Application of high performance computing in calculating second virial coefficient and predicting protein aggregation with different coarse-graining representations

Pin-Kuang Lai

Department of Chemical Engineering, University of Delaware

Introduction

Protein is of primary importance in biology, pharmaceutical industry. However, many behaviors of proteins are not fully understood. Atomistic modeling of proteins like many other soft matter systems is constrained from the computational time.

Many coarse-graining (CG) representations of proteins have been proposed to simplify the system. One bead per amino acid (1bAA) and four beads per amino acid (4bAA) representations are used in this work.

Second virial coefficient (B_{22}) represents the interaction between two particles. B_{22}/B_{22}^S is an indicator of protein aggregation. B_{22}^S means steric second virial coefficient. However, B_{22}^S cannot be obtained from experiment. Kofke et al. [1] proposed an algorithm named Mayer-Sampling to solve virial coefficient. This method has been used to calculate for Lennard-Jones fluid and water and the results agree with theoretical values.

The purpose of this study is to analyze the validity of using Mayer-Sampling to calculate for protein systems. Moreover, attractive interactions are applied which hasn't been studied before. In addition, the relationship between 1bAA and 4bAA representations is also of interest.

Theory

From statistical thermodynamics, the second osmotic virial coefficient can be expressed as:

$$B_{22} = -\frac{1}{2}\int f_{12} r_{12} \qquad (1)$$

In the integral, $f_{12} = [\exp(-\beta W_{12}) - 1]$ is the Mayer-function,
Where $\beta = (k_B T)^{-1}$ is the reciprocal of temperature in energy units and W_{12} is the orientationally-averaged potential of mean force.

(A) Mayer-Sampling

Mayer-Sampling can solve integral of any order and can be applied to potential of any complexity. The approach is based on umbrella sampling and the general form is

$$\Gamma = \Gamma_0 \frac{\Gamma}{\Gamma_0} = \Gamma_0 \frac{\langle \gamma / \pi \rangle_\pi}{\langle \gamma_0 / \pi \rangle_\pi} \qquad (2)$$

Γ is the general cluster integral. The subscript 0 denotes reference system. is the integrand. The angle brackets with subscript π indicates ensemble average governed by the probability distribution $\pi := A >_\pi = \int dr^n A\pi / \int dr^n \pi$. In order to get second osmotic second virial coefficient, γ is set as f_{12}, and π is defined as the absolute value of $\gamma = |f_{12}|$.

In this work, gamma d-crystalline is used and the structure coordinates are taken from Protein data bank (PDB). This protein has 173 amino acids. In addition, the center of protein is set at the centroid and each amino acid is placed relative to the center.

(B) Coarse-Graining Structure

Modeling protein structure atomically is computationally expensive. Coarse graining (CG) representation provides a more tractable approach. In this work, two CG approaches are applied.

In one bead per amino acid (1bAA) representation, a single bead is used to substitute for one residual. Each bead is assigned with a characteristic diameter based on the residual. The center of bead is set at the centroid of central carbon C_α.

Another CG approach is four beads per amino acid (4bAA) representation. In this method, amino group N, central carbon C_α, carbonyl group C', and (for non-glycine residual) side chain are replaced with one bead , respectively.

Reference

[1]. J.K. Singh, D.A. Kofke, Phys. Rev. Lett., *92*, 220601-1, 2004.

(c) Interactions

This work is aimed to study the attractive interaction between proteins, and its effect to under different CG representations. The potential of mean force of CG representation is shown below:

$$W_{ij} = 4\varepsilon^* \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{128} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \qquad (3)$$

where σ_{ij} is the arithmetic average diameters of beads *i* and *j*. r_{ij} is the center to center distance between beads *i* and *j*. $\varepsilon^* (= \varepsilon / kT)$ is the reduced energy parameter.

The hard sphere system is usually chosen as the reference system due to its simplicity. A sphere is used to represent a protein. Moreover, the center of the sphere is placed at the centroid of the protein. The potential of mean force for the reference system is given below.

$$W^{ref} = \left(\frac{\sigma^{ref}}{r^{ref}} \right)^{128} \qquad (4)$$

where σ_{ij} is the diameter of the reference sphere. r_{ij} is the center to center distance of the reference spheres.

Results and discussion

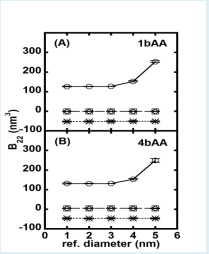

Figure 1. The relationship between B_{22} and the diameter of reference sphere with different reference potential forms. In subfigure (A) ○ denotes to steric system with HS ref. □ denotes to LJ system (0.058)with HS ref. ◇ denotes to LJ system (0.058) with LJ ref. potential. × denotes to LJ system (0.063) with HS ref. + denotes to LJ system (0.063) with LJ ref. In (B) ○ denotes to steric system with hard sphere (HS) ref. □ denotes to LJ system (0.045) with HS ref. ◇ denotes to LJ system (0.045) with LJ ref. × denotes to LJ system (0.047) with HS ref. + denotes to LJ system (0.047) with LJ ref. The value in parenthesis denotes reduced energy parameter.

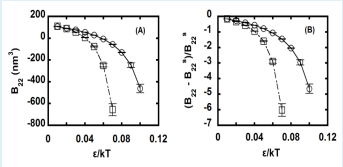

Figure 2. The relationship between B_{22} or reduced B_{22} and . (○ denotes to 1bAA representation, □ denotes to 4bAA representation.)

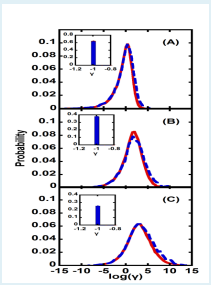

Figure 3. The distribution of gamma for 1bAA and 4bAA with equivalent B22. The main plot is the distribution of attractive interaction (solid line represents 1bAA and dash line represents 4bAA), and the inset plot is the distribution of repulsive interaction (red is 1bAA and blue is 4bAA). ((A): 0.04 for 1bAA and 0.03 for 4bAA; (B): 0.073 for 1bAA and 0.05 for 4bAA; (C): 0.09 for 1bAA and 0.06 for 4bAA.) The numerical value represents reduced energy parameter.

Fig1. shows Mayer-Sampling can be applied to attractive system for both 1bAA and 4bAA representations regardless of the reference diameter and potential used. Fig2. shows that the B_{22} and reduced B_{22} value of 4bAA model drops more rapidly than that of 1bAA model. It is because of the effective area is larger for 4bAA. Fig 3. shows that if we can find a equivalent energy parameter for 1bAA model so that the B22 values for both models are the same. The integrated probability distribution of gamma value is similar for both 1bAA and 4bAA models. It argues that if the energy parameter is shifted, one can use 1bAA model to obtain the result from 4bAA model, which is computationally more effective.