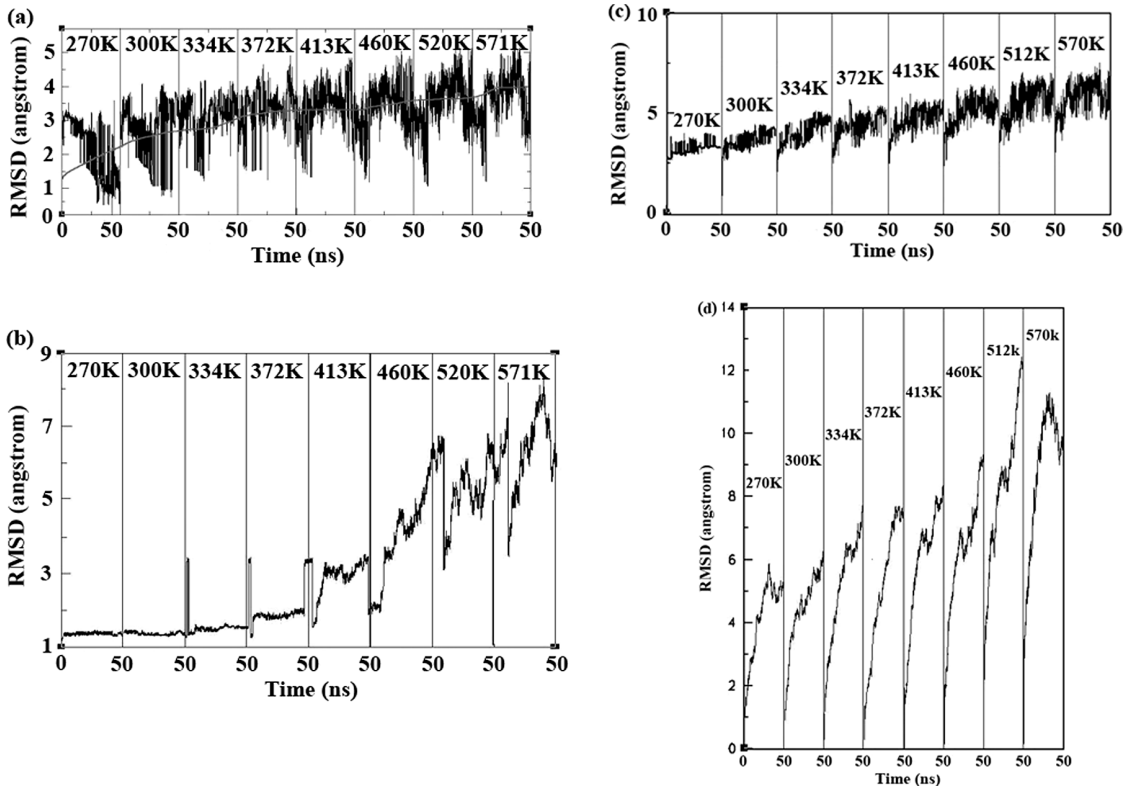

Figure 6. The evolution of RMSD during the simulation at different temperatures: (a) CG algorithm. (b) All-atom algorithm. The reference structures of (a) and (b) for RMSD calculation is the lowest energy structure during the simulation among all the temperatures. (c) CG algorithm. (d) All-atom algorithm. The reference structures of (c) and (d) for RMSD calculation are the initial structure during the simulation for all the temperatures.

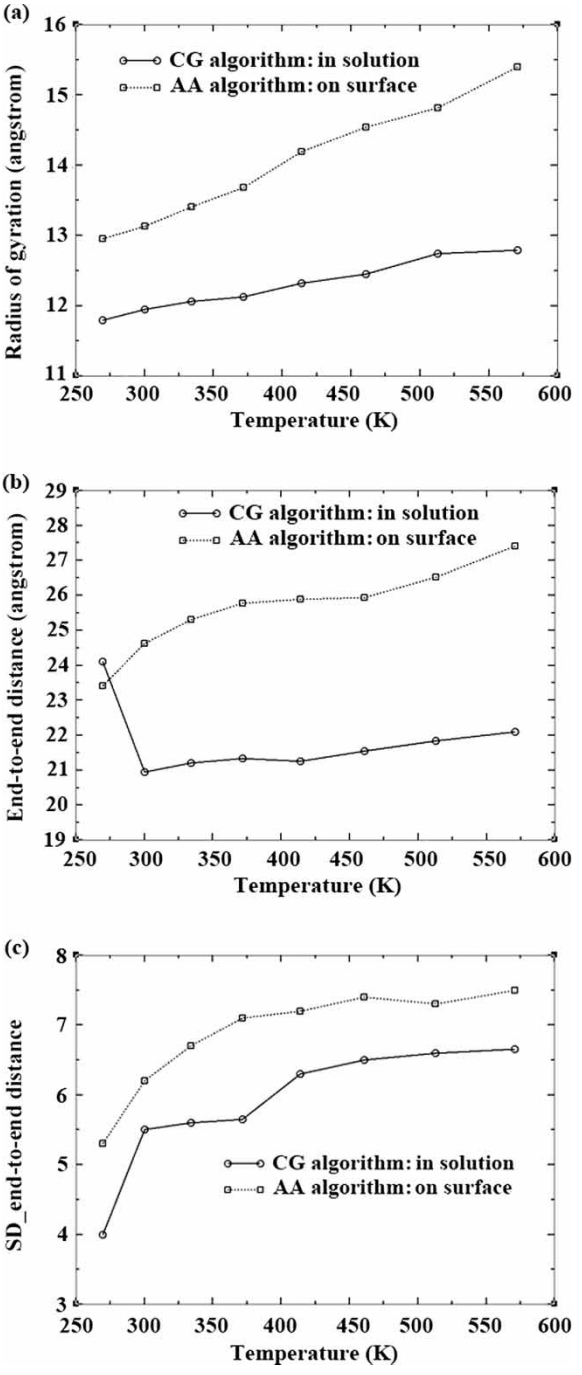

Figure 7. Comparison of the parameters for tertiary structure of lysozyme between the lowest energy state in solution and on silica surface at different temperatures: (a) RG (Å). (b) End-to-end distance (Å). (c) SD of end-to-end distance.

intial structure. It is to be noted that the RMSD values for lysozyme adsorbed on silica surface are higher (4–8 Å) compared to the values for lysozyme in solution (2.5–6 Å, see Figure 6(c)) as a result of more unfolding on the surface.

It is of interest to investigate the orientation of a protein molecule adsorbed on the surface to better

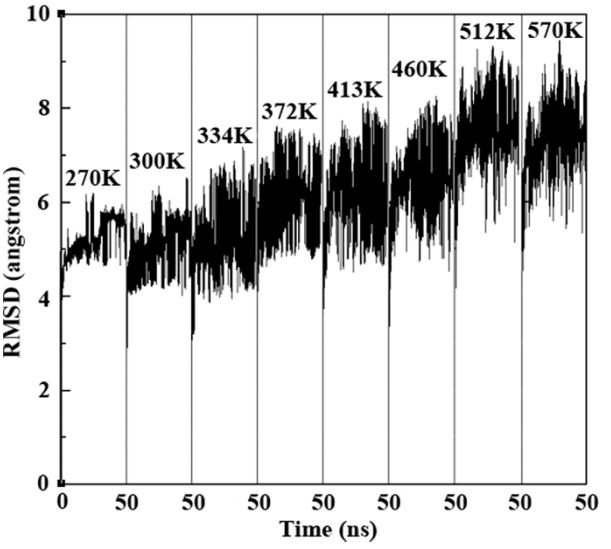

Figure 8. The evolution of RMSD at different temperatures for CG algorithm of lysozyme adsorbed on the silica surface. The reference structures for RMSD calculation are the initial structure for all the temperatures.

elucidate protein surface interaction. Figure 9(a) shows the change of the height of a protein molecule (vertical to the surface) during its adsorption to the silica surface at 300 K. The vertical height from the surface was found to decrease rapidly during the initial time up to 0.5 ns from 3.1 to 1.9 nm (around 1/3 decrease in height) followed by a small fluctuation around the equilibrium value of around 2 nm. Figure 9(b) shows the effect of temperature on the height of the protein molecule after the molecule reaches its equilibrium state. The vertical height from the silica surface was found to be lower at higher temperature thus indicating more unfolding.

Figure 9(c) shows the distances between each residue of the lysozyme molecule and the silica surface at several different times, which provides more detailed information about the interaction of the residues with the silica surface. The distance of residues from the surface decreases rapidly from 0.01 to 0.1 ns consistent with the results presented in Figure 9(a). Even though the vertical distance of lysozyme from the silica surface does not change significantly at longer times (0.1–1 ns), the distances of specific residues from the surface change thus indicating conformational change. Although it was found that the overall distances between the residues and the surface may be quite similar at 0.1 and 1 ns, such a distance is quite different for a specific residue during the simulation, thus indicating a conformational change. Figure 9(d) shows the projection of a lysozyme molecule on the surface at different times, which reveals the conformational change of lysozyme on the surface. Besides the height, the projected area is another indicator of the orientation of the protein molecule on the surface. At an initial time

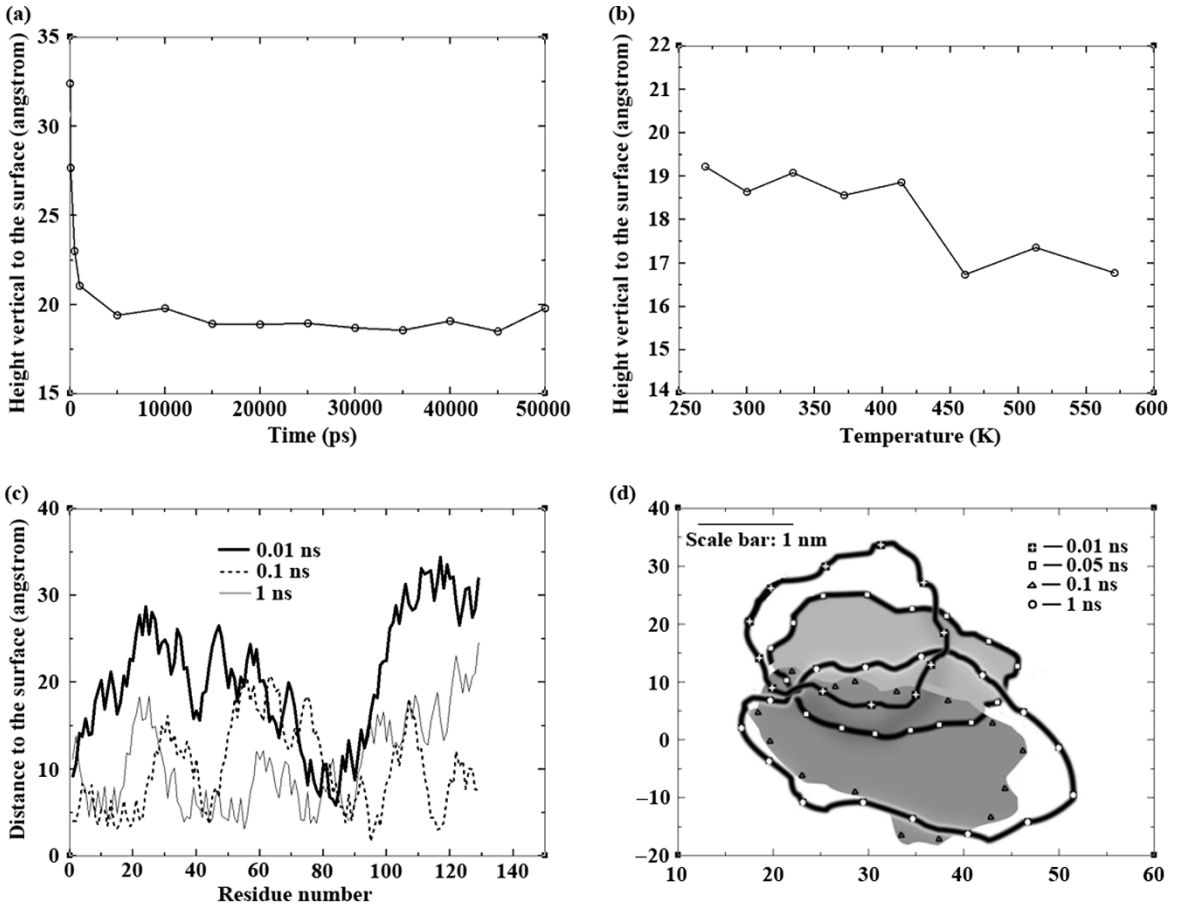

Figure 9. Orientation of lysozyme molecule on the silica surface. (a) Kinetics of the change of the height of lysozyme molecule at 300 K. (b) The height of lysozyme molecule at the equilibrium state for different temperatures. (c) Comparison of the distance of each residue of lysozyme to the silica surface for different simulation time at 300 K. (d) Comparison of the projection shape of the lysozyme molecule on the silica surface for different simulation time at 300 K. The direction of the height is vertical to the silica surface.

of 0.01 ns, the projected area is small as a result of the compact original structure. The projected area increases with time with a corresponding decrease in the vertical height from the surface which suggests unfolding. The RG does not fully describe the unfolding of the protein molecule because of the opposing effects of the increased projected area and decreased vertical height from the surface.

Figure 10(a),(b) shows the 3D structures of lysozyme in solution and on silica surface, respectively, at the lowest energy state at 300 K. Direct visual comparison of these conformations of lysozyme reveals that lysozyme on the surface has a flatter structure compared to that in solution. It was also found that the RMSD between the structure of lysozyme on the surface and that in solution is around 2.7 Å, which indicated a small unfolding of lysozyme on the surface. The end-to-end distance values as shown in the figures for these two structures are 24.48 Å (on surface) and 21.10 Å (in solution), respectively.

Figure 11(a) shows the comparison of calculated values of RG of lysozyme on silica surface using CG algorithm for two different ionic strengths. A higher extent of unfolding was observed for higher ionic strength with a higher value of RG. Higher ionic strength compresses the electrical double layer in the vicinity of the silica surface as well as the adsorbed protein molecule thereby reducing the electrostatic interaction between them. In addition, compression of the electrical double layer will also result in suppression of intra molecular electrostatic interactions for lysozyme. Both these effects will result in more unfolding. As shown in the Figure 11(b), the calculated values of end-to-end distance for lysozyme adsorbed on silica surface using CG algorithm is much higher for higher ionic strengths. This effect of ionic strength was found to be less pronounced at higher temperature. Such a behaviour is due to larger thermal energy of molecules at higher temperature, which reduces the effect of electrostatic interactions.

4. Conclusions

A CG algorithm was developed for MD simulation of proteins in which the polypeptide backbone as well as side

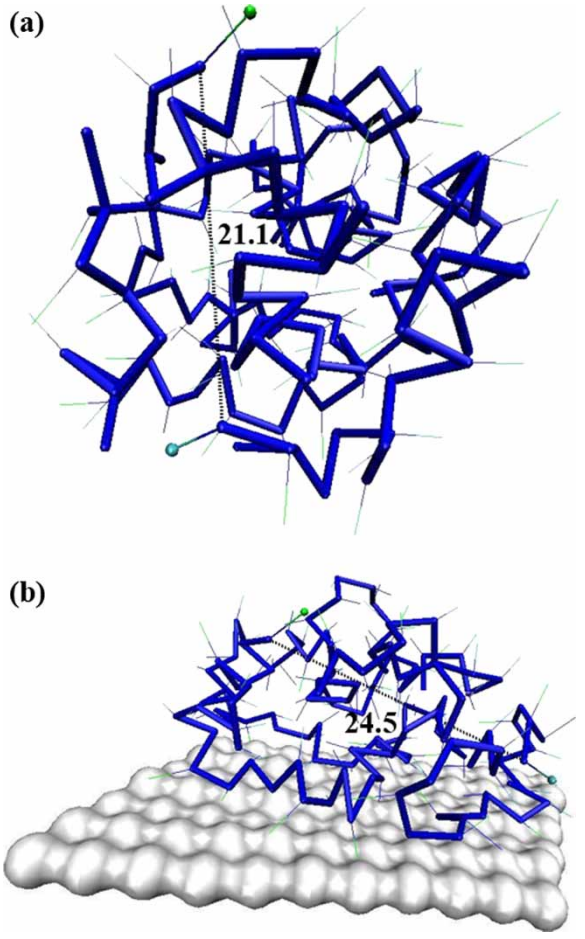

Figure 10. The 3D structure of lysozyme in solution and on the silica surface at the lowest energy state at 300 K. (a) In solution, (b) on silica surface.

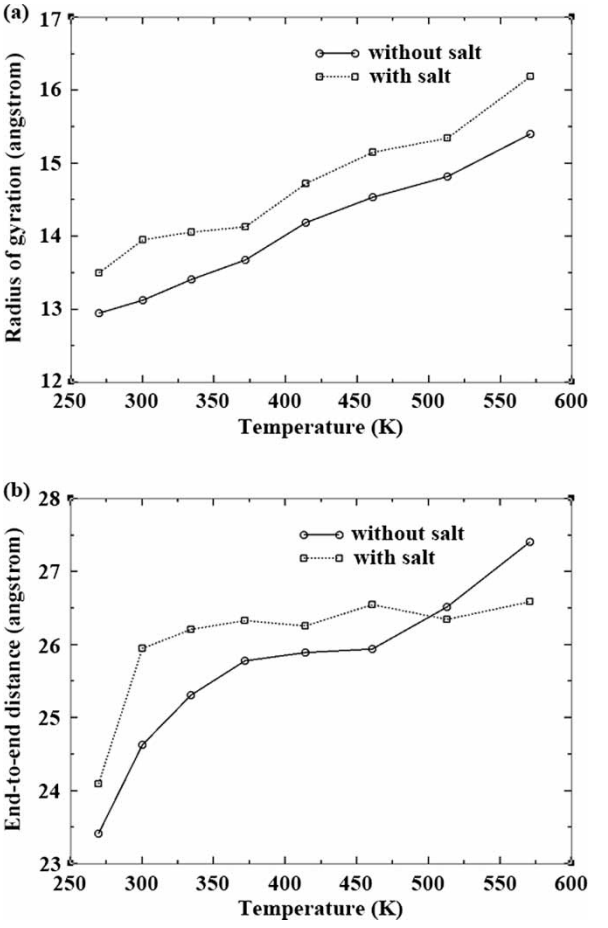

Figure 11. Comparison of the parameters for tertiary structure of lysozyme between the lowest energy states on silica surface for two different ionic strengths at different temperatures: (a) RG (Å). (b) End-to-end distance (Å). The higher ionic strength is 0.12.

chains were mapped into spherical interaction centres. The force field parameters for non-bonded interactions between beads were evaluated from the evaluation of interaction of all atoms corresponding to these beads using pairwise additivity by employing AMBER simulation software. The validity of CG algorithm was demonstrated by comparing the simulation result with that obtained by AA method for a small protein molecule Trp-cage and lysozyme in solution. The ΔG, RG and end-to-end distance were found to compare favourably for these two methods. Therefore, employing the CG molecular simulation method to replace the conventional all-atomic method is an efficient tool to speed up the simulation rate, especially for a large system. Our simulation results show that using the CG method for the prediction of tertiary conformation of lysozyme in solution costs less than 1/50 CPU-time compared to that by the AA method. Also the CG method converges to equilibrium potential energy much faster than AA simulation.

It was shown that a lysozyme molecule adsorbed on silica surface is more unfolded with a less compact tertiary structure compared to that in the solution. The projected area of the lysozyme on silica surface was found to increase with time during adsorption whereas the vertical height of the molecule from the surface decreased, thus indicating unfolding of the molecule. This unfolding was rapid during the initial time interval of 0.5 ns, which was followed by a small fluctuation around the equilibrium state. At longer times, even though the RG of the lysozyme does not change significantly, the distance of some specific residues from the silica surface does change, which suggests reorientation. Higher ionic strength was shown to result in a more extended structure for lysozyme on silica surface. These results will be helpful in elucidating protein unfolding pathway on the surface.

Acknowledgements

We would like to acknowledge the Purdue Research Foundation for providing financial support to X.W. We also would like to acknowledge Dr Janaswamy and Dr R. Chandrasekaran for their help in the construction of silica surface.

References

[1] X. Wu and G. Narsimhan, *Characterization of secondary and tertiary conformational changes of β lactoglobulin on silica nanoparticle surfaces*, Langmuir 24 (2008), pp. 4989–4998.

[2] C.B. Anfinsen, *Principles that govern the folding of protein chains*, Science 181 (1973), pp. 223–230.

[3] E. Alm and D. Baker, *Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures*, Proc. Natl Acad. Sci. USA 96 (1999), pp. 11305–11310.

[4] C.M. Dobson and M. Karplus, *The fundamentals of protein folding: bringing together theory and experiment*, Curr. Opin. Struct. Biol. 9 (1999), pp. 92–101.

[5] H. Lei and Y. Duan, *Improved sampling methods for molecular simulation*, Curr. Opin. Struct. Biol. 17 (2007), pp. 187–191.

[6] D.O. Alonso and V. Daggett, *Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains*, Proc. Natl Acad. Sci. USA 97 (2000), pp. 133–138.

[7] W. Wang, O. Donnii, C.M. Reyes, and P.A. Kollman, *Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions*, Annu. Rev. Biophys. Biomol. Struct. 30 (2001), pp. 211–243.

[8] T. Hansson, C. Oostenbrink, and W.F. van Gunsteren, *Molecular dynamics simulations*, Curr. Opin. Struct. Biol. 12 (2002), pp. 190–196.

[9] M. Karplus and J.A. McCammon, *Molecular dynamics simulations of biomolecules*, Nat. Struct. Biol. 9 (2002), pp. 646–652.

[10] H.X. Lei and Y. Duan, *Two-stage folding of HP-35 from* ab initio *simulations*, J. Mol. Biol. 370 (2007), pp. 196–206.

[11] K. Leonhard, J.M. Prausnitz, and C.J. Radke, *3D-Lattice Monte Carlo simulations of model proteins. Size effects on folding thermodynamics and kinetics*, Biophys. Chem. 106 (2003), pp. 81–89.

[12] K. Leonhard, J.M. Prausnitz, and C.J. Radke, *Solvent-amino acid interaction energies in three-dimensional-lattice Monte Carlo simulations of a model 27-mer protein: folding thermodynamics and kinetics*, Protein Sci. 13 (2004), pp. 358–369.

[13] K. Leonhard, J.M. Prausnitz, and C.J. Radke, *Three-dimensional lattice Monte Carlo simulations of model proteins. IV. Proteins at an oil-water interface*, Langmuir 22 (2006), pp. 3265–3272.

[14] S.J. Marrink, A.H. de Vrij, and A.E. Mark, *Coarse grain model for semiquantitative lipid simulations*, J. Phys. Chem. B 108 (2004), pp. 750–760.

[15] N. Basdevant, D. Borgis, and T. Ha-Duong, *A coarse-grained protein–protein potential derived from an all-atom force field*, J. Phys. Chem. B 111 (2007), pp. 9390–9399.

[16] A.Y. Shih, A. Arkhipov, P.L. Freddolino, and K. Schulten, *Coarse grained protein-lipid model with application to lipoprotein particles*, J. Phys. Chem. B 110 (2006), pp. 3674–3684.

[17] P.J. Bond and M.S. Sansom, *Insertion and assembly of membrane proteins via simulation*, J. Am. Chem. Soc. 128 (2006), pp. 2697–2704.

[18] Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chem. Phys. Lett. 314 (1999), pp. 141–151.

[19] A. Fleming, *On a remarkable bacteriolytic element found in tissues and secretions*, Proc. R. Soc. Ser. B 93 (1922), pp. 306–317.

[20] C.C. Blake, D.F. Koenig, G.A. Mair, A.C.T. North, D.C. Phillips, and V.R. Sarma, *Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Ångstrom resolution*, Nature 206 (1965), pp. 757–761.

[21] J.C. Cheetham, P.J. Artymiuk, and D.C. Phillips, *Refinement of an enzyme complex with inhibitor bound at partial occupancy. Hen egg-white lysozyme and tri-N-Acetylchitotriose at 1.75 Å resolution*, J. Mol. Biol. 224 (1992), p. 613.

[22] L.N. Johnson and D.C. Phillips, *Structure of some crystalline lysozyme-inhibitor complexes determined by X-ray analysis at 6 Ångstrom resolution*, Nature 206 (1965), pp. 761–763.

[23] A.G. Murzin, S.E. Brenner, T. Hubbard, and C. Chothia, *SCOP: a structural classification of proteins database for the investigation of sequences and structures*, J. Mol. Biol. 247 (1995), pp. 536–540.

[24] P. Bai and Z. Peng, *Cooperative folding of the isolated alpha-helical domain of hen egg-white lysozyme*, J. Mol. Biol. 314 (2001), pp. 321–329.

[25] B. Fischer, *Folding of lysozyme*, EXS 75 (1996), pp. 143–161.

[26] H. Tachibana, O. Akasaka and K. Takashi, *Native-like tertiary structure formation in the alpha-domain of a hen lysozyme two-disulfide variant*, J. Mol. Biol. 314 (2001), pp. 311–320.

[27] V.K. Sharma and D.S. Kalonia, *Steady state tryptophan fluorescence spectroscopy study to probe tertiary structure of proteins in solid powders*, J. Pharm. Sci. 92 (2003), pp. 890–899.

[28] M.H. Baron, M. Revault, S. Servagent-Noinville, J. Abadie and H. Quiquampoix, *Chymotrypsin adsorption on montmorillonite: enzymatic activity and kinetic FTIR structural analysis*, J. Colloid Interf. Sci. 214 (1999), pp. 319–332.

[29] S. Noinville, M. Revault, M. Baron, A. Tiss, S. Yapoudjian, M. Ivanova, and R. Verger, *Conformational changes of enzymes adsorbed at liquid-solid interface: relevance to enzymatic activity*, Biopolymers 67 (2002), pp. 323–326.

[30] A. Sethuraman, G. Vedantham, T. Imoto, T. Przybycien, and G. Belfort, *Protein unfolding at interfaces: slow dynamics of alpha helix to beta sheet transition*, Proteins Struct. Funct. Biointerf. 56 (2004), pp. 669–678.

[31] P. Billsten, P.O. Freskgard, U. Carlsson, B.H. Jonsson, and H. Elwing, *Adsorption to silica nanoparticles of human carbonic anhydrase II and truncated forms induce a molten-globule-like structure*, FEBS Lett. 402 (1997), pp. 67–72.

[32] A.A. Vertegel, et al., *Silica nanoparticle size influences the structure and enzymatic activity of adsorbed lysozyme*, Langmuir 20 (2004), pp. 6800–6807.

[33] X. Wu and G. Narsimhan, *Effect of surface concentration on secondary and tertiary conformational changes of lysozyme adsorbed on silica nanoparticles*, Biochim. Biophys. Acta (BBA) Proteins Proteomics 1784 (2008), pp. 1694–1701.

[34] D.A. Case, T.A. Darden, T.E. Cheatham III, C. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, and P.A. Kollman, http://www.ambermd.org/doc8/amber8.pdf,AMBER 8, Univesity of California, San Francisco, 2004.

[35] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *Development and testing of a general amber force field*, J. Comput. Chem. 25 (2004), pp. 1157–1174.

[36] C. Simmerling, W.C. Still, A. Tempczyk, R.C. Hawley, and T. Hendrickson, *All-atom structure prediction and folding simulations of a stable protein*, J. Am. Chem. Soc. 124 (2002), pp. 11258–11259.

[37] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola and J.R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. 81 (1984), pp. 3684–3690.

[38] R. Zhou, *Trp-cage: folding free energy landscape in explicit water*, Proc. Natl Acad. Sci. USA 100 (2003), pp. 13280–13285.

[39] J.W. Pitera and W. Swope, *Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins*, Proc. Natl Acad. Sci. USA 100 (2003), pp. 7587–7592.

[40] W.C. Still, A. Tempczyk, R.C. Hawley, and T. Hendrickson, *Semianalytical treatment of solvation for molecular mechanics and dynamics*, J. Am. Chem. Soc. 112 (1990), pp. 6127–6129.

[41] D. Qiu, P.S. Shenkin, F.P. Hollinger, and W.C. Still, *The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate born radii*, J. Phys. Chem. A 101 (1997), pp. 3005–3014.

[42] B. Hess, H. Bekker, H.J.C. Berendsen, and J. Fraaije, *LINCS: a linear constraint solver for molecular simulations*, J. Comput. Chem. 18 (1997), pp. 1463–1472.

[43] W. Humphrey, A. Dalke, and K. W.C. Still, *VMD: visual molecular dynamics*, J. Mol. Graph. 14 (1996), pp. 33–38.

[44] Monoclinic, *C12/c1 Chemical handbook*, p. 101.

[45] K.L. Geisinger, et al., *Exploration of structure, electron density distribution, and bonding in coesite with Fourier and pseudoatom refinement methods using single-crystal X-ray diffraction data*, J. Phys. Chem. 91 (1987), p. 3237.