Chemistry Publications

Chemistry

1-2011

Calculations of the Second Virial Coefficients of Protein Solutions with an Extended Fast Multipole Method

Bongkeun Kim
Iowa State University

Xueyu Song
Iowa State University, xsong@iastate.edu

Calculations of the Second Virial Coefficients of Protein Solutions with an Extended Fast Multipole Method

Abstract

The osmotic second virial coefficients B2 are directly related to the solubility of protein molecules in electrolyte solutions and can be useful to narrow down the search parameter space of protein crystallization conditions. Using a residue level model of protein-protein interaction in electrolyte solutions B2 of bovine pancreatic trypsin inhibitor and lysozyme in various solution conditions such as salt concentration, pH and temperature are calculated using an extended fast multipole method in combination with the boundary element formulation. Overall, the calculated B2 are well correlated with the experimental observations for various solution conditions. In combination with our previous work on the binding affinity calculations it is reasonable to expect that our residue level model can be used as a reliable model to describe protein-protein interaction in solutions.

Disciplines
Chemistry

Calculations of the second virial coefficients of protein solutions with an extended fast multipole method

Bongkeun Kim and Xueyu Song

Department of Chemistry, Iowa State University, Ames, Iowa 50011, USA

The osmotic second virial coefficients B_2 are directly related to the solubility of protein molecules in electrolyte solutions and can be useful to narrow down the search parameter space of protein crystallization conditions. Using a residue level model of protein-protein interaction in electrolyte solutions B_2 of bovine pancreatic trypsin inhibitor and lysozyme in various solution conditions such as salt concentration, pH and temperature are calculated using an extended fast multipole method in combination with the boundary element formulation. Overall, the calculated B_2 are well correlated with the experimental observations for various solution conditions. In combination with our previous work on the binding affinity calculations it is reasonable to expect that our residue level model can be used as a reliable model to describe protein-protein interaction in solutions.

I. INTRODUCTION

In a remarkable observation, George and Wilson found that there is a correlation between slightly negative second virial coefficient of a protein solution and its successful crystallization condition [1]. There is also a correlation between the solubility of a protein in an electrolyte solution and the osmotic second virial coefficient B_2 of the solution (Veesler *et al.* [2], Boistelle *et al.* [3]). These observations have led to numerous studies on the second virial coefficients of protein solutions with the hope to use this property to narrow down the parameter space of protein solutions for the search of optimal crystallization conditions. For example, even for membrane proteins, rapid screening of small molecules and detergents as crystallization additives are achieved to improve the crystallization conditions of light harvesting protein complexes [4].

Experimentally, the osmotic second virial coefficients B_2 can be measured by using Static Light Scattering (SLC) [1,5,6], Small Angle X-ray Scattering (SAXS) [7], Small Angle Neutron Scattering (SANS) [8], or Self-Interaction Chromatography (SIC) [9]. All of these methods, however, are quite demanding due to large amounts of proteins used in the measurements. So far, using the B_2 of protein solutions as a tool to screen the solution conditions is not a routine practice yet in most crystallographers' labs.

To overcome the protein consumption problem in B_2 measurements, one possible alternative is to use computational methods to calculate the second virial coefficients of protein solutions. B_2 is related to molecular interactions in terms of the orientationally averaged potential of mean force (PMF), $W(r_{12})$, where r_{12} is the center-to-center distance,

$$B_2 = -2\pi \int_0^\infty (e^{-W(r_{12})/k_B T} - 1) r_{12}^2 \, \mathrm{d}r_{12}, \tag{1}$$

where W is the interaction free energy between two proteins, k_B is the Boltzmann constant, and T the temperature. Previous efforts to model the interaction free energy between two protein molecules and to compute B_2 have been based on idealized descriptions of proteins. The protein molecules are mostly treated as spheres, although Vilker *et al.* [10] modeled a protein (bovine serum albumin) as an ellipsoid. For spherical

model approaches, the interaction is normally divided into two parts: The first part is due to the excluded volume to account for the size of protein molecules and the second part accounts for the solution dependent effective interaction between protein molecules. Due to the spherical shape approximation of protein molecules, the thickness of the hydration layer is often considered as an adjustable parameter for B_2 calculations. The solution dependent contributions to B_2 are modeled using standard colloidal methods [11]. Namely the van der Waals interaction is treated in the Lifshitz-Hamaker framework and the electrostatic interaction [10,12–14] is obtained using the Poisson-Boltzmann approach. For such idealized spherical models, with adjustable parameters such as the Hamaker constant the computed B_2 have been partially successful to capture the trend of experimental data at various solution conditions.

Neal *et al.* [15] calculated the second virial coefficients by applying orientational dependence protein-protein interaction models. Electrostatic interactions in their study were obtained by distributing charges to the ionizable residues, thus an orientationally dependent charge distribution but treating the protein as a spherical dielectric body. The van der Waals interactions were calculated by a semiempirical approach. When the intermolecular distance is large enough, the Lifshitz-Hamaker approach [16] was implemented with the realistic shape of proteins in mind. At shorter distance, the Optimized Potentials for Liquid Simulations (OPLS) parameter set [17] was used to capture the short-range interaction. Even though the comparison between their calculations and experimental measurements yields large errors for some B_2 calculations, this approach did not use any further adjustable parameters.

The goal of our work is to develop a protein-protein interaction model to account for the realistic shape of proteins and at the same time to capture the effect of solutions without adjustable parameters. To this end, a residue level model of protein-protein interaction had been introduced [18,19].

In this model, each residue of a protein is represented by a sphere located at the geometric center of the residue determined by its native or approximate structure. The

011915-1

diameter of the sphere is determined by the molecular volume of a residue in a solution environment [20]. The molecular surface of our model protein is defined as the Richard-Connolly surface spanned by the union of these residue spheres using the MSMS program from Sanner [21]. Each residue carries a permanent dipole moment located at the center of its sphere and the direction of the dipole is given by the amino acid type from the protein's native structure. If a residue is charged the amount of charge is given by the Henderson-Hasselbalch equation using the generic pK_a values of residues if the local environmental effects on pK_a values are neglected. Alternatively experimental or calculated pK_a values of residues can be used to account for the local environments as it was done in this paper. For each residue there is also a polarizable dipole at the center of the sphere, whose nuclear polarizability had been determined from our recent work [22] and the electronic polarizability is estimated from optical dielectric constant augmented with quantum chemistry calculations [23]. There are three kinds of interactions in this model: the electrostatic interaction due to the electric double layer effect, the van der Waals attraction due to the polarizable dipoles and a short-range correction term to account for the short-range interactions such as the desolvation energy, hydrophobic interaction, and so on. In this article, we only consider the electrostatic interaction which gives the most contribution to the protein-protein interaction [24,25], the van der Waals interaction, and the short-range interaction which is accounted for using the excluded volume based upon the realistic shape of a protein.

The electrostatic problem in the electrostatic and the van der Waals interaction is solved using the Poisson-Boltzmann equation where the realistic shapes of protein molecules are considered. The Boundary Element Method (BEM) in combination with the Fast Multipole Method [26,27] is implemented to circumvent the extensive memory problem similar to the recent work by Lu *et al.* [28]. The validity of our model was already tested by binding affinity calculations of several protein complexes [29]. Direct comparisons between our calculations of B_2 and experimental measurements under various solution conditions were made and reasonable agreements from these comparisons provide further concrete evidence that our model can be used as a universal model for studies of nonspecific protein-protein interactions in aqueous solutions.

II. THEORETICAL DEVELOPMENT

A. General formulation for the second virial coefficient calculation using a residue level patch model

The osmotic second virial coefficient (B_2) can be expressed in terms of the interaction energy between two proteins [30]:

$$B_2 = -\frac{V}{2}\left[\frac{2Q_2(T)}{Q_1^2(T)} - 1\right], \qquad (2)$$

where Q_1 and Q_2 are one-protein and two-protein partition functions and V is the volume. Noting that the partition function involves the integration of the center of mass

$\mathbf{R} = (x,y,z)$ in a space-fixed Cartesian coordinate and the rotational coordinates $\Omega = (\alpha,\beta,\gamma)$ in Euler angles, we have

$$B_2 = -\frac{1}{128\pi^4 V}\int\cdots\int[e^{-W(\mathbf{R}_1,\Omega_1,\mathbf{R}_2,\Omega_2)/k_B T} - 1]$$
$$\times d\mathbf{R}_1 d\Omega_1 d\mathbf{R}_2 d\Omega_2, \qquad (3)$$

where the interaction potential W describes the anisotropic interaction between two proteins. $d\mathbf{R}_i = dx_i dy_i dz_i$ and $d\Omega_i = d\alpha_i \sin\beta_i d\beta_i d\gamma_i$. After a transformation to the center of mass of the protein pair and to the relative coordinates $\mathbf{R} = \mathbf{R}_1 - \mathbf{R}_2$, we assume that protein 1 is in the space-fixed coordinate, thus, the interaction potential W is now independent of Ω_1. The integration over the center of mass of the protein pair coordinate and Ω_1 for B_2 yields

$$B_2 = -\frac{1}{16\pi^2}\int_0^\infty\int_0^\pi\int_0^{2\pi}\int_0^{2\pi}\int_0^\pi\int_0^{2\pi}$$
$$\times [e^{-W(R,\theta,\phi,\alpha_2,\beta_2,\gamma_2)/k_B T} - 1]$$
$$\times R^2 dR \sin\theta d\theta d\phi d\alpha_2 \sin\beta_2 d\beta_2 d\gamma_2. \qquad (4)$$

In this expression of the orientation dependent potential $W(R,\theta,\phi,\alpha_2,\beta_2,\gamma_2)$, protein 2 moves around protein 1 and $(\alpha_2,\beta_2,\gamma_2)$ capture all of the orientations of protein 2 relative to the space-fixed coordinate (R,θ,ϕ) of protein 1 (Fig. 1).

In our residue level protein-protein interaction model, there are three contributions to the interaction energy $W(R,\theta,\phi,\alpha_2,\beta_2,\gamma_2)$. The electrostatic interaction and the van der Waals interaction will be obtained in the next two sections when the protein molecules are not in contact with each other. When the protein molecules are in contact a simple excluded volume model will be used. Thus, Eq. (4) can be split into two
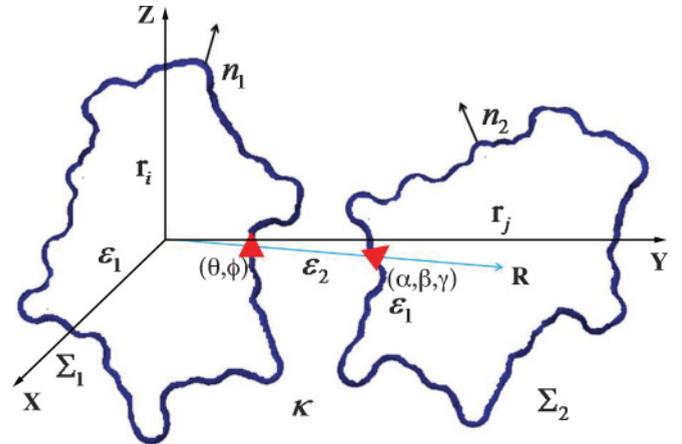
FIG. 1. (Color online) Schematic illustration showing the formulation of the electrostatic interaction between two proteins. The orientation of protein 1 (signified by a triangle on protein 1) is defined by two spherical coordinate angles (θ,ϕ) in a space-fixed coordinate ($\mathbf{X},\mathbf{Y},\mathbf{Z}$). The orientation of protein 2 (signified by a triangle on protein 2) is specified by Euler angles (α,β,γ) relative to the space-fixed coordinate. The molecular surfaces are defined by $\sum_1$ and $\sum_2$ for each protein and the n_1 and n_2 are the outward unit normals on $\sum_1$ and $\sum_2$. ε_1 and ε_2 are the dielectric constants of the protein cavity and the solution, respectively. κ represents the inverse Debye screening length. Charge q_i and dipole μ_i are located at the geometric center $\mathbf{r}_i$ of residue i.
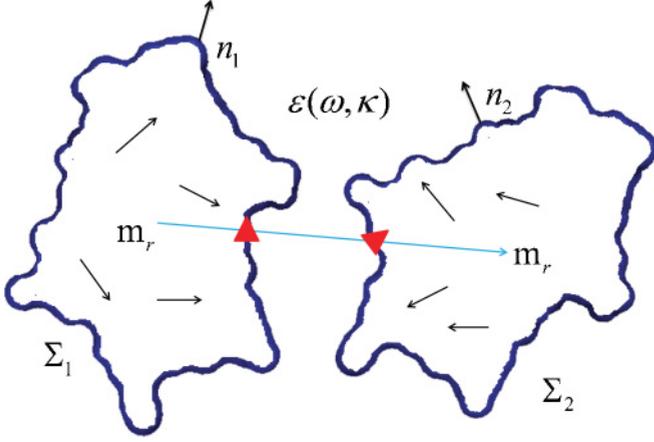
FIG. 2. (Color online) Schematic illustration showing the formulation of the van der Waals interaction of two proteins. The molecular surfaces are defined by $\sum_1$ and $\sum_2$ for each protein and the n_1 and n_2 are the outward unit normals on $\sum_1$ and $\sum_2$ and $\varepsilon(\omega,\kappa)$ is the dielectric constant of the outside solution as a function of the frequency ω and the inverse Debye screening length κ. The orientations of two proteins are defined by two surface patches(triangles) at the center-to-center distance R. $\mathbf{m}_r$s stand for the polarizable dipoles located at the residue centers.

parts, the hard core contribution and the rest:

$$
\begin{aligned}
B_2 = \frac{1}{16\pi^2} \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} &\left\{ \frac{1}{3} r_c^3 \right. \\
&\left. - \int_{r_c}^\infty (e^{-W(R,\theta,\phi,\alpha_2,\beta_2,\gamma_2)/k_B T} - 1) R^2 dR \right\} \\
&\times \sin\theta d\theta d\phi d\alpha_2 \sin\beta_2 d\beta_2 d\gamma_2,
\end{aligned}
\tag{5}
$$

where $r_c = r_c(\theta,\phi,\alpha_2,\beta_2,\gamma_2)$ is the distance between two protein molecules when the interaction becomes really high.

As the electrostatic and the van der Waals interactions are computed based upon the Boundary Element Method (BEM) of solving the Poisson-Boltzmann equations, a natural way to capture the detailed protein orientations is to set up a patch model by utilizing discrete elements of the molecular surface used in the BEM. To this end, let us assume there is N_1 triangular elements to represent protein 1, thus there are N_1 patches and each patch is specified by surface area σ_{1l} and r_{1l} centered at $(r_{1l},\theta_{1l},\phi_{1l})$. For protein 2, there are N_2 patches and each patch is specified by surface area σ_{2m} and r_{2m} centered at $(\alpha_{2m},\beta_{2m},\gamma_{2m})$, where r_{2m} is the distance from patch m to the center of mass of protein 2. For such a patch model, calculation of the interaction can be done explicitly,

$$
\begin{aligned}
B_2 = 2\pi \sum_{l=1}^{N_1} \sum_{m=1}^{N_2} \frac{\sigma_{1l}\sigma_{2m}}{\sigma_1 \sigma_2} \\
\times \left[\frac{1}{3} r_{clm}^3 - \int_{r_{clm}}^\infty (e^{-W_{lm}(R)/k_B T} - 1) R^2 dR \right],
\end{aligned}
\tag{6}
$$

where $r_{clm} = r_{1l} + r_{2m}$ is the distance between two proteins when the surface element l on protein 1 and the surface element m on protein 2 are in contact. σ_i is the surface area of protein i. $W_{lm}(R)$ is the interaction potential between two patches l

and m from proteins 1 and 2. In this article, the interaction energy between two protein molecules can be calculated by the sum of the electrostatic interaction energy and the van der Waals interaction energy at various patch combinations and distances,

$$
W_{lm}(R) = \Delta E_{\text{elec},lm}(R) + \Delta E_{\text{vdw},lm}(R),
\tag{7}
$$

which will be presented in the following sections.

B. General formulation of the electrostatic interaction free energy between two proteins with the boundary element method

Integral equations of the linearized Poisson-Boltzmann equation for two protein model were derived [19] following previous work from Juffer *et al.* [31] on the single protein problem. Consider the molecular surfaces $\sum_1$ and $\sum_2$ which cover two protein molecules, respectively. There are N charges q_i and dipoles $\vec{\mu}_i$ at position $\mathbf{r}_i$ enclosed by the surface $\sum_1$ and also there are N charges q_j and dipoles $\vec{\mu}_j$ at position $\mathbf{r}_j$ enclosed by the surface $\sum_2$. Inside each dielectric cavity the dielectric constant is ε_1 and the dielectric constant of the solution is given as ε_2 (see Fig. 1). The inverse Debye screening length κ is given by the solution's ionic strength and the temperature, $\kappa = \sqrt{\frac{2IF^2}{4\pi\varepsilon_0\varepsilon RT}} = \sqrt{\frac{I}{T}} \times (1.586\,115\,104)\dot{A}^{-1}$, where ε_0 is the permittivity of free space, ε is the dielectric constant of water, R is the gas constant, F is the Faraday constant, and I is the ionic strength of the electrolyte solution. The integral equations for the potential $\varphi_1(\mathbf{r})$ and $\varphi_2(\mathbf{r})$ and their gradient $\partial\varphi_1(\mathbf{r})/\partial(n_1)$ and $\partial\varphi_2(\mathbf{r})/\partial(n_2)$ on the molecular surfaces are [19]

$$
\begin{aligned}
&\frac{1}{2}\left(1 + \frac{\varepsilon_2}{\varepsilon_1}\right)\varphi_1(\mathbf{r_{01}}) + \iint_{\sum_1} L_1(\mathbf{r_1},\mathbf{r_{01}})\varphi_1(\mathbf{r_1})\,d\mathbf{r_1} \\
&+ \iint_{\sum_1} L_2(\mathbf{r_1},\mathbf{r_{01}})\frac{\partial\varphi_1(\mathbf{r_1})}{\partial n_1}d\mathbf{r_1} \\
&- \iint_{\sum_2} L_1(\mathbf{r_2},\mathbf{r_{01}})\varphi_2(\mathbf{r_2})d\mathbf{r_2} \\
&+ \iint_{\sum_2} L_2(\mathbf{r_2},\mathbf{r_{01}})\frac{\partial\varphi_2(\mathbf{r_2})}{\partial n_2}d\mathbf{r_2} \\
&= \sum_{i=1}^{2N}\{q_i F(\mathbf{r_i},\mathbf{r_{01}}) + \vec{\mu}_i \cdot \nabla F(\mathbf{r_i},\mathbf{r_{01}})\}/\varepsilon_1,
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
&\frac{1}{2}\left(1 + \frac{\varepsilon_1}{\varepsilon_2}\right)\frac{\partial\varphi_1(\mathbf{r_{01}})}{\partial n_1} + \iint_{\sum_1} L_3(\mathbf{r_1},\mathbf{r_{01}})\varphi_1(\mathbf{r_1})\,d\mathbf{r_1} \\
&+ \iint_{\sum_1} L_4(\mathbf{r_1},\mathbf{r_{01}})\frac{\partial\varphi_1(\mathbf{r_1})}{\partial n_1}d\mathbf{r_1} \\
&- \iint_{\sum_2} L_3(\mathbf{r_2},\mathbf{r_{01}})\varphi_2(\mathbf{r_2})d\mathbf{r_2} \\
&+ \iint_{\sum_2} L_4(\mathbf{r_2},\mathbf{r_{01}})\frac{\partial\varphi_2(\mathbf{r_2})}{\partial n_2}d\mathbf{r_2} \\
&= \sum_{i=1}^{2N}\left\{q_i \frac{\partial F}{\partial n_{01}}(\mathbf{r_i},\mathbf{r_{01}}) + \vec{\mu}_i \cdot \nabla \frac{\partial F}{\partial n_{01}}(\mathbf{r_i},\mathbf{r_{01}})\right\}\bigg/\varepsilon_1,
\end{aligned}
\tag{9}
$$

$$\frac{1}{2}\left(1+\frac{\varepsilon_2}{\varepsilon_1}\right)\varphi_2(\mathbf{r_{02}}) - \iint_{\sum_1} L_1(\mathbf{r_1},\mathbf{r_{02}})\varphi_1(\mathbf{r_1})\,d\mathbf{r_1}$$

$$+ \iint_{\sum_1} L_2(\mathbf{r_1},\mathbf{r_{02}})\frac{\partial\varphi_1(\mathbf{r_1})}{\partial n_1}d\mathbf{r_1}$$

$$+ \iint_{\sum_2} L_1(\mathbf{r_2},\mathbf{r_{02}})\varphi_2(\mathbf{r_2})d\mathbf{r_2}$$

$$+ \iint_{\sum_2} L_2(\mathbf{r_2},\mathbf{r_{02}})\frac{\partial\varphi_2(\mathbf{r_2})}{\partial n_2}d\mathbf{r_2}$$

$$= \sum_{i=1}^{2N}\{q_i F(\mathbf{r_i},\mathbf{r_{02}}) + \vec{\mu}_i\cdot\nabla F(\mathbf{r_i},\mathbf{r_{02}})\}/\varepsilon_1, \quad (10)$$

$$\frac{1}{2}\left(1+\frac{\varepsilon_1}{\varepsilon_2}\right)\frac{\partial\varphi_2(\mathbf{r_{02}})}{\partial n_2} - \iint_{\sum_1} L_3(\mathbf{r_1},\mathbf{r_{02}})\varphi_1(\mathbf{r_1})\,d\mathbf{r_1}$$

$$+ \iint_{\sum_1} L_4(\mathbf{r_1},\mathbf{r_{02}})\frac{\partial\varphi_1(\mathbf{r_1})}{\partial n_1}d\mathbf{r_1}$$

$$+ \iint_{\sum_2} L_3(\mathbf{r_2},\mathbf{r_{02}})\varphi_2(\mathbf{r_2})d\mathbf{r_2}$$

$$+ \iint_{\sum_2} L_4(\mathbf{r_2},\mathbf{r_{02}})\frac{\partial\varphi_2(\mathbf{r_2})}{\partial n_2}d\mathbf{r_2}$$

$$= \sum_{i=1}^{2N}\left\{q_i\frac{\partial F}{\partial n_{02}}(\mathbf{r_i},\mathbf{r_{02}}) + \vec{\mu}_i\cdot\nabla\frac{\partial F}{\partial n_{02}}(\mathbf{r_i},\mathbf{r_{02}})\right\}\Big/\varepsilon_1, \quad (11)$$

where

$$L_1(\mathbf{r},\mathbf{r_0}) = \frac{\partial F}{\partial n}(\mathbf{r},\mathbf{r_0}) - \frac{\varepsilon_2}{\varepsilon_1}\frac{\partial P}{\partial n}(\mathbf{r},\mathbf{r_0}), \quad (12)$$

$$L_2(\mathbf{r},\mathbf{r_0}) = P(\mathbf{r},\mathbf{r_0}) - F(\mathbf{r},\mathbf{r_0}), \quad (13)$$

$$L_3(\mathbf{r},\mathbf{r_0}) = \frac{\partial^2 F}{\partial n_0\partial n}(\mathbf{r},\mathbf{r_0}) - \frac{\partial^2 P}{\partial n_0\partial n}(\mathbf{r},\mathbf{r_0}), \quad (14)$$

$$L_4(\mathbf{r},\mathbf{r_0}) = -\frac{\partial F}{\partial n_0}(\mathbf{r},\mathbf{r_0}) + \frac{\partial P}{\partial n_0}(\mathbf{r},\mathbf{r_0})\frac{\varepsilon_1}{\varepsilon_2}, \quad (15)$$

and

$$F(\mathbf{r},\mathbf{r_0}) = \frac{1}{4\pi|\mathbf{r}-\mathbf{r_0}|},$$
$$P(\mathbf{r},\mathbf{r_0}) = \frac{e^{-\kappa|r-\mathbf{r_0}|}}{4\pi|\mathbf{r}-\mathbf{r_0}|}. \quad (16)$$

Although the traditional boundary element method such as Atkinson and his coworkers [32] can be used to solve the above integral equations, the memory requirement is too costly for current computers using either a direct linear system solver or an iterative solver, such as the Generalized Minimal Residual Method (GMRES) for a moderate size protein. In the current work the Fast Multipole Method is used and the details of our implementation will be outlined in Secs. II D and II E. Once the above integral equations are solved the potentials inside the dielectric cavity are

$$\varphi_1(\mathbf{r_1}) = -\iint_{\sum_1} L_1(\mathbf{r_1},\mathbf{r_{01}})\varphi_1(\mathbf{r_{01}})\,d\mathbf{r_{01}}$$

$$- \iint_{\sum_1} L_2(\mathbf{r_1},\mathbf{r_{01}})\frac{\partial\varphi_1(\mathbf{r_{01}})}{\partial n_{01}}d\mathbf{r_{01}}, \quad (17)$$

$$\varphi_2(\mathbf{r_2}) = -\iint_{\sum_2} L_1(\mathbf{r_2},\mathbf{r_{02}})\varphi_2(\mathbf{r_{02}})\,d\mathbf{r_{02}}$$

$$- \iint_{\sum_2} L_2(\mathbf{r_2},\mathbf{r_{02}})\frac{\partial\varphi_2(\mathbf{r_{02}})}{\partial n_{02}}d\mathbf{r_{02}}, \quad (18)$$

$$\nabla_1\varphi_1(\mathbf{r_1}) = -\iint_{\sum_1} \nabla_1 L_1(\mathbf{r_1},\mathbf{r_{01}})\varphi_1(\mathbf{r_{01}})\,d\mathbf{r_{01}}$$

$$- \iint_{\sum_1} \nabla_1 L_2(\mathbf{r_1},\mathbf{r_{01}})\frac{\partial\varphi_1(\mathbf{r_{01}})}{\partial n_{01}}d\mathbf{r_{01}}, \quad (19)$$

$$\nabla_2\varphi_2(\mathbf{r_2}) = -\iint_{\sum_2} \nabla_2 L_1(\mathbf{r_2},\mathbf{r_{02}})\varphi_2(\mathbf{r_{02}})\,d\mathbf{r_{02}}$$

$$- \iint_{\sum_2} \nabla_2 L_2(\mathbf{r_2},\mathbf{r_{02}})\frac{\partial\varphi_2(\mathbf{r_{02}})}{\partial n_{02}}d\mathbf{r_{02}}. \quad (20)$$

The electrostatic free energy between the protein molecules at a center-to-center distance, R, and relative orientations, $\Omega_1 = (\theta,\phi)$ and $\Omega_2 = (\alpha_2,\beta_2,\gamma_2)$, is given by

$$E_{\text{ele}}(R,\Omega_1,\Omega_2) = \sum_{i=1}^{N}\left\{\frac{q_i}{\epsilon_1}\varphi_1(\mathbf{r_i}) + \frac{1}{\epsilon_1}\vec{\mu}_i\cdot\nabla\varphi_1(\mathbf{r_i})\right\}$$

$$+ \sum_{j=1}^{N}\left\{\frac{q_j}{\epsilon_1}\varphi_2(\mathbf{r_j}) + \frac{1}{\epsilon_1}\vec{\mu}_j\cdot\nabla\varphi_2(\mathbf{r_j})\right\}. \quad (21)$$

Finally, the effective electrostatic interaction between two proteins is

$$\Delta E_{\text{ele}}(R,\Omega_1,\Omega_2) = E_{\text{ele}}(R,\Omega_1,\Omega_2) - E_{\text{ele}}(R\to\infty,\Omega_1,\Omega_2)$$

$$+ \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\varepsilon_1}\left\{q_i T_{ij}q_j - q_i\sum_{\alpha}T_{ij}^{\alpha}\mu_{j,\alpha}\right.$$

$$\left.+ \sum_{\alpha}\mu_{i,\alpha}T_{ij}^{\alpha}q_j - \sum_{\alpha\beta}\mu_{i,\alpha}T_{ij}^{\alpha\beta}\mu_{j,\beta}\right\}, \quad (22)$$

where the interaction tensors for charge-charge, charge-dipole, and dipole-dipole are given by

$$T_{ij} = \frac{e^{-\kappa r_{ij}}}{r_{ij}},$$

$$T_{ij}^{\alpha} = \nabla_\alpha T_{ij} = e^{-\kappa r_{ij}}\frac{(1+\kappa r_{ij})}{r_{ij}^3}r_{ij,\alpha},$$

$$T_{ij}^{\alpha\beta} = \nabla_\alpha\nabla_\beta T_{ij} = e^{-\kappa r_{ij}}\left\{\left(\frac{3}{r_{ij}^5} + \frac{3\kappa}{r_{ij}^4} + \frac{\kappa^2}{r_{ij}^3}\right)r_{ij,\alpha}r_{ij,\beta}\right.$$

$$\left.- \left(\frac{1}{r_{ij}^3} + \frac{\kappa}{r_{ij}^2}\right)\delta_{\alpha\beta}\right\}. \quad (23)$$

Here ∇_α is $\frac{\partial}{\partial r_{ij,\alpha}}$ for each $\alpha = x,y,z$ and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. The last summation terms in Eq. (22) are the interaction energy between charges and dipoles in two proteins when the solution has the same dielectric constant ε_1 as inside the protein and with the inverse Debye screening length κ.

C. General formulation of the van der Waals interaction free energy

The van der Waals interaction energy between two proteins is defined as

$$\Delta E_{vdw}(R,\Omega_1,\Omega_2) = E_{vdw}(R,\Omega_1,\Omega_2)$$
$$- E_{vdw}(R \to \infty, \Omega_1, \Omega_2). \quad (24)$$

Song and Zhao [18] formulated the van der Waals interaction between the protein molecules in an electrolyte solution using the following effective action in Fourier space for polarizable dipoles $\mathbf{m_{r,n}}$:

$$S[\mathbf{m_{r,n}}] = -\frac{\beta}{2} \sum_{\mathbf{r}} \sum_{n=-\infty}^{n=\infty} \frac{1}{\alpha_{\mathbf{r,n}}} \mathbf{m_{r,n}} \cdot \mathbf{m_{r,-n}}$$

$$+ \frac{\beta}{2} \sum_{\mathbf{r} \neq \mathbf{r'}} \sum_{n=-\infty}^{n=\infty} \frac{1}{\alpha_{\mathbf{r,n}}} \mathbf{m_{r,n}} \cdot T(\mathbf{r-r'}) \cdot \mathbf{m_{r,-n}}$$

$$+ \frac{\beta}{2} \sum_{\mathbf{r,r'}} \sum_{n=-\infty}^{n=\infty} \frac{1}{\alpha_{\mathbf{r,n}}} \mathbf{m_{r,n}} \cdot R_n(\mathbf{r-r'}) \cdot \mathbf{m_{r,-n}}, \quad (25)$$

where $\alpha_{\mathbf{r,n}}$ is the frequency-dependent polarizability of a residue located at position $\mathbf{r}$. $T(\mathbf{r} - \mathbf{r'})$ is the dipole-dipole interaction tensor between $\mathbf{r}$ and $\mathbf{r'}$. $R_n(\mathbf{r} - \mathbf{r'})$ is the reaction field tensor at the Matsubara frequency $\omega_n = 2\pi n/\beta\hbar$ (see Fig. 2). If the electrolyte solvent is treated by the Debye-Hückel theory, this reaction field tensor can be calculated by solving the Poisson-Boltzmann equation with the dielectric constant $\varepsilon(i\omega_n)$. The quantum partition function from this effective action of the system is

$$Q(R,\Omega_1,\Omega_2) = \prod_n \left[\frac{2\pi}{\beta \det A_n}(R,\Omega_1,\Omega_2) \right]^{1/2}, \quad (26)$$

where A_n's matrix element is given by

$$A_n(\mathbf{r,r'}) = \frac{1}{\alpha_{\mathbf{r,n}}}\delta_{\mathbf{r,r'}} - T(\mathbf{r} - \mathbf{r'}) - R_n(\mathbf{r} - \mathbf{r'}), \quad (27)$$

where $\mathbf{r}$ and $\mathbf{r'}$ represent residues in each protein. For the rigid residue level model in our work, the residue positions $\mathbf{r}$ and $\mathbf{r'}$ are completely determined by (R,Ω_1,Ω_2) in the spaced-fixed coordinate. The symbol "det" represents the determinant of the matrix. Therefore, the van der Waals interaction free energy is given by

$$\Delta E_{vdw} = \frac{1}{2}k_B T \sum_{n=-\infty}^{n=\infty} [\ln\{\det A_n(R,\Omega_1,\Omega_2)\}$$
$$- \ln\{\det A_n(R \to \infty)\}]. \quad (28)$$

In order to evaluate the van der Waals interaction in our model, the reaction field matrix $R_n(\mathbf{r} - \mathbf{r'})$ has to be calculated using the properties of the proteins and the solution. The boundary element formulation which is used to evaluate the electrostatic free energy can also be used to calculate the reaction field matrix. Consider two molecular surfaces $\sum_1$ and $\sum_2$ spanned by two protein molecules. There are N polarizable dipoles $\mathbf{m_r}$ at position $\mathbf{r}$ enclosed by each surface $\sum_1$ and $\sum_2$. Inside this dielectric cavity the dielectric constant is one and the dielectric constant of the solution is $\varepsilon(i\omega_n)$ at the Matsubara frequency ω_n. The inverse Debye screening length κ is given by the solution's ionic strength and the temperature. If we recognize that in order to calculate the potential at the molecular surface

a dipole $\mathbf{m}$ at position $\mathbf{r}_0$ can be described by an effective charge density $\rho_{eff}(\mathbf{r}) = -\mathbf{m}\nabla\delta(\mathbf{r} - \mathbf{r}_0)$ [33], the reaction field matrix involving residues $\mathbf{r}_i$ and $\mathbf{r}_j$ can be given as

$$R(\mathbf{r}_i,\mathbf{r}_j) = \iint_{\sum_p} [\nabla_i F(\mathbf{r}_i,\mathbf{r}_j)$$

$$- \nabla_i P(\mathbf{r}_i,\mathbf{r}_j)]\frac{\partial \varphi_p}{\partial n_p}(\mathbf{r}_j,\mathbf{r}_p)d\mathbf{r}_p$$

$$+ \iint_{\sum_p} \left[-\nabla_i \frac{\partial F}{\partial n_j} F(\mathbf{r}_i,\mathbf{r}_j) \right.$$

$$\left. + \nabla_i \frac{\partial P}{\partial n_j}(\mathbf{r}_i,\mathbf{r}_j)\varepsilon \right] \varphi_p(\mathbf{r}_j,\mathbf{r}_p)d\mathbf{r}_p, \quad (29)$$

where F and P are defined in Eq. (16) and for p to be 1 or 2 depends upon $\mathbf{r}_j$ in $\sum_1$ or $\sum_2$. The potential and its gradient φ_p and $\partial\varphi_p$ on the molecular surface p due to residue i can be obtained by solving the following integral equations [19,31]:

$$\frac{1}{2}[1 + \varepsilon(i\omega_n)]\varphi_1(\mathbf{r}_i,\mathbf{r}_{01})$$

$$+ \iint_{\sum_1} L_1(\mathbf{r}_1,\mathbf{r}_{01})\varphi_1(\mathbf{r}_i,\mathbf{r}_1)\,d\mathbf{r}_1$$

$$+ \iint_{\sum_1} L_2(\mathbf{r}_1,\mathbf{r}_{01})\frac{\partial\varphi_1}{\partial n_1}(\mathbf{r}_i,\mathbf{r}_1)\,d\mathbf{r}_1$$

$$- \iint_{\sum_2} L_1(\mathbf{r}_2,\mathbf{r}_{01})\varphi_2(\mathbf{r}_i,\mathbf{r}_2)\,d\mathbf{r}_2$$

$$+ \iint_{\sum_2} L_2(\mathbf{r}_2,\mathbf{r}_{01})\frac{\partial\varphi_2}{\partial n_2}(\mathbf{r}_i,\mathbf{r}_2)\,d\mathbf{r}_2$$

$$= \nabla_i F(\mathbf{r}_i,\mathbf{r}_{01}), \quad (30)$$

$$\frac{1}{2}\left(1 + \frac{1}{\varepsilon(i\omega_n)}\right)\frac{\partial\varphi_1}{\partial n_1}(\mathbf{r}_i,\mathbf{r}_{01})$$

$$+ \iint_{\sum_1} L_3(\mathbf{r}_1,\mathbf{r}_{01})\varphi_1(\mathbf{r}_i,\mathbf{r}_1)\,d\mathbf{r}_1$$

$$+ \iint_{\sum_1} L_4(\mathbf{r}1,\mathbf{r}_{01})\frac{\partial\varphi_1}{\partial n_1}(\mathbf{r}_i,\mathbf{r}_1)\,d\mathbf{r}_1$$

$$- \iint_{\sum_2} L_3(\mathbf{r}_2,\mathbf{r}_{01})\varphi_2(\mathbf{r}_i,\mathbf{r}_2)\,d\mathbf{r}_2$$

$$+ \iint_{\sum_2} L_4(\mathbf{r}2,\mathbf{r}_{01})\frac{\partial\varphi_2}{\partial n_2}(\mathbf{r}_i,\mathbf{r}_2)\,d\mathbf{r}_2$$

$$= \nabla_i \frac{\partial F}{\partial n_{01}}(\mathbf{r}_i,\mathbf{r}_{01}), \quad (31)$$

$$\frac{1}{2}[1 + \varepsilon(i\omega_n)]\varphi_2(\mathbf{r}_i,\mathbf{r}_{02})$$

$$- \iint_{\sum_1} L_1(\mathbf{r}_1,\mathbf{r}_{02})\varphi_1(\mathbf{r}_i,\mathbf{r}_1)\,d\mathbf{r}_1$$

$$+ \iint_{\sum_1} L_2(\mathbf{r}_1,\mathbf{r}_{02})\frac{\partial\varphi_1}{\partial n_1}(\mathbf{r}_i,\mathbf{r}_1)\,d\mathbf{r}_1$$

$$+ \iint_{\sum_2} L_1(\mathbf{r}_2,\mathbf{r}_{02})\varphi_2(\mathbf{r}_i,\mathbf{r}_2)\,d\mathbf{r}_2$$

$$+ \iint_{\sum_2} L_2(\mathbf{r}_2,\mathbf{r}_{02})\frac{\partial\varphi_2}{\partial n_2}(\mathbf{r}_i,\mathbf{r}_2)\,d\mathbf{r}_2$$

$$= \nabla_i F(\mathbf{r}_i,\mathbf{r}_{02}), \quad (32)$$

$$\frac{1}{2}\left(1 + \frac{1}{\varepsilon(i\omega_n)}\right)\frac{\partial \varphi_2}{\partial n_2}(\mathbf{r}_i, \mathbf{r}_{02})$$

$$- \iint_{\sum_1} L_3(\mathbf{r}_1, \mathbf{r}_{02})\varphi_1(\mathbf{r}_i, \mathbf{r}_1)\, d\mathbf{r}_1$$

$$+ \iint_{\sum_1} L_4(\mathbf{r}1, \mathbf{r}_{02})\frac{\partial \varphi_1}{\partial n_1}(\mathbf{r}_i, \mathbf{r}_1)\, d\mathbf{r}_1$$

$$+ \iint_{\sum_2} L_3(\mathbf{r}_2, \mathbf{r}_{02})\varphi_2(\mathbf{r}_i, \mathbf{r}_2)\, d\mathbf{r}_2$$

$$+ \iint_{\sum_2} L_4(\mathbf{r}2, \mathbf{r}_{02})\frac{\partial \varphi_2}{\partial n_2}(\mathbf{r}_i, \mathbf{r}_2)\, d\mathbf{r}_2$$

$$= \nabla_i \frac{\partial F}{\partial n_{02}}(\mathbf{r}_i, \mathbf{r}_{02}), \tag{33}$$

where L_1, L_2, L_3, and L_4 are defined in Eqs. (12), (13), (14), and (15). To evaluate the van der Waals interaction energy in Eq. (28), the reaction field matrix should be built corresponding to the dielectric constant $\varepsilon(i\omega_n)$ for each frequency ω_n. The total polarizability of a residue in a protein is

$$\alpha_n = \alpha(i\omega_n) = \frac{\alpha_{nu}}{1 + \omega_n/\omega_{rot}} + \frac{\alpha_{el}}{1 + (\omega_n/\omega_I)^2}, \tag{34}$$

where α_{nu} is the static nuclear polarizability of a residue [22] and ω_{rot} is a characteristic frequency of nuclear collective motion from a generalization of the Debye model. α_{el} is the static electronic polarizability of a residue and ω_I is the ionization frequency of a residue as in the Drude oscillator model of electronic polarizabilities. $\omega_{\mathrm{rot}} = 20$ cm^{-1} for this calculation which is typical rotational frequency of molecules [34]. Further improvements may be archived if individual rotational frequencies are used for each amino acid type used in the method described in Ref. [22]. Other properties listed in Table I from Kim *et al.* [29] are based on the calculated results from Millefiori *et al.* [23] An accurate parametrization of the dielectric function $\varepsilon(i\omega)$ of water based on the experimental data is taken from Parsegian's work [35].

D. Solving the linear system: The iterative double-tree fast multipole method

The integral equations Eqs. (8), (9), (10), and (11) for the electrostatic interaction energy and Eqs. (30), (31), (32), and (33) for the van der Waals interaction energy will become a linear system once a basis set is constructed over molecular surfaces,

$$(I - L)A = B, \tag{35}$$

where A and B are single column vectors with the size of $2N$, where N is the number of surface elements on the protein molecules for the electrostatic energy calculation and will be the $(2M) \times (2N)$ matrix for the reaction field calculation of the van der Waals energy calculation, where M is the number of residues in a protein. More explicitly,

$$I\begin{pmatrix} \varphi_{00} \\ \varphi_{11} \\ \varphi_{22} \\ \varphi_{33} \end{pmatrix} - \begin{pmatrix} L_1^{00} & L_2^{01} & L_1^{02} & L_2^{03} \\ L_3^{10} & L_4^{11} & L_3^{12} & L_4^{13} \\ L_1^{20} & L_2^{21} & L_1^{22} & L_2^{23} \\ L_3^{30} & L_4^{31} & L_3^{32} & L_4^{33} \end{pmatrix}\begin{pmatrix} \varphi_{00} \\ \varphi_{11} \\ \varphi_{22} \\ \varphi_{33} \end{pmatrix} = \begin{pmatrix} F_{00} \\ F_{11} \\ F_{22} \\ F_{33} \end{pmatrix},$$
$$\tag{36}$$

where I is the identity matrix with the size of $(2N) \times (2N)$. φ_{00} and φ_{11} are the potential and the gradient of potential on surface 1, and φ_{22} and φ_{33} are the corresponding ones on surface 2. The matrix element, L_1, L_2, L_3, and L_4 are defined in Eqs. (12), (13), (14), and (15), and the upper indices are the equation indices from Eqs. (8) to (11) or from Eqs. (30) to (33) for the electrostatic interaction and the van der Waals interaction, respectively, according to the φ's indices. If the distance between two proteins is large, the contribution from the matrix elements in indices 02, 03, 12, 13 and 20, 21, 30, 31 to the matrix-vector multiplications is relatively small in comparison with other matrix elements. Thus, the one-body problem can be solved first and the cross-body contributions can be treated perturbatively,

$$\begin{pmatrix} \bar{\varphi}_{00} & 0 \\ 0 & \bar{\varphi}_{11} \end{pmatrix} - \begin{pmatrix} L_1^{00} & L_2^{01} \\ L_3^{10} & L_4^{11} \end{pmatrix}\begin{pmatrix} \bar{\varphi}_{00} \\ \bar{\varphi}_{11} \end{pmatrix} = \begin{pmatrix} F_{00} \\ F_{11} \end{pmatrix}, \tag{37}$$

$$\begin{pmatrix} \bar{\varphi}_{22} & 0 \\ 0 & \bar{\varphi}_{33} \end{pmatrix} - \begin{pmatrix} L_1^{22} & L_2^{23} \\ L_3^{32} & L_4^{33} \end{pmatrix}\begin{pmatrix} \bar{\varphi}_{22} \\ \bar{\varphi}_{33} \end{pmatrix} = \begin{pmatrix} F_{22} \\ F_{33} \end{pmatrix}, \tag{38}$$

where

$$\varphi_{ii} = \bar{\varphi}_{ii} + \delta\varphi_{ii} \tag{39}$$

and $i = 0, 1, 2, 3$. Substituting Eq. (39) into Eq. (36) and using the definition from Eqs. (37) and (38) yield a new system of linear equations,

$$I\begin{pmatrix} \delta\varphi_{00} \\ \delta\varphi_{11} \\ \delta\varphi_{22} \\ \delta\varphi_{33} \end{pmatrix} - \begin{pmatrix} L_1^{00} & L_2^{01} & L_1^{02} & L_2^{03} \\ L_3^{10} & L_4^{11} & L_3^{12} & L_4^{13} \\ L_1^{20} & L_2^{21} & L_1^{22} & L_2^{23} \\ L_3^{30} & L_4^{31} & L_3^{32} & L_4^{33} \end{pmatrix}\begin{pmatrix} \delta\varphi_{00} \\ \delta\varphi_{11} \\ \delta\varphi_{12} \\ \delta\varphi_{13} \end{pmatrix}$$
$$= \begin{pmatrix} L_1^{02}\bar{\varphi}_{22} + L_2^{03}\bar{\varphi}_{33} \\ L_3^{12}\bar{\varphi}_{22} + L_4^{13}\bar{\varphi}_{33} \\ L_1^{20}\bar{\varphi}_{00} + L_2^{21}\bar{\varphi}_{11} \\ L_3^{30}\bar{\varphi}_{00} + L_4^{31}\bar{\varphi}_{11} \end{pmatrix}. \tag{40}$$

The same argument can be made for the above linear system, hence this linear system can be reduced to the following two linear systems with the order $O(N^2)$:

$$\begin{pmatrix} \delta\varphi_{00} & 0 \\ 0 & \delta\varphi_{11} \end{pmatrix} - \begin{pmatrix} L_1^{00} & L_2^{01} \\ L_3^{10} & L_4^{11} \end{pmatrix}\begin{pmatrix} \delta\varphi_{00} \\ \delta\varphi_{11} \end{pmatrix}$$
$$= \begin{pmatrix} L_1^{02}\varphi_{22} + L_2^{03}\varphi_{33} \\ L_3^{12}\varphi_{22} + L_4^{13}\varphi_{33} \end{pmatrix}, \tag{41}$$

$$\begin{pmatrix} \delta\varphi_{22} & 0 \\ 0 & \delta\varphi_{33} \end{pmatrix} - \begin{pmatrix} L_1^{22} & L_2^{23} \\ L_3^{32} & L_4^{33} \end{pmatrix}\begin{pmatrix} \delta\varphi_{22} \\ \delta\varphi_{33} \end{pmatrix}$$
$$= \begin{pmatrix} L_1^{20}\varphi_{00} + L_2^{21}\varphi_{11} \\ L_3^{30}\varphi_{00} + L_4^{31}\varphi_{11} \end{pmatrix}. \tag{42}$$

To solve the system of linear equations in Eq. (36), we first solve the one-body linear systems in Eqs. (37) and (38), then the right-hand side vectors in Eqs. (41) and (42) are obtained from the previous solutions of the one-body problem and the cross-matrix elements from the two-bodies. The perturbations $\delta\varphi$ are computed after solving two linear systems in Eqs. (41)
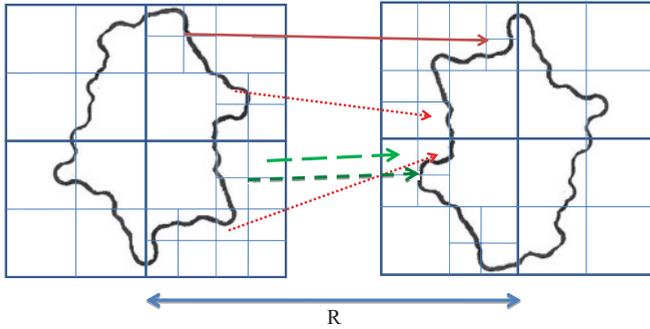
FIG. 3. (Color online) Schematic illustration showing the double tree Fast Multiple Method (dt-FMM). Two tree structures are set up with the center-to-center distance (R). On level = 2, all the Multipole-to-Local (M2L) translations are computed for far-field interactions. On level = 3, the long interaction (solid line) is not allowed in the M2L translation list (the interaction list) but the interaction (dashed line) is allowed. On level = 4, long interactions (dotted lines) are not allowed but the interaction within the interaction list (long dashed line) is computed.

and (42). The new solution φ is the sum of the one-body solution and the perturbation solutions from Eqs. (41) and (42). By solving Eqs. (41) and (42) using the new φ a close loop is set up to solve the problem iteratively. In this iterative method, we only need one matrix-vector product operation between two separated bodies in each iteration. This iteration is called the "outer" iteration to separate the term with the "inner" iteration which is used to solve the one-body linear system with an iterative solver, such as GMRES. The "outer" iteration can reduce the size of system from $O(2N \times 2N)$ to $O(N \times N)$ and the "inner" iteration can be accelerated by introducing the Fast Multipole Method (FMM) [29]. Figure 3 shows how the double tree structures are defined to cover one body in one tree and the interactions between two separated bodies are allowed in the FMM algorithm to calculate matrix-vector products in Eqs. (41) and (42) to calculate the right-hand side vectors.

This double-tree FMM with "outer" iterative method has an advantage that can reduce the computational cost from the traditional direct Boundary Element Method, $O[(2N)^2]$ to the one of the single-body problem, $O(N)$. But the drawback is that the closest distance between two bodies has to be that there is no overlap of trees in this double-tree FMM. For example, the closest center-to-center distance between two BPTI proteins in the crystal lattice structure is about the range of 24–28 Å, but it should be more than 33Å in double-tree FMM to avoid the tree overlapping. The accuracy of the double-tree FMM is going to be worse if two trees are getting close (as will be seen in Fig. 7). In this case, the number of the "outer" iteration is also getting larger, thus, the overall performance will be slower. In general, the double-tree FMM is useful when the center-to-center distance is about 1.5–2 times longer than the size of the tree.

E. Solving the linear system: The single-tree fast multipole method

In order to calculate the interaction energy when two bodies are too close to be reliable using the double-tree FMM, we introduce the single-tree FMM in Fig. 4 . This method is based on the single-body FMM [29]. The system of linear equations
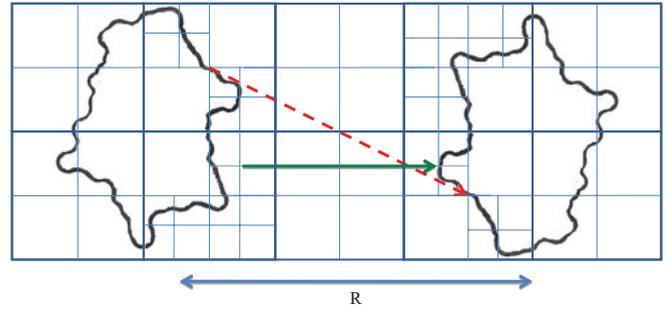

FIG. 4. (Color online) Schematic illustration showing the single tree Fast Multiple Method (st-FMM). Only one tree is set up to cover two surfaces of proteins with the center-to-center distance (R). On level = 2, only the Multipole-to-Local (M2L) translations which are in the interaction list (solid line) are computed but the long interaction (dashed line) is not allowed for the M2L translation.

from Eqs. (8), (9), (10), and (11) for the electrostatic interaction and Eqs. (30), (31), (32), and (33) for the van der Waals interaction can be described by the equations of a single body. One subtle complication is the additional negative signs of L_1^{02}, L_3^{12}, L_1^{20}, and L_3^{30} in Eq. (36) where the signs of gradients are changed because of the convention used for outside normal at the cavity surfaces. Thus we need to consider this sign change when the integral is performed on the surface of one body when the source is in another body. In the traditional single-body FMM, there is no way to deal with this conventional change,
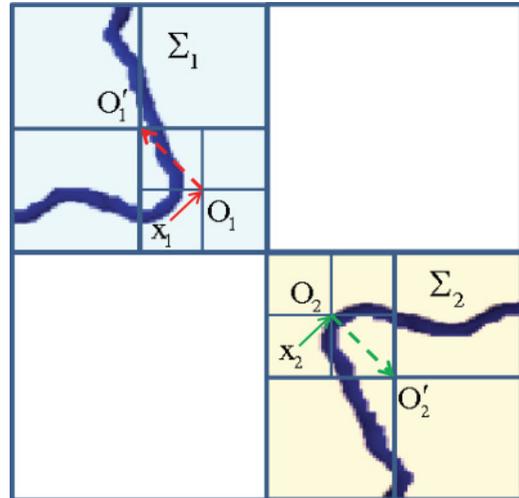

FIG. 5. (Color online) Schematic illustration showing the single-tree Fast Multiple Method (st-FMM) in level = 2 to level = 5. $\sum_1$ and $\sum_2$ are the surfaces of two proteins. All cells with light shade belong to the surface $\sum_1$ and cells with lighter shade belong to the surface $\sum_2$, respectively. From the lowest level, level = 5, the surface index (either 1 or 2) is transferred from the level = 5 center x_1 or x_2 to the level = 4 center O_1 or O_2 by Multipole-to-Multiple (M2M) translations. This index also can be transferred to the upper level's cell. For example, on level = 3 the center O_1' or O_2' has the surface index during the process of M2M translation. The arrows in O_1 cell indicate the flow of the surface index 1 and the arrows in O_2 cell for the surface index 2. The dashed arrows represent level = 5 to level = 4 M2M translations and solid arrows represent level = 4 to level = 3 M2M translations, respectively.

but this problem can be solved by transferring the additional information of the ownership of surface elements during the process of Multipole-to-Multipole(M2M) and Local-to-Local (L2L) translations. Figure 5 shows the details how the ownership of each surface element in a leaf cell can be transferred to the parent's cell in FMM.

Because this single-tree FMM is based on the single-body FMM, the computational cost follows the order $O(2N)$, that is about twice more than the one of the double-tree FMM algorithm. Even though the single-tree FMM takes twice more memory than the double-tree FMM, this cost is still highly competitive compared with the traditional direct Boundary Element Method. Figure 6 shows that the direct BEM follows the quadratic increase as a function of the number of surface elements and two FMMs follow only the linear increase via order $O(N)$ or $O(2N)$ for the double and single-tree FMM, respectively.

To test both FMM methods, we applied them to the electrostatic interaction energy calculation of two identical spheres. According to Fig. 7 , both solutions gave correct effective electrostatic interaction energies compared with the analytic solution of two identical spheres based on Eq. (A13) in the Appendix. Furthermore, we had the consistent results by two FMM methods when the effective electrostatic interaction energies between the two BPTI molecules are computed. Also these results were compared to the result from the direct BEM solver and we found that the single-tree FMM is slightly more accurate when two particles are getting closer and the double-tree FMM is more accurate when two particles are farther than twice of the size of a particle. So we used both FMM methods to calculate the effective interaction energy between two protein molecules.

F. Preparation of protein structures

The bovine pancreatic trypsin inhibitor (BPTI) is used to validate our model by calculating the osmotic second virial coefficients because it is a relatively small protein (the
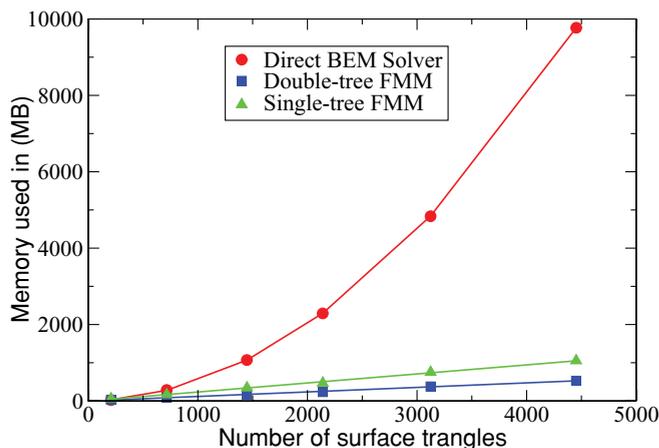

FIG. 6. (Color online) Memory cost comparison between the direct Boundary Element Method (BEM) in solid circles, the double-tree FMM (solid squares), and the single-tree FMM (solid upper triangles). The number of surface elements is the number of surface elements from a single protein (N). So the order of each method is $O[(2N)^2]$ for the direct BEM, $O(N)$ for the double-tree FMM and $O(2N)$ for the single-tree FMM, respectively.
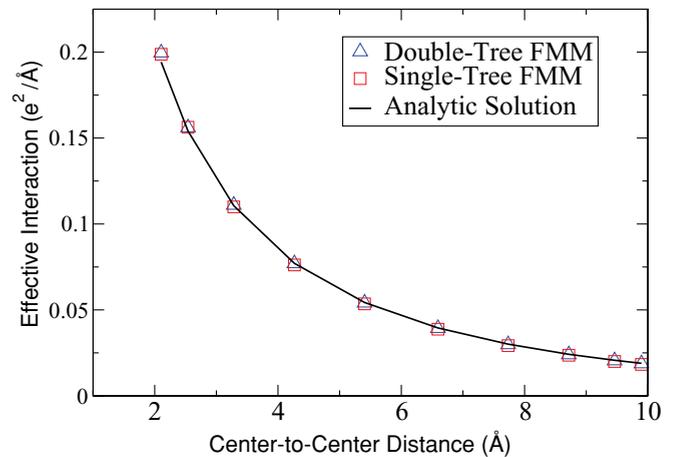

FIG. 7. (Color online) Effective electrostatic interaction energy comparison between the analytic solution (solid line) from Eq. (A13) and the solutions of the double-tree FMM (upper triangle) and the single-tree FMM (square). The radius of both spheres is 1.0Å and the unit charge is located at the center of each sphere. The inverse Debye screening length is 0.1Å^{-1} and the dielectric constant is 1.0 inside the spheres and 10.0 outside spheres.

number of residues is 58), the structure is well known and the experimental B_2 data are known from Farnum and Zukoski [5]. We will use the anisotropic patch model introduced in Sec. II A by treating surface elements as patches to define the anisotropic interactions between two protein molecules. Because of the large number of patches on the protein surface, it is really time consuming to compute interaction energies of all patch pairs. To reduce the number of calculations for patch pairs between two protein molecules, we only consider the most probable configurations of pair interactions between two protein molecules. To this end, a natural starting point is to consider the patch pairs appearing in the crystal structure (PDB code = 6PTI). The crystal space group of BPTI for this structure is $P2_12_12$. Using the transformation matrix given in the PDB file, other unit cell elements, B, C, and D can be obtained from the original structure, A (Fig. 8). For example, B is generated from the symmetry operation $(\bar{x}, \bar{y}, z)$, which leads to an AB pair configuration. The opposite direction $(x, y, \bar{z})$ leads to an additional AB′ pair configuration. From this PDB structural information we have all six pairs of interactions, AB, AC, AD, AB′, AC′, and AD′. Figure 9 describes the relative orientations of BPTI elements in a unit cell.

Using our residue level model and the CHARMMING web portal [38], the positions of residues of protein pairs, the charges, and the dipole moments can be generated. The calculations of the osmotic second virial coefficients of the BPTI protein in solutions are performed using the solution conditions from Farnum and Zukoski [5]. The temperature of the solution is 20°C which is used both in the calculation of B_2 from Eq. (6) and in the inverse Debye screening length. The pH of the solution, 4.9, is used to calculate the charge of each amino acid residue in the protein using the Henderson-Hasselbalch equation and the pK_a of the residues are calculated by PROPKA 2.0 [39]. The generic pK_a values of amino acids are not used because the local pK_a of a residue which is either burred inside the protein or on the surface of
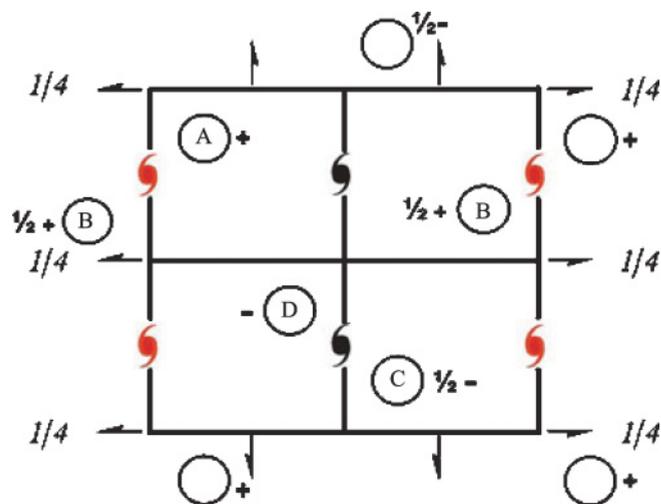
FIG. 8. (Color online) 2D illustration shows the unit cell of the point group $P2_12_12$. In unit cell, there are four elements indicated by the capital letters: A is at the origin of coordinate system and its symmetrical operation is (x,y,z), B can be obtained by the operation $(\bar{x},\bar{y},z)$, C can be obtained by the operation $(1/2 + x,1/2 + \bar{y},\bar{z})$, and D can be obtained by the operation $(1/2 + \bar{x},1/2 + y,\bar{z})$. All the notations follows the Hermann-Mauguin symmetry notation and the style of Wondratschek and Müller [36]. This diagram is adapted from Jasinski and Foxman [37].

the protein may have a shifted pK_a as the case of P_1 Glu and P_1 Asp mutations in the BPTI-trypsin complexes [29] and the PROPKA 2.0 is an accurate program for the pK_a prediction [40]. The dependence of B_2 of BPTI molecules on the concentration of the sodium chloride solution and the comparison with the experimental B_2 data will be described in Sec. III.

To test the reliability of the small sampling in relative orientations of proteins, we increased the number of relative orientations up to 10 and converged results are obtained for all the NaCl concentration dependence of BPTI B_2.
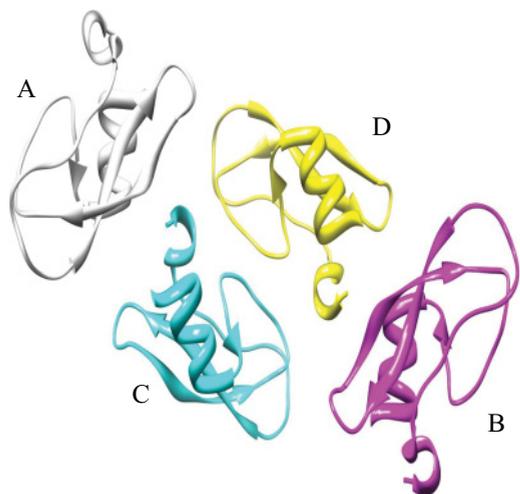

FIG. 9. (Color online) 3D illustration shows the relative orientations of all BPTI molecules in a unit cell of $P2_12_12$. Ribbon structures labeled element A, B, C, and D are shown. UCSF Chimera [41] was used to draw this figure.

The converged results are a little bit different from the six orientations' results, but the comparison with experiential results remains the same. Thereafter, all of our calculations are done with six orientations sampled from the protein's crystal structure.

In addition to the calculations of the second virial coefficients of BPTI as a function of the concentration of the sodium chloride solution, we also calculated the osmotic second virial coefficients of lysozyme in various solution conditions. To generate the most probable configurations of pair interactions between two lysozyme molecules, the crystal structure (PDB code = 2ZQ3) is used. In this case, the crystal space group of lysozyme is $P2_12_12_1$. Again, we apply the transformation matrix given in the PDB file to the original structure, A, to generate other unit cell elements, B, C, and D. As in the BPTI case, six pairs of relative orientations, AB, AC, AD, AB′, AC′, and AD′ are generated.

The calculations of the osmotic second virial coefficients of the lysozyme in solutions are performed using the same conditions as in [6] and [9]. The concentration dependence from 2% to 7% of salt concentration, the pH dependence from pH = 4.0 to pH = 5.4 and the temperature dependence from 25°C to 5°C are used for the sodium chloride solution. The concentration dependence from $0.50M$ to $1.10M$ of the ammonium chloride solution is used at pH = 4.5 and temperature 18°C. The concentration dependence from $0.10M$ to $0.70M$ of the magnesium bromide solution at pH = 7.8 and temperature 23°C are also calculated. Comparisons between calculated B_2 and experimental ones will be presented in Sec. III using the experimental data from Guo *et al.* [6] and additional data for the magnesium bromide salt from Tessier *et al.* [9].

III. RESULTS

The electrostatic interaction energies and the van der Waals interaction energies between two BPTI molecules are calculated by the single-tree FMM algorithm when the center-to-center distance R between two proteins is less than twice the size of the protein and by the double-tree FMM when the center-to-center distance is greater. Figure 10 shows the interaction energy changes as a function of R, relative orientations, and the inverse Debye screening length κ. The results agreed with our previous findings [18,19] that the electrostatic and van der Waals interactions are sensitive to the relative orientations. The ionic strength affects the electrostatic interactions much more than the van der Waals interactions. From these calculated interaction energies, B_2 can be obtained from Eq. (6), where the contact distances and patch surface areas can be obtained from the molecular surfaces used in the BEM calculations.

The soft interaction contribution (the electrostatic and the van der Waals contribtion) to B_2 is calculated using the six pair configurations to represent all orientational dependence of the soft interaction potential. Figure 11 shows the NaCl concentration dependence of the osmotic second virial coefficients of the BPTI from the experimental data and our calculations. The error bars of the experimental data are from [5].

It is well known that the electrostatic contribution depends on the choice of the molecular surface [42], in our model
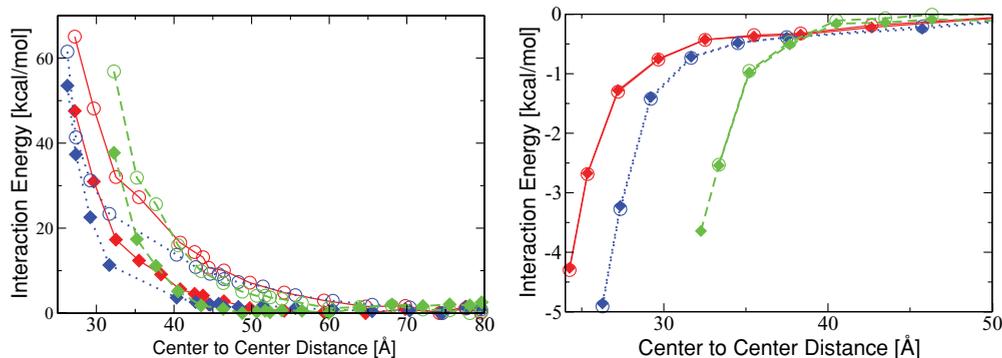
FIG. 10. (Color online) The electrostatic interaction energies (left) and the van der Waals interaction energies (right) between two BPTI molecules at various solution conditions. Each pair configuration is represented by a solid line, dotted line, and dashed line for AB, AC, AD configuration, respectively (the lines are only to guide the eye). Using the same code, two curves for each pair configuration are shown: the open circle indicates the interaction energies for 2% NaCl solution and the filled diamond indicates the energy for 7% NaCl solution. Because of the three-dimensional structure of the BPTI protein, the starting distance of the single-tree FMM calculation for each pair interaction is different as the contact distance varies.

there is a coupling between the hard core contribution to B_2 and the electrostatic contribution as both of them are related to the choice of the molecular surface. As for the van der Waals contribution, our model's attraction strength at contact is very similar to the estimate from other ones [43], thus, we will treat the electrostatic contribution with a scaling factor which is determined by matching the calculated B_2 with the experimental one at one solution condition (0.75M NaCl in this case, other matching solution conditions yield similar correlations). Besides this scaling factor, there is no other adjustable parameter in our calculations. For the BPTI case, the hard core contribution is about 38 500Å^3 and thus there is a substantial contribution to overall B_2 from the soft interactions.

The variations of the calculated B_2 from observed values are relatively large at high concentrations of NaCl solution. This is because the calculated B_2 data above $1M$ of NaCl concentration are overestimated by our model. This is an
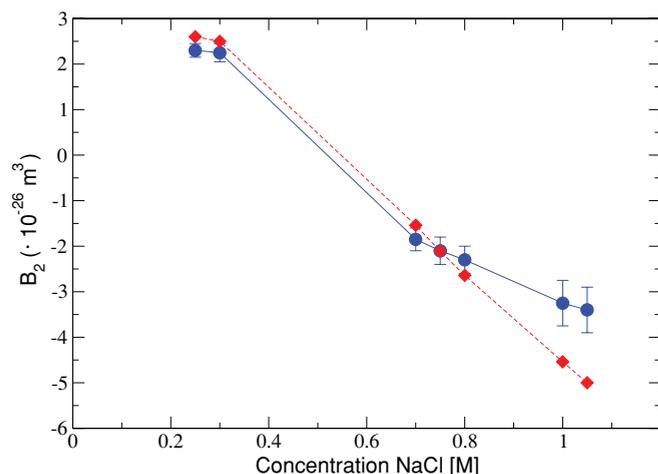

FIG. 11. (Color online) The NaCl concentration dependence of the osmotic second virial coefficients of BPTI. The solid line with circles is the experimental B_2 from Farnum and Zukoski and the dashed line with diamonds is our calculated result. The error bars for the experimental data are taken from [5].

indication of the limitation of our model as the Debye-Hückel theory will break down at high salt concentrations.

The second virial coefficients of lysozyme are also calculated in a similar manner. When compared with experimental data, the second virial coefficient is scaled as $B_2(ml \cdot mol/g^2) = B_2(m^3)N_A/M_w^2$ which is used in reporting the experimental data [6], where N_A is the Avogadro constant and M_w is the molar mass of the protein. Again averaged B_2 is calculated by using Eq. (6) with six different pair configurations based on the crystal space group operations of $P2_12_12_1$. Figure 12 shows the comparisons between the experimental data and calculated results from various solution conditions.

In Fig. 12(a), the experimental and the calculated B_2 are given as a function of the concentration of the NaCl solution and other conditions remain constant at pH = 4.2 and 25°C. In general the correlation between the experimental and calculated results are good, but we also can see the limitation of our model for high concentrations of electrolyte solutions, at $7\%(w/v)$ of NaCl solution just as the same behavior of BPTI.

The B_2 as a function of the pH of solution in NaCl solution in Fig. 12(b) shows a reasonable agreement between the experimental and calculated data even though experiments show a slight increase at pH = 5.2. The experiments and calculations are performed at 25°C and 2.0% NaCl concentration. The temperature dependence of B_2 clearly shows that the calculated result has good correlation with the experimental data. This dependence also has an exception point for the low temperature 5°C. According to the correlation between observed B_2 values and the solubilities of the lysozyme in solutions [44], the solubility of lysozyme shows clearly decrease as the calculated B_2 decreases with temperature as the other solution conditions remain constant at pH = 4.2 and the concentration of NaCl being 2.0%.

The temperature of a solution affects the second virial coefficients of protein solutions either via the inverse Debye screening length κ or the integrand in Eq. (4). Furthermore, temperature effect is represented by the change of the dielectric constant of water which enters our calculations via the Debye
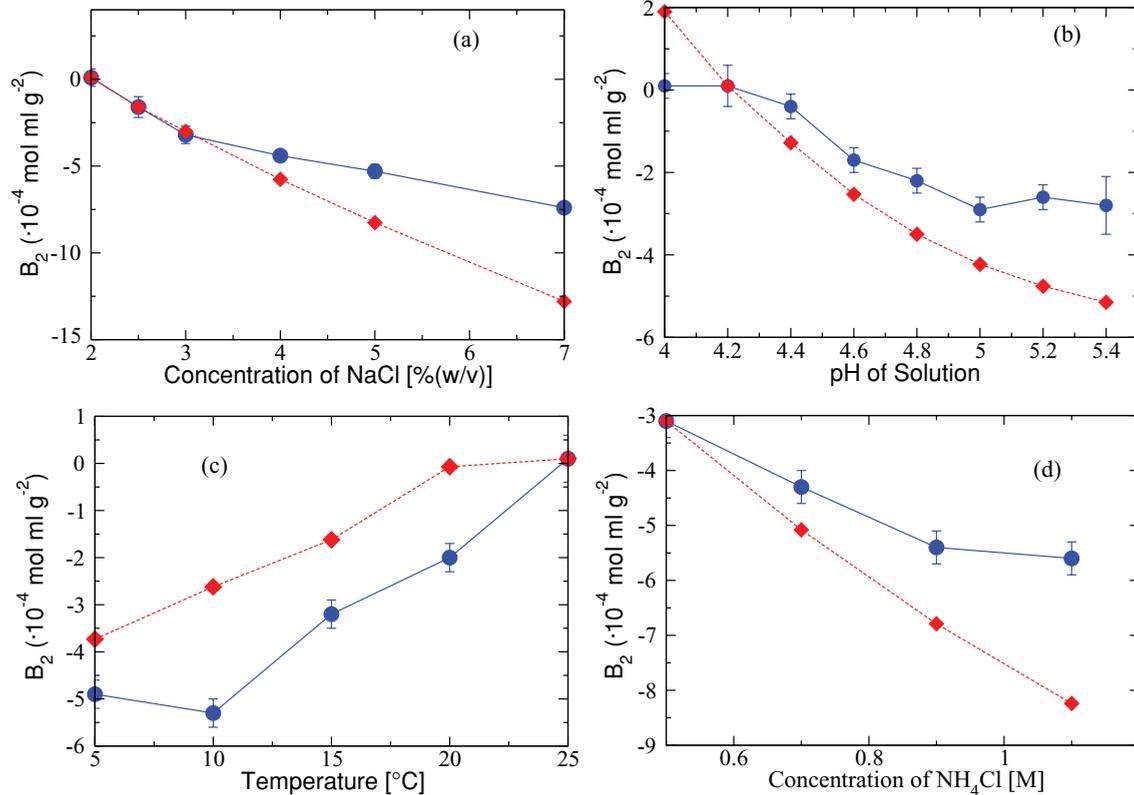
FIG. 12. (Color online) Comparisons between the experimental B_2 [6] and the calculated B_2 of lysozyme at various solution conditions. The dependence of B_2 on NaCl concentration is shown in (a). The pH dependence is in (b). The temperature dependence is in (c). The dependence upon ammonium chloride concentration is shown in (d). The solid lines with circles indicate the experimental data and the dashed lines with diamonds indicate our calculated results. For the first three panels a single solution condition (2% NaCl, pH = 4.2 and temperature is 25°C) is used to determine the scaling parameter for the electrostatic contribution. For the (d), the solution condition (0.5M ammonium chloride, pH = 4.5 and temperature 25°C) is used to determine the scaling parameter, which is essential the same as the NaCl solutions since our model cannot differentiate the nature of the salt except the ionic strength. The experimental data error bars from the literature [6] are also shown.

screening length and direct dielectric screening. From 25°C to 0°C, the dielectric constant increases from 80 to 88 [45] and according to Harvey and Lemmon this increase also gives a decreasing effect on the second virial coefficients under low temperatures, $T < 350$ K [46]. The predicted B_2 from our calculations shows the correct correlation with observed data [6] of lysozyme solutions. But the observed second virial coefficient shows unusual effect at the temperature 5°C.

From the structural study of the lysozyme crystal, the unusual effect of temperature was seen at the 280 K structure [47]. The number of water molecules under 4Å, the cutoff distance between the lysozyme surface, and the water molecules in the 280K structure, are smaller than in either the higher temperature($T > 295K$) or the lower temperature ($T < 250$) K structures. The lower number of waters may cause the smaller interactions between water molecules and atoms on the protein surface. This could be a possible reason that the second virial coefficient at 5°C is observed to have an abnormal behavior considering the overall trend with the temperature changes.

Finally, in Fig. 12(d), the experimental and calculated B_2 are given as a function of the concentration of the ammonium chloride solution. We also can see the limitation of this model for the high concentration above 1M of NH$_4$Cl solution, which will be further discussed in the next section.

IV. LIMITATION OF THE MODEL: BEYOND DEBYE-HÜCKEL THEORY

In Figs. 11, 12(a) and 12(d), the calculated B_2 at high concentrations of both sodium chloride and ammonium chloride are overestimated and the linear fit correlations to the experimental values deteriorate. According to our calculations this overestimation occurs at the high concentration of an ionic solution whose ionic strength is greater than $1M$ and the inverse Debye-Hückel screening length κ is large (>0.1). At such high concentrations, the Debye-Hückel theory fails, which affects our electrostatic and the van der Waals calculations.

This limitation leads to qualitative wrong correlations for divalent ion solutions such as magnesium bromide. Figure 13 shows the failure of our model which is based on the Debye-Hückel theory. The observed second virial coefficients of lysozyme show a minimum at the concentration of MgBr$_2 \sim$ $0.3M$, and start increasing as the ionic strength increases. Both experimental results from the Static Light Scattering (SLS) [6] and the Self-Interaction Chromatography(SIC) [9] show the same behavior. The calculations predict decrease of B_2 as the concentration increases and agree with the experimental data only up to the minimum point from the experiments. But at
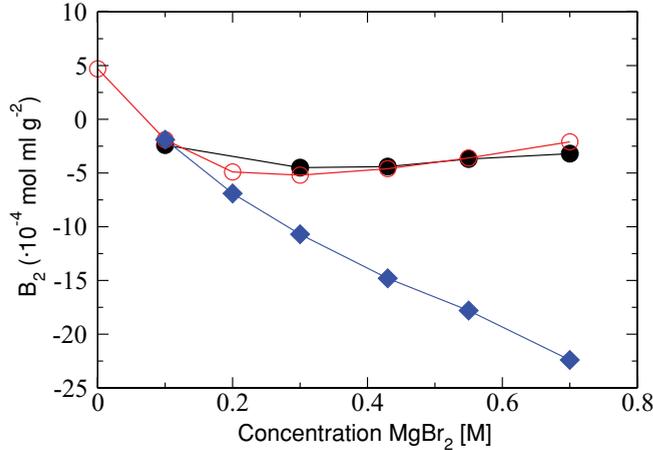
FIG. 13. (Color online) The $MgBr_2$ concentration dependence of the osmotic second virial coefficients of lysozyme solution at pH 7.8 is shown above. The solid line with filled circles are measured by the Static Light Scattering (SLS) [6], and the solid line with open circles are from the Self-Interaction Chromatography (SIC) [9]. The solid lines with diamonds are our calculations. Both observed results of B_2 become more positive at higher ionic strength. But the calculated results do not show the increase of the second virial coefficients at high ionic strength of magnesium bromide solutions. In this figure the scaling factor for the electrostatic contribution is determined at the following solution condition: 0.1M MgBr, pH = 7.8, and $25°C$.

high concentration of $MgBr_2$, the calculations only predict the second virial coefficients decrease to large negative values and at this point the inverse Debye-Hückel screening length κ is already greater than 0.1.

Recently, a molecular Debye-Hückel theory was developed [48,49] to address such a limitation of the traditional Debye-Hückel theory. The new theory is not only formulated for the static case, but also for the dynamical case. Therefore, using the new theory may improve the calculation of the electrostatic contribution to the interaction energy and at the same time can also improve the calculation of the van der Waals energies. The frequency dependent dielectric function is already applied to the dynamical Poisson-Boltzmann equation in Eqs. (30), (31), (32), and (33) for the van der Waals interaction. It will be interesting to see how the results from the new theory correlate with the experimental ones.

At the molecular level, the binding affinity of Mg^{2+} ions to the surface of lysozyme increases as the concentration of $MgCl_2$ increases [50,51]. The extent of Mg^{2+} ion binding increases as the pH of the solution increases to the isoelectric point of the protein (for lysozyme, 9.2) because the net positive charge on the protein surface approaches zero at this point. The open active site residues of lysozyme are glutamic acid (E53) and aspartic acid (D70) and both are negatively charged at this pH condition and the overall net charge of lysozyme decreases from 13.3 at pH = 4.0 to 7.65 at pH = 7.8 under $23°C$ which is the condition used in the experiments and our calculations. Due to the binding of Mg^{2+} divalent cations to the acidic residues of lysozyme, the repulsive interactions between lysozyme molecules increase, hence, cause more positive second virial coefficients observed in both SLS and SIC experiments.

V. CONCLUDING REMARKS

The extended Fast Multipole Method for two bodies are implemented to solve the system of linear equations from the linearized Poisson-Boltzmann equation to calculate the effective interaction energy of both electrostatic and van der Waals contributions. The traditional Boundary Element Method [32] implementation following Juffer *et al.* [31] requires the computational cost both in term of memory and time with the order of $O[(2N)^2]$ if the number of surface elements is N. This computational cost is the bottleneck for comprehensive studies on the interactions between large proteins. The extended FMM algorithm circumvents this computational bottleneck to reduce the cost to order of $O(N)$ for the double-tree FMM with additional outer iteration method and the order of $O(2N)$ for the single-tree FMM. The double-tree FMM is suitable at the relatively large distance and the single-tree method is good at shorter distance, where the transition point is roughly twice the size of protein molecule. The accuracy and performance of both methods can be controlled by adjusting the depth of trees, the number of expansion terms and the tolerance factor of iteration [52].

The osmotic second virial coefficients B_2 calculations of bovine pancreatic trypsin inhibitor and lysozyme solutions are used to validate our protein-protein interaction model. To reduce the computational cost the orientational dependence of the interaction energy in the integral of Eq. (6) is simplified by using the pair configurations from the crystal structure, which is a reasonable way to sample the most probable configurations in orientational space. The calculated B_2 generally agrees well with observed values from various solution conditions such as salt concentrations, pH, and temperature.

The model breaks down at high concentrations of monovalent salts and moderate concentration of multivalent salts such as Mg^{2+}. Our results show the overestimation of B_2 when the ionic strength is greater than $0.1M$ in general and do not show the repulsive effect of the magnesium ion upon binding to the negatively charged amino acid residues, which causes the positive increase of B_2 even if the ionic salt concentration increases. This clearly indicates the limitation of the Debye-Hückel theory used in our model. Possible improvements using the newly developed molecular Debye-Hückel theory [48,49] are under way.

Overall, the calculated B_2 are well correlated with the experimental observations for various solution conditions. In combination with our previous work on the binding affinity calculations [29] it is reasonable to expect that our residue level model can be used as a reliable model to describe protein-protein interactions in solutions. Naturally there are several immediate ways to improve the model, such as bettering the nuclear polarizability model of amino acids and improving treatment of the electrolyte solution modeling beyond Debye-Hückel theory. Given the simplicity of the model, the overall agreements between our calculations and experimental measurements are worth exploring so that a reliable model of protein-protein interactions in electrolyte solutions can be developed.

Since our approach needs the approximate structure of a protein at the residue level as initial input we will briefly discuss possible ways to obtain this information.

Experimentally there are other ways to provide partial structural information, which can also be used as the starting point of our model. Even though a reliable and accurate structure prediction from sequence is not yet available, approximate structures (resolution 6 to 8 Å, which corresponds to the residue level resolution) from such predictions [53] could offer a reasonable starting point for our approach, naturally an iterative process in collaboration with crystallographers is essential. For example, using the initial approximate structure a comparison of the second virial coefficient between the model calculation as shown in the current contribution and the light scattering experiments will lead to some insights into the geometric shape of the approximation structure and the result over all interactions between protein molecules. Thus, a combination of our strategy and the structure prediction from primary sequence may be exploited for the search of optimal crystallization condition. The predicted crystallization conditions can then be used to guide experimental design for the search of optimal conditions.

APPENDIX: ELECTROSTATIC INTERACTION FREE ENERGY BETWEEN TWO CHARGED SPHERICAL PARTICLES

In order to validate our boundary element solvers either based on the direct solver or the fast multipole method, we derived the analytic solution for a simple model, two identically charged spheres in an electrolyte solution. We follow the approach described in [54] for linearized Poisson-Boltzmann equations by adding a charge at the center of each sphere. In the linearized Poisson-Boltzmann model, the electrostatic potential ψ outside the spheres and φ_i inside the sphere i satisfy the following equations:

$$\nabla^2 \psi = \kappa^2 \psi \qquad \text{outside the spheres,}$$
$$\nabla^2 \varphi_i = -\frac{q_i \delta(r - r_i)}{\varepsilon_1} \quad \text{inside the sphere i = 1 or 2,} \quad \text{(A1)}$$
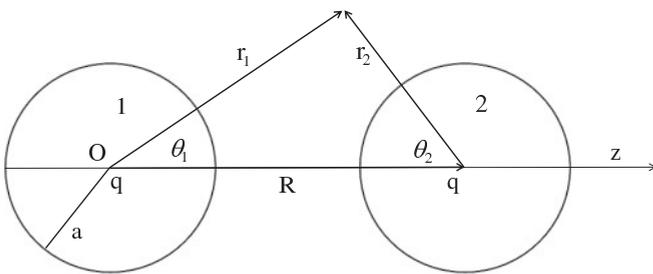

FIG. 14. Schematic diagram of the coordinate system of two sphere problem. a is the radius of sphere, R is the center-to-center distance, and r_1, θ_1, r_2 and θ_2 are the coordinate system from spheres 1 and 2, respectively [54]. The charge q is located at the center of each sphere.

where κ is the inverse Debye screening length of the electrolyte solution and q_i is the charge located at the center of each sphere i and ε_1 is the dielectric constant inside the sphere. The solution of Eq. (A1) in an electrolyte solution (outside of the spheres) can be written as [55] (and the coordinate system of the two spheres are shown in Fig. 14)

$$\psi(r_1, \theta_1, R) = \sum_{n=0}^{\infty} a_n \left\{ k_n(\kappa r_1) P_n(\cos \theta_1) \right.$$
$$\left. + \sum_{m=0}^{\infty} (2m+1) B_{nm} i_m(\kappa r_1) P_n(\cos \theta_1) \right\}, \quad \text{(A2)}$$

where

$$B_{nm} = \sum_{\nu=0}^{\infty} A_{nm}^{\nu} k_{n+m-2\nu}(\kappa R) \quad \text{(A3)}$$

$$A_{nm}^{\nu} = \frac{\begin{array}{c}\Gamma(n - \nu + 1/2)\Gamma(m - \nu + 1/2)\Gamma(\nu + 1/2) \\ \times (n + m - \nu)!(n + m - 2\nu + 1/2)\end{array}}{\pi \Gamma(m + n - \nu + 3/2)(n - \nu)!(m - \nu)!\nu!}, \quad \text{(A4)}$$

$i_n(x)$ and $k_n(x)$ are the modified spherical Bessel functions of the first and third kind, respectively [56], $\Gamma(z)$ is the γ function. The solution of Eq. (A1) inside the spheres has the following general form:

$$\varphi_i(r_i, \theta_i) = \sum_{n=0}^{\infty} b_n r_i^{\ n} P_n(\cos \theta_i) + \frac{q_i}{r_i}. \quad \text{(A5)}$$

The unknown coefficients a_n and b_n can be determined by applying the boundary conditions of the potential on the surface of the sphere at $r_1 = a$,

$$\psi|_{r_1=a} = \varphi_1|_{r_1=a}$$
$$\varepsilon_2 \frac{\partial \psi}{\partial r}\bigg|_{r_1=a} = \varepsilon_1 \frac{\partial \varphi_1}{\partial r}\bigg|_{r_1=a}, \quad \text{(A6)}$$

where ε_2 is the dielectric constant of the solution, and $\varepsilon = \varepsilon_2/\varepsilon_1$ will be used for further derivation. Applying boundary conditions Eq. (A6) on Eqs. (A2) and (A5) the coefficients b_n and the potential function inside sphere 1 is (the subscript to denote spheres are dropped due to the symmetry of the problem as $\varphi_1 = \varphi_2$)

$$\varphi(r, \theta) = \sum_{n=0}^{\infty} \left[\left(\frac{r}{a}\right)^n a_n \left\{ k_n(\kappa a) \right. \right.$$
$$\left. + \sum_{m=0}^{\infty} (2m+1) B_{nm} i_m(\kappa a) \right\} P_n(\cos\theta) - \frac{q}{a}\left(\frac{r}{a}\right)^n \right]$$
$$+ \frac{q}{r}. \quad \text{(A7)}$$

In order to evaluate the electrostatic solvation energy at the charge position $r = 0$, $r \to 0$ limit means that only $n = 0$ term

survives,

$$\varphi(r=0) = a_0 \left\{ k_0(\kappa a) + \sum_{m=0}^{\infty} (2m+1) B_{0m} i_m(\kappa a) \right\} - \frac{q}{a}. \tag{A8}$$

To find another unknown coefficient a_0, we only need the $m=0$ term after applying the second boundary condition in Eq. (A6) using the $n=0$ term in the solvation energy calculation,

$$a_0 = -\frac{q}{a} \frac{1}{\varepsilon \kappa a} \frac{1}{k'_0(\kappa a) + B_{00} i'_0(\kappa a)}. \tag{A9}$$

So the potential at the charge position can be written as

$$\varphi(r=0) = -\frac{q}{a} \frac{1}{\varepsilon \kappa a} \frac{k_0(\kappa a) + B_{00} i_0(\kappa a)}{k'_0(\kappa a) + B_{00} i'_0(\kappa a)} - \frac{q}{a}, \tag{A10}$$

where $B_{00} = \sum_{\nu=0}^{\infty} A_{00}^{\nu} k_{-2\nu}(\kappa R) = k_0(\kappa R)$.

The exact analytic expression of the solvation energy of a single sphere with a charge at the center of the sphere is reproduced by taking the $R \to \infty$ limit and using

$$B_{00}(R \to \infty) = \lim_{R \to \infty} k_0(\kappa R) = \lim_{R \to \infty} \frac{\pi}{2} \frac{e^{-\kappa R}}{\kappa R} = 0, \tag{A11}$$

thus, the solvation energy W of a single sphere,

$$W(R \to \infty) = \frac{1}{2} q \varphi(r=0)$$

$$= \frac{1}{2} \left\{ -\frac{q^2}{a} \frac{1}{\varepsilon \kappa a} \frac{k_0(\kappa a)}{k'_0(\kappa a)} - \frac{q^2}{a} \right\}$$

$$= \frac{1}{2} \frac{q^2}{a} \frac{1 - (1 + \kappa a)\varepsilon}{(1 + \kappa a)\varepsilon}. \tag{A12}$$

To calculate the electrostatic interaction free energy of the two identical spheres, we need to subtract the interaction potential of the infinitely separated spheres from the potential between two spheres at a finite distance, that is, $\varphi_{12} = \varphi(R) - \varphi(R \to \infty)$,

$$\varphi_{12} = \varphi(R) - \varphi(R \to \infty)$$

$$= -\frac{q}{a} \frac{1}{\varepsilon \kappa a} \left\{ \frac{k_0(\kappa a) + k_0(\kappa R) i_0(\kappa a)}{k'_0(\kappa a) + k_0(\kappa R) i'_0(\kappa a)} - \frac{k_0(\kappa a)}{k'_0(\kappa a)} \right\}. \tag{A13}$$

This expression is used to validate our solution based on the fast multipole method.

[1] A. George, Y. Chiang, B. Guo, A. Arabshahi, Z. Cai, and W. W. Wilson, in *Methods in Enzymology*, edited by J. C. W. Carter (Academic Press, New York, 1997), Vol. 276, pp. 100–110.

[2] S. Veesler, S. Lafont, S. Marcq, J. Astier, and R. Boistelle, J. Cryst. Growth **168**, 124 (1996).

[3] R. Boistelle, S. Lafont, S. Veesler, and J. Astier, J. Cryst. Growth **173**, 132 (1997).

[4] M. Gabrielsen, L. A. Nagy, L. J. DeLucas, and R. J. Cogdell, Acta Crystallogr. Sect. D **66**, 44 (2010).

[5] M. Farnum and C. Zukoski, Biophys. J. **76**, 2716 (1999).

[6] B. Guo, S. Kao, H. McDonald, A. Asanov, L. L. Combs, and W. William Wilson, J. Cryst. Growth **196**, 424 (1999).

[7] F. Bonneté, N. Ferté, J. Astier, and S. Veesler, J. Phys. IV (France) **118**, 3 (2004).

[8] O. D. Velev, E. W. Kaler, and A. M. Lenhoff, Biophys. J. **75**, 2682 (1998).

[9] P. M. Tessier, A. M. Lenhoff, and S. I. Sandler, Biophys. J. **82**, 1620 (2002).

[10] V. L. Vilker, C. K. Colton, and K. A. Smith, J. Colloid Interface Sci. **79**, 548 (1981).

[11] R. J. Hunter, *Foundations of Colloid Science* (Oxford University Press, Oxford, 1987).

[12] W. H. Gallagher and C. K. Woodward, Biopolymers **28**, 2001 (1989).

[13] M. Muschol and F. Rosenberger, J. Chem. Phys. **103**, 10424 (1995).

[14] D. Kuehner, C. Heyer, C. Ramsch, U. Fornefeld, H. Blanch, and J. Prausnitz, Biophys. J. **73**, 3211 (1997).

[15] B. L. Neal, D. Asthagiri, and A. M. Lenhoff, Biophys. J. **75**, 2469 (1998).

[16] C. Roth, B. Neal, and A. Lenhoff, Biophys. J. **70**, 977 (1996).

[17] W. L. Jorgensen and J. Tirado Rives, J. Am. Chem. Soc. **110**, 1657 (1988).

[18] X. Song and X. Zhao, J. Chem. Phys. **120**, 2005 (2004).

[19] X. Song, Mol. Simul. **29**, 643 (2003).

[20] A. A. Zamyatnin, Annu. Rev. Biophys. Bioengineering **13**, 145 (1984).

[21] Sanner, [http://www.scripps.edu/sanner/html/msms_home.html].

[22] X. Song, J. Chem. Phys. **116**, 9359 (2002).

[23] S. Millefiori, A. Alparone, A. Millefiori, and A. Vanella, Biophysical Chemistry **132**, 139 (2008).

[24] F. Dong and H.-X. Zhou, Proteins: Struct., Funct., Bioinformatics **65**, 87 (2006).

[25] K. Brock, K. Talley, K. Coley, P. Kundrotas, and E. Alexov, Biophys. J. **93**, 3340 (2007).

[26] L. Greengard, *The rapid evaluation of potential fields in particle systems*, ACM distinguished dissertations (MIT Press, Cambridge, MA, 1988).

[27] L. Greengard and V. Rokhlin, J. Comput. Phys. **135**, 280 (1997).

[28] B. Lu, X. Cheng, and J. Andrew McCammon, J. Comput. Phys. **226**, 1348 (2007).

[29] B. Kim, J. Song, and X. Song, J. Chem. Phys. **133**, 095101 (2010).

[30] D. A. McQuarrie, *Statistical Mechanics* (Harper Collins, New York, 1976).

[31] A. J. Juffer, E. F. F. Botta, B. A. M. van Keulen, A. van der Ploeg, and H. J. C. Berendsen, J. Comput. Phys. **97**, 144 (1991).

[32] K. Atkinson and W. Han, *Numerical Solution of Fredholm Integral Equations of the Second Kind*, 3rd ed., Texts Applied in Mathematics (Springer, New York, 2009).

[33] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley and Sons, New York, 1999).

[34] J. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, New York, 1985).

[35] V. Parsegian, *Physical Chemistry: Enriching Topic from Colloid and Surface Science* (Theorex, La Jolla, CA, 1975).

[36] H. Wondratschek and U. Müller, *International Tables for Crystallography, Volume A: Space Group Symmetry*, 5th ed. (Springer, New York, 2002).

[37] J. P. Jasinski and B. M. Foxman, [http://people.brandeis.edu/~foxman1/teaching/indexpr.html] (2007).

[38] B. T. Miller, R. P. Singh, J. B. Klauda, M. Hodoscek, B. R. Brooks, and H. L. Woodcock, Journal of Chemical Information and Modeling **48**, 1920 (2008).

[39] C. B. Delphine, M. R. David, and H. J. Jan, Proteins: Struct., Funct., Bioinformatics **73**, 765 (2008).

[40] M. Davies, C. Toseland, D. Moss, and D. Flower, BMC Biochemistry **7**, 18 (2006).

[41] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, J. Comput. Chem. **25**, 1605 (2004).

[42] C. H. Tan, L. J. Yang, and R. Luo, J. Phys. Chem. B **110**, 18680 (2006).

[43] V. A. Parsegian, *Van der Waals Forces: A Handbook for Biologists, Chemists, Engineers, and Physicists* (Cambridge University Press, New York, 2006).

[44] C. Gripon, L. Legrand, I. Rosenman, O. Vidal, M. C. Robert, and F. Boue, J. Cryst. Growth **178**, 575 (1997).

[45] J. N. Murrell and A. D. Jenkins, *Properties of Liquids and Solutions*, 2nd ed. (John Wiley and Sons, Chichester, UK, 1994).

[46] A. H. Harvey and E. W. Lemmon, J. Phys. Chem. Ref. Data **33**, 369 (2004).

[47] I. V. Kurinov and R. W. Harrison, Acta Crystallogr. Sect. D **51**, 98 (1995).

[48] X. Song, J. Chem. Phys. **131**, 044503 (2009).

[49] T. Xiao and X. Song (manuscript in preparation).

[50] J. J. Grigsby, H. W. Blanch, and J. M. Prausnitz, J. Phys. Chem. B **104**, 3645 (2000).

[51] T. Arakawa, R. Bhat, and S. N. Timasheff, Biochemistry **29**, 1924 (1990).

[52] K.-i. Yoshida, Ph.D. thesis, Department of Global Environment Engineering, Kyoto University (2001).

[53] See, [http://predictioncenter.gc.ucdavis.edu].

[54] S. L. Carnie and D. Y. Chan, J. Colloid Interface Sci. **155**, 297 (1993).

[55] A. B. Glendinning and B. W. Russel, J. Colloid Interface Sci. **93**, 95 (1982).

[56] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th Dover printing, 10th GPO printing ed. (Dover, New York, 1964).