2010

Calculations of the binding affinities of protein-protein complexes with the fast multipole method

Bongkeun Kim
Iowa State University, bkkim@iastate.edu

Jiming Song
Iowa State University, jisong@iastate.edu

Xueyu Song
Iowa State University

Calculations of the binding affinities of protein-protein complexes with the fast multipole method

Abstract

In this paper, we used a coarse-grained model at the residue level to calculate the binding free energies of three protein-protein complexes. General formulations to calculate the electrostatic binding free energy and the van der Waals free energy are presented by solving linearized Poisson-Boltzmann equations using the boundary element method in combination with the fast multipole method. The residue level model with the fast multipole method allows us to efficiently investigate how the mutations on the active site of the protein-protein interface affect the changes in binding affinities of protein complexes. Good correlations between the calculated results and the experimental ones indicate that our model can capture the dominant contributions to the protein-protein interactions. At the same time, additional effects on protein binding due to atomic details are also discussed in the context of the limitations of such a coarse-grained model.

Keywords

CNDE, Chemistry, active site, binding affinities, coarse grained models, dominant contributions, electrostatic binding, fast multipole method, good correlations, Poisson-Boltzmann equations, protein binding, protein complexes, residue level, Boltzmann equation, boundary element method, Van der Waals forces, basillus amyloliquefaciens ribonuclease

Disciplines

Chemistry | Electrical and Computer Engineering

Comments

The following article is from *Journal of Chemical Physics* 133 (2010): 095101 and may be found at doi: 10.1063/1.3474624.

Calculations of the binding affinities of protein-protein complexes with the fast multipole method

Bongkeun Kim, Jiming Song, and Xueyu Song

Citation: The Journal of Chemical Physics **133**, 095101 (2010); doi: 10.1063/1.3474624
View online: http://dx.doi.org/10.1063/1.3474624
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/133/9?ver=pdfcov
Published by the AIP Publishing

Calculations of the binding affinities of protein-protein complexes with the fast multipole method

Bongkeun Kim,[1] Jiming Song,[2] and Xueyu Song[1,a)]
[1]*Department of Chemistry, Iowa State University, Ames, Iowa 50011, USA*
[2]*Department of Electrical and Computer Engineering, Iowa State University, Ames, Iowa 50011, USA*

(Received 15 May 2010; accepted 12 July 2010; published online 2 September 2010)

In this paper, we used a coarse-grained model at the residue level to calculate the binding free energies of three protein-protein complexes. General formulations to calculate the electrostatic binding free energy and the van der Waals free energy are presented by solving linearized Poisson–Boltzmann equations using the boundary element method in combination with the fast multipole method. The residue level model with the fast multipole method allows us to efficiently investigate how the mutations on the active site of the protein-protein interface affect the changes in binding affinities of protein complexes. Good correlations between the calculated results and the experimental ones indicate that our model can capture the dominant contributions to the protein-protein interactions. At the same time, additional effects on protein binding due to atomic details are also discussed in the context of the limitations of such a coarse-grained model.
 [doi:10.1063/1.3474624]

I. INTRODUCTION

The atomic resolved structure of proteins from x-ray crystallography relies on the production of diffraction quality crystals. Recent extensive studies from structural genomic project clearly indicate that even though purified proteins can be successfully obtained, only about 16% of them are crystallized with suitable quality for diffraction.[1] Thus the crystallization process is still the bottleneck for protein structure determination using crystallography. Currently protein crystallization conditions are screened from traditional trial-and-error procedure. Namely, the optimal condition is obtained from extensive searches of a large parameter space of protein solutions such as pH, buffer, temperature, salt concentration, and precipitating agent. Even with remarkable progress of automation the poor successful rate of obtaining protein single crystals is in no small part due to the lack of fundamental understanding of the crystallization process.[2] For example, recent studies using time-controlled microfluidic seeding heavily rely on knowledge of solution conditions during the nucleation stage and crystal growth stage.[3] In other recent high-throughput experimental studies using microfluidics, it is clearly demonstrated that knowledge of the phase behavior of a protein allows one to create a rational screening that increases the success rate of crystallization of challenging proteins.[4] It is therefore useful to understand what kind of solution conditions might lead toward the optimal crystallization conditions and why.

As a first step toward a reliable and practical theory of protein crystallization, a realistic model of protein-protein interactions needs to be developed. In general, there are two types of protein-protein interactions in nature. One type of interaction is responsible for the protein-protein recognition to perform specific biological functions. In this case there are complimentary regions on both proteins to recognize each other and hydrophobicity is the dominating factor.[5] On the other hand, the protein-protein interaction in protein crystallization does not necessarily involve complimentary regions to establish protein-protein contacts. For example, we have analyzed protein contacts for five lysozyme protein crystals from the protein data bank under five different crystallization conditions.[6] What we found is that protein contacts can be formed from different parts of lysozyme surface residues depending on the solution conditions. Similar conclusions can be drawn from other studies as well. For example, Crosio *et al.*[7] found that pancreatic ribonuclease uses nearly the entire protein surface residues to establish crystal-packing contacts under various crystallization conditions. An extensive analysis on 78 protein crystals indicates that the amino acid composition involved in the protein contacts is indistinguishable from that of the protein surface accessible to the solvent.[8] These studies also suggest that crystal-packing contacts formed are sensitive to the solution conditions in contrast with the type of protein-protein interaction where hydrophobic residues are favored.[9,10] Therefore a universal model to capture the effective interaction between protein molecules in solutions can be developed based on the Derjaguin–Landau–Verwey–Overbeek (DLVO) picture given the protein-protein interaction at short range can be accounted for appropriately since the DLVO picture will fail when the protein molecules are separated by several solvent molecular layers.

The key feature of such a model is that it should be based on the generic properties of 20 amino acids in nature and experimentally accessible properties of electrolyte solutions and crystallization agents and is therefore portable to all of the protein-protein interactions in aqueous solutions. Our recent studies[11,12] are such efforts toward this goal.

In this model, each residue of a protein is represented by

a)Electronic mail: xsong@iastate.edu.

a sphere located at the geometric center of the residue determined by its native or approximate structure. The diameter of the sphere is determined by the molecular volume of the residue in solution environment.[13] 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.[14] 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 a protein's native structure. If a residue is charged, the amount of charge can be obtained from the Henderson–Hasselbalch equation using the generic pK_a values of residues, thus the local environmental effects on pK_a values are neglected. In the calculations presented in this work, pK_a of the residues are calculated from PROPKA 2.0 (Ref. 15) to capture some of the local environmental effects. 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[16] and the electronic polarizability is estimated from optical dielectric constants augmented with quantum chemistry calculations.[17] There are three kinds of interactions in this model: the electrostatic interaction due to 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 report, we consider the electrostatic interactions which give the most contribution to the protein-protein interaction,[18,19] and the van der Waals interactions, which are the major contributors to the binding affinity calculations.

The electrostatic problem in the electrostatic interaction and the van der Waals interaction is solved using the Poisson–Boltzmann equation where realistic shapes of protein molecules are considered. The boundary element method (BEM) in combination with the fast multipole method[20,21] is implemented to circumvent the extensive memory requirements similar to a recent work.[22] In order to test the validity of our model, the binding affinities of several protein-protein complexes are calculated using our model and direct comparisons are made against experimental measurements. Reasonable agreements from these comparisons provide the first concrete evidence that our model can be used as a universal model for studies of nonspecific protein-protein interactions in aqueous solutions.

II. THEORETICAL DEVELOPMENTS

A. The statistical thermodynamics of binding affinities

To set up the computational framework for calculating the binding affinities, the statistical thermodynamic analysis of "double-decoupling" method from Gilson and his co-workers[23] is used. This approach is based on the change in the free energy of protein-protein binding when one protein and the other react and then a single complex is formed. The final result for the binding affinity is given as

$$\Delta G^{\circ} = \Delta G^{\circ}_{\mathrm{sol}}(\mathrm{AB}) - \Delta G^{\circ}_{\mathrm{sol}}(\mathrm{A}) - \Delta G^{\circ}_{\mathrm{sol}}(\mathrm{B}), \tag{1}$$

where $\Delta G^{\circ}_{\mathrm{sol}}$ is the free energy change when a molecule is introduced into a solution from vacuum. In this report, the binding free energy calculations are for single mutations at the binding site, thus free energy change due to translational, rotational, and vibrational contributions of the proteins upon binding remains relatively constant. In our model the free energy change is the sum of the electrostatic solvation energy and the van der Waals energy of the molecule. For protein A,

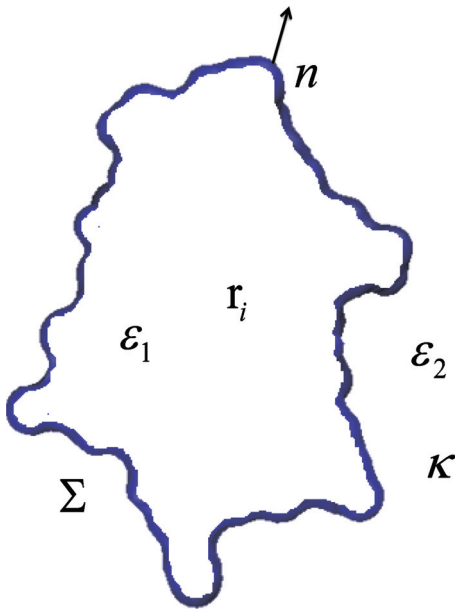

FIG. 1. Schematic illustration showing the electrostatic formulation of a protein. Σ is the molecular surface of a protein, $\mathbf{n}$ is outward unit normal, ε_1 and ε_2 are dielectric constants inside the cavity and outside solvent, respectively. κ is the inverse Debye screening length for the electrolyte solution. $\mathbf{r}_i$ is the geometric center of residue i, and charge q_i and dipole μ_i are located on residue i.

$$\Delta G^{\circ}_{\mathrm{sol}}(\mathrm{A}) = \Delta G^{\mathrm{elec}}_{\mathrm{sol}}(\mathrm{A}) + \Delta G^{\mathrm{vdw}}_{\mathrm{sol}}(\mathrm{A}). \tag{2}$$

B. The electrostatic solvation energy calculation

The electrostatic binding free energy between proteins A and B is defined as

$$\Delta G^{\mathrm{binding}}_{\mathrm{elec}} = \Delta G_{\mathrm{elec}}(\mathrm{AB}) - \Delta G_{\mathrm{elec}}(\mathrm{A}) - \Delta G_{\mathrm{elec}}(\mathrm{B}). \tag{3}$$

The electrostatic interaction is estimated from the Poisson–Boltzmann (PB) equation. To solve the PB equation, we use the boundary element method based on the integral equation formulation of the linearized PB equation for a single protein.[24,25]

Consider the molecular surface Σ which covers a protein molecule. There are N charges q_i and dipoles $\vec{\mu}_i$ at position $\mathbf{r_i}$ enclosed by the surface Σ. Inside this dielectric cavity the dielectric constant is ε_1 and the dielectric constant of the solution is ε_2 (see Fig. 1). The inverse Debye screening length κ is given by the solution's ionic strength. The integral equations for the potential $\varphi(\mathbf{r})$ and its gradient $\partial\varphi(\mathbf{r})/\partial n$ are given by the following integral equations:[11,25]

$$\frac{1}{2}\left(1+\frac{\varepsilon_2}{\varepsilon_1}\right)\varphi(\mathbf{r}_0)+\int\int_{\sum} L_1(\mathbf{r},\mathbf{r}_0)\varphi(\mathbf{r})d\mathbf{r} + \int\int_{\sum} L_2(\mathbf{r},\mathbf{r}_0)\frac{\partial\varphi(\mathbf{r})}{\partial n}d\mathbf{r} = \sum_{i=1}^{N}\{q_i F(\mathbf{r_i},\mathbf{r}_0)+\vec{\mu}_i\cdot\nabla F(\mathbf{r_i},\mathbf{r}_0)\}/\varepsilon_1, \quad (4)$$

$$\frac{1}{2}\left(1+\frac{\varepsilon_1}{\varepsilon_2}\right)\frac{\partial\varphi(\mathbf{r}_0)}{\partial n}+\int\int_{\sum} L_3(\mathbf{r},\mathbf{r}_0)\varphi(\mathbf{r})d\mathbf{r} + \int\int_{\sum} L_4(\mathbf{r},\mathbf{r}_0)\frac{\partial\varphi(\mathbf{r})}{\partial n}d\mathbf{r} = \sum_{i=1}^{N}\left\{q_i\frac{\partial F}{\partial n_0}(\mathbf{r_i},\mathbf{r}_0)+\vec{\mu}_i\cdot\nabla\frac{\partial F}{\partial n_0}(\mathbf{r_i},\mathbf{r}_0)\right\}/\varepsilon_1, \quad (5)$$

where

$$L_1(\mathbf{r},\mathbf{r}_0)=\frac{\partial F}{\partial n}(\mathbf{r},\mathbf{r}_0)-\frac{\varepsilon_2}{\varepsilon_1}\frac{\partial P}{\partial n}(\mathbf{r},\mathbf{r}_0), \quad (6)$$

$$L_2(\mathbf{r},\mathbf{r}_0)=P(\mathbf{r},\mathbf{r}_0)-F(\mathbf{r},\mathbf{r}_0), \quad (7)$$

$$L_3(\mathbf{r},\mathbf{r}_0)=\frac{\partial^2 F}{\partial n_0\,\partial n}(\mathbf{r},\mathbf{r}_0)-\frac{\partial^2 P}{\partial n_0\,\partial n}(\mathbf{r},\mathbf{r}_0), \quad (8)$$

$$L_4(\mathbf{r},\mathbf{r}_0)=-\frac{\partial F}{\partial n_0}(\mathbf{r},\mathbf{r}_0)+\frac{\partial P}{\partial n_0}(\mathbf{r},\mathbf{r}_0)\frac{\varepsilon_1}{\varepsilon_2}, \quad (9)$$

and

$$F(\mathbf{r},\mathbf{r}_0)=\frac{1}{4\pi|\mathbf{r}-\mathbf{r}_0|}, \quad (10)$$

$$P(\mathbf{r},\mathbf{r}_0)=\frac{e^{-\kappa|r-\mathbf{r}_0|}}{4\pi|\mathbf{r}-\mathbf{r}_0|}. \quad (11)$$

Although the traditional boundary element method such as Atkinson and his co-workers'[26] can be used to solve above integral equations, the memory requirement is too costly even on the newest computers using either a direct linear system solver or iterative solver, such as generalized minimal residual method (GMRES) (Ref. 27) for a moderate size protein. In the current work the fast multipole method is implemented and the details will be outlined in the Appendix. Once the above integral equations are solved the potential inside the dielectric cavity is

$$\varphi(\mathbf{r}_0)=-\int\int_{\sum} L_1(\mathbf{r},\mathbf{r}_0)\varphi(\mathbf{r})d\mathbf{r} - \int\int_{\sum} L_2(\mathbf{r},\mathbf{r}_0)\frac{\partial\varphi(\mathbf{r})}{\partial n}d\mathbf{r}, \quad (12)$$

$$\nabla_0\varphi(\mathbf{r}_0)=-\int\int_{\sum} \nabla_0 L_1(\mathbf{r},\mathbf{r}_0)\varphi(\mathbf{r})d\mathbf{r} - \int\int_{\sum} \nabla_0 L_2(\mathbf{r},\mathbf{r}_0)\frac{\partial\varphi(\mathbf{r})}{\partial n}d\mathbf{r}. \quad (13)$$

Finally, the electrostatic solvation free energy is given by

$$\Delta G_{\text{ele}}=\sum_{i=1}^{N}\left\{\frac{q_i}{\epsilon_1}\varphi(\mathbf{r_i})+\frac{1}{\epsilon_1}\mu_i\cdot\nabla\varphi(\mathbf{r_i})\right\}. \quad (14)$$

C. The van der Waals energy contribution

The van der Waals binding free energy between proteins A and B is defined as in Eq. (15),

$$\Delta G_{\text{vdw}}^{\text{binding}}=\Delta G_{\text{vdw}}(\text{AB})-\Delta G_{\text{vdw}}(\text{A})-\Delta G_{\text{vdw}}(\text{B}). \quad (15)$$

Song and Zhao[12] had developed a theory to calculate the van der Waals interaction between protein molecules in an electrolyte solution using the following effective action in Fourier space of the polarizable dipoles:

$$S[\mathbf{m}_{\mathbf{r},n}]=-\frac{\beta}{2}\sum_{\mathbf{r}}\sum_{n=-\infty}^{n=\infty}\frac{1}{\alpha_{\mathbf{r},n}}\mathbf{m}_{\mathbf{r},n}\cdot\mathbf{m}_{\mathbf{r},-n} + \frac{\beta}{2}\sum_{\mathbf{r}\neq\mathbf{r}'}\sum_{n=-\infty}^{n=\infty}\frac{1}{\alpha_{\mathbf{r},n}}\mathbf{m}_{\mathbf{r},n}\cdot T(\mathbf{r}-\mathbf{r}')\cdot\mathbf{m}_{\mathbf{r},-n} + \frac{\beta}{2}\sum_{\mathbf{r},\mathbf{r}'}\sum_{n=-\infty}^{n=\infty}\frac{1}{\alpha_{\mathbf{r},n}}\mathbf{m}_{\mathbf{r},n}\cdot R_n(\mathbf{r}-\mathbf{r}')\cdot\mathbf{m}_{\mathbf{r},-n}, \quad (16)$$

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 dipoles at $\mathbf{r}$ and $\mathbf{r}'$, where the retardation is neglected. $R_n(\mathbf{r}-\mathbf{r}')$ is the reaction field tensor at Matsubara frequency $\omega_n=2\pi n/\beta\hbar$, which captures the effect of surrounding dielectric medium. If the electrolyte solvent is treated by the Debye–Hückel theory, this reaction field tensor can be calculated by solving the PB equation with dielectric function $\varepsilon(i\omega_n)$ and the Debye screening length κ.[12] The quantum partition function from this effective action of the system is

$$Q(\text{A})=\prod_n\left[\frac{2\pi}{\beta\det[A_n(\text{A})]}\right]^{1/2}, \quad (17)$$

where A represents the protein A and A_n matrix element is given by

$$A_n(\mathbf{r},\mathbf{r}')=\frac{1}{\alpha_{\mathbf{r},n}}\delta_{\mathbf{r},\mathbf{r}'}-T(\mathbf{r}-\mathbf{r}')-R_n(\mathbf{r}-\mathbf{r}'). \quad (18)$$

The symbol "det" represents the determinant of a matrix. Finally the van der Waals binding free energy is given by[12]

$$\Delta G_{\text{vdw}}^{\text{binding}}=\frac{1}{2}k_B T\sum_{n=-\infty}^{n=\infty}[\ln\{\det[A_n(\text{AB})]\}-\ln\{\det[A_n(\text{A})]\} - \ln\{\det[A_n(\text{B})]\}]. \quad (19)$$

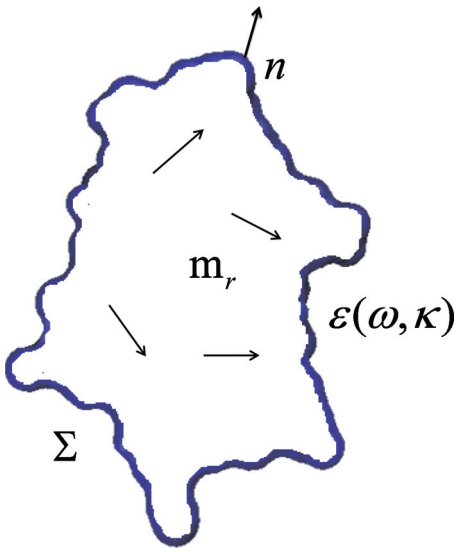

FIG. 2. Schematic illustration showing the van der Waals energy formulation of a single protein. Σ is the molecular surface of a protein, **n** is the outward unit normal, and ε is the dielectric constant outside solvent. $\mathbf{m_r}$ stands for the polarizable dipole located on the residue center.

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 with the properties of the proteins and of 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. Again consider the molecular surface Σ spanned by a protein molecule (Fig. 2). There are N polarizable dipoles $\mathbf{m_r}$ at position $\mathbf{r}$ enclosed by the surface Σ. Inside this dielectric cavity the dielectric constant is one and the dielectric function 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. 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 $\rho_{\text{eff}}(\mathbf{r})=-\mathbf{m}\nabla\delta(\mathbf{r}-\mathbf{r}_0)$,[28] the reaction field matrix involving residues $\mathbf{r}_i$ and $\mathbf{r}_j$ can be written as

$$R_n(\mathbf{r}_i,\mathbf{r}_j)=\int\int_{\sum}[\nabla_i F(\mathbf{r}_i,\mathbf{r}_j)-\nabla_i P(\mathbf{r}_i,\mathbf{r}_j)]\frac{\partial\varphi}{\partial n}(\mathbf{r}_j,\mathbf{r})d\mathbf{r}+\int\int_{\sum}\left[-\nabla_i\frac{\partial F}{\partial n_j}F(\mathbf{r}_i,\mathbf{r}_j)+\nabla_i\frac{\partial P}{\partial n_j}(\mathbf{r}_i,\mathbf{r}_j)\varepsilon\right]\varphi(\mathbf{r}_j,\mathbf{r})d\mathbf{r}, \quad (20)$$

where F and P are defined in Eqs. (10) and (11). φ and $\partial\varphi/\partial n$ can be obtained by solving the following integral equations at each frequency ω_n:[12,25]

$$\frac{1}{2}(1+\varepsilon(i\omega_n))\varphi(\mathbf{r}_i,\mathbf{r}_0)+\int\int_{\sum}L_1(\mathbf{r},\mathbf{r}_0)\varphi(\mathbf{r}_i,\mathbf{r})d\mathbf{r}+\int\int_{\sum}L_2(\mathbf{r},\mathbf{r}_0)\frac{\partial\varphi}{\partial n}(\mathbf{r}_i,\mathbf{r})d\mathbf{r}=\nabla_i F(\mathbf{r}_i,\mathbf{r}_0), \quad (21)$$

TABLE I. Intrinsic nuclear polalizability (α_{nu}) (Ref. 16), electronic polarizability (α_{el}), and ionization frequency of amino acids in unit of Å^3 from Millefiori *et al.* (Ref. 17).

Amino acid	α_{nu}	α_{el}	ω_I
Ala	2.09	8.25	75 650
Arg	4.38	18.01	63 880
Asn	5.05	11.66	71 740
Asp	3.08	10.86	75 250
Cys	2.18	11.40	70 900
Gln	4.40	13.54	70 450
Glu	3.10	12.79	73 400
Gly	2.01	6.44	77 110
His	2.53	15.14	65 730
Ile	1.98	13.67	73 710
Leu	1.90	13.80	74 320
Lys	2.96	15.39	67 510
Met	2.07	15.33	66 380
Phe	2.01	18.33	67 550
Pro	1.47	11.07	71 340
Ser	3.52	8.94	74 890
Thr	4.19	10.72	73 640
Trp	3.06	23.35	58 430
Tyr	3.50	19.25	63 070
Val	1.99	11.82	74 160

$$\frac{1}{2}\left(1+\frac{1}{\varepsilon(i\omega_n)}\right)\frac{\partial\varphi}{\partial n}(\mathbf{r}_i,\mathbf{r}_0)+\int\int_{\sum}L_3(\mathbf{r},\mathbf{r}_0)\varphi(\mathbf{r}_i,\mathbf{r})d\mathbf{r}+\int\int_{\sum}L_4(\mathbf{r},\mathbf{r}_0)\frac{\partial\varphi}{\partial n}(\mathbf{r}_i,\mathbf{r})d\mathbf{r}=\nabla_i\frac{\partial F}{\partial n_0}(\mathbf{r}_i,\mathbf{r}_0), \quad (22)$$

where L_1, L_2, L_3, and L_4 are defined in the electrostatic free energy calculation in Eqs. (6)–(9). To evaluate the van der Waals binding free energy in Eq. (19), the reaction field matrix is built using the dielectric function $\varepsilon(i\omega_n)$ for each frequency ω_n. The polarizability of a residue in a protein is given by

$$\alpha_n=\alpha(i\omega_n)=\frac{\alpha_{\text{nu}}}{1+\omega_n/\omega_{\text{rot}}}+\frac{\alpha_{\text{el}}}{1+(\omega_n/\omega_I)^2}, \quad (23)$$

where α_{nu} is the static nuclear polarizability of a residue[16] 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. $\omega_{\text{rot}}=20$ cm^{-1} is used for this calculation which is typical rotational frequency of molecules.[29] Other properties are listed in Table I based on the calculated results from Millefiori *et al.*[17] An accurate parametrization of the dielectric function $\varepsilon(i\omega)$ of water based on the experimental data is taken from Parsegian's work.[30]

In order to present a simple argument to show that our model indeed captures the dispersion interaction, a Drude model in one dimension can be used to illustrate the basic physics of the dispersion interaction. In the one-dimensional Drude model[31] an atom with a fluctuating dipole moment is described by a linear harmonic oscillator in which positive

and negative charges are joined by a spring with force constant k. We consider two identical neutral atoms with a separation, R, which is much larger than the separations between charges, $|x_1|$ and $|x_2|$, on the two atoms. Under these conditions, the lowest-order Coulomb interaction between the two atoms (oscillators) becomes a dipole-dipole interaction. Thus, the total Hamiltonian of the system is written as

$$H_e = \frac{m\dot{x}_1^2}{2} + \frac{kx_1^2}{2} + \frac{m\dot{x}_2^2}{2} + \frac{kx_2^2}{2} - \frac{2e^2x_1x_2}{R^3}.$$

If the coupled oscillators are decoupled using a normal-mode transformation, the zero-point energy of the system is reduced from that of the sum of the two uncoupled oscillators, $\hbar\omega_0$, where $\omega_0 = k/m$, by

$$\Delta U = -\frac{1}{2}\hbar\omega_0\frac{e^4}{k^2R^6} = -\frac{1}{2}\hbar\omega_0\frac{\alpha^2}{R^6}.$$

This energy difference, which is inversely proportional to the sixth power of the separation of the two oscillators, is the dispersion interaction energy. The second equality is obtained by recognizing that the polarizability α is e^2/k for a Drude model. Our model is the generalization of the above simple model to account for the thermal effect and the nuclear polarizability contributions in addition to the intervening dielectric medium.

D. Implementation of the fast multipole method to the boundary element method

The major drawback of the traditional BEM is the order $O(N^2)$ dependence of the matrix size on the number of surface elements N. The large size of a matrix not only requires larger usage of memory but also takes longer time to solve the corresponding linear system. An efficient algorithm developed by Greengard and Rokhlin,[21] the fast multipole method (FMM), is implemented to avoid storing matrix elements and to speed up matrix-vector multiplications which is the most time consuming step in solving linear equations. To apply the FMM algorithm to the BEM, surface elements on a protein surface are distributed to different three-dimensional rectangular boxes at different levels based on a hierarchical oct-tree, and a divide-conquer strategy is applied to the far field interactions at each level in the tree structure (see Fig. 7 in the Appendix).

The fundamental observation in FMM is that the multipole moment expansion of the far field interaction, which is roughly $O(N^2)$ in the direct BEM, can be approximated by low numbers of summation depending on the designated accuracy to lower computational cost. The integral elements of matrices in the electrostatic and the van der Waals interaction formulations are described by two different interactions, Columbic interaction and Debye–Hückel (screened Columbic) interaction. The detail formulations of the fast multipole method is described in the Appendix.

E. Preparation of protein complex structures

Three protein complex systems, where extensive experimental data are available, are used to test our protein-protein interaction model. The bovine pancreatic trypsin inhibitor (BPTI)-trypsin system where the crucial P_1 residue had been mutated to various residues, and the binding constants and mutated protein complex structures have been extensively documented.[32] The other two systems are well studied barnase-barstar complex[33] and the *Streptomyces griseus* proteinase B (SGPB)-turkey ovomucoid third domain complex (OMTKY3).[34]

In our preparations for mutant protein complex structures without experimental PDB structures, the Swiss-PDB viewer (Ref. 35) is first used to make a single mutant on the binding site and select the best rotamer based on its lowest score according to the formula

$$\begin{aligned}\text{score} = &(4 \times \text{Nb clash with backbone N, C}\alpha\text{, and C atoms}) + (3 \times \text{Nb clash with backbone O atoms}) \\ &+ (2 \times \text{Nb clash with side-chain atoms}) - \text{NbH bonds} - 4 \times \text{NbSS bonds},\end{aligned}$$

where "Nb" is the abbreviation of "number." Then molecular dynamics (MD) simulations using CHARMM force field[36] are performed to determine the final mutant structure used in our calculations. After the mutation using the Swiss-PDB viewer, an energy minimization in vacuum is performed using CHARMM, then further minimization is performed with water molecules. Over 200 ps NPT simulations at 300 K are performed using CHARMM. Either the average structure over 100 ps or the last snapshot of the structure are used, but there is no difference for our calculations due to the coarse-grained nature of the model. In order to test the validity of the simulated mutant structures as compared to the experimental mutant structures, 10 P_1 mutants of BPTI-trypsin complexes based on the wild-type PDB (PDB ID=3BTK) are used to validate our procedure for the simulated mutant structures. For the BPTI-trypsin system, the crystal structures of complexes between bovine β-trypsin and ten P_1 variants of BPTI (Ref. 37) are known experimentally (PDB code: 3BTD, 3BTE, 3BTF, 3BTG, 3BTH, 3BTK, 3BTM, 3BTQ, 3BTT, and 3BTW). The RMSD studies between simulated structures and experimental structures are within 1.3 Å (the average value from all ten mutants). Figure 3 shows the correlation of calculated binding affinities between experimental PDB structures and simulated PDB structures. Thus, the simple mutant PDB structure from Swiss PDB viewer mutation followed by MD simulations can be used as the mutant

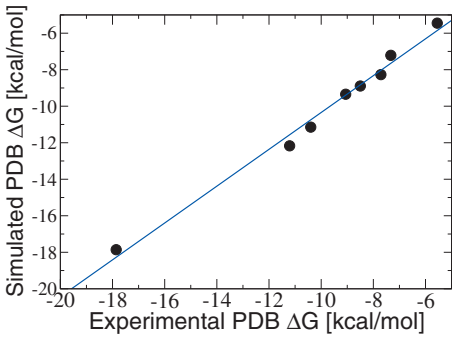

FIG. 3. The comparison of the electrostatic binding free energies between the experimental structures and MD simulated structures for the P_1 variants of BPTI-trypsin complexes. The linear fit correlation coefficient is 0.989.

structure to calculate the binding free energy of mutant complexes. This method is used to generate all mutant structures for barnase-barstar complexes and SGPB-OMTKY3 complexes for calculations.

For the barnase-barstar complex system, the crystal structure of the pseudo-wild-type barnase-barstar complex[38] (PDB code=1B27) is used as a template of the mutant complexes. This complex contains three sets of barnase-barstar complex, chain A is used for barnase and its binding site mutations and chain D is used for modeling the wild-type barstar. In order to make a wild-type protein, A40 and A82 in barstar are mutated to Cys. As the experimental results of barnase-barstar binding measurements indicate that the deletion of N-terminal Met residue in barstar, thus in our calculation the N-terminal Met is deleted in the template protein structure. So the final template has 110 barnase residues and 89 barstar residues. To make comparisons with experimental binding affinities,[33] seven mutant complexes (Ala, Cys, Phe, Gln, Ser, Trp, and Tyr) on the Glu73 residue in barnase are generated by the Swiss PDB viewer and followed by MD simulations.

Finally, the crystallographic structure of the SGPB and OMTKY3 complexes[39] (PDB code=3SGB) is used as the wild-type template for the mutant complexes. The following experimental PDB structures, PDB code 1CSO, 1CT0, 1CT2, and 1CT4 (Ref. 40) for P_1 Ile, Ser, Thr, and Val mutant complexes and PDB code 1SGP (Ref. 41) for P_1 Ala mutant complex and PDB code 2NU0, 2NU1, 2NU2, 2NU3, and 2SGF for P_1 Trp, His, Arg, Lys, and Phe mutant complexes already existed. However these structures were not used to calculate the binding affinities of the mutant complexes, instead the simulated structures from the wild-type template are used as RMSDs between experimental PDB structures and the simulated ones are all within 0.49 Å. At the same time, binding affinity calculations of BPTI-trypsin complexes already indicated the validity of using MD simulated mutant complex structures shown in Fig. 3. Since the wild-type template, PDB code 3SGB, has the first six residues in OMTKY3 inhibitor chain on the wild-type structure while the protein complexes used in the binding affinity measurements do not, thus, they are deleted from our simulated structures. The final templates for the SGPB and OMTKY3 complex contains 185 SGPB residues and 50 OMTKY3 residues.

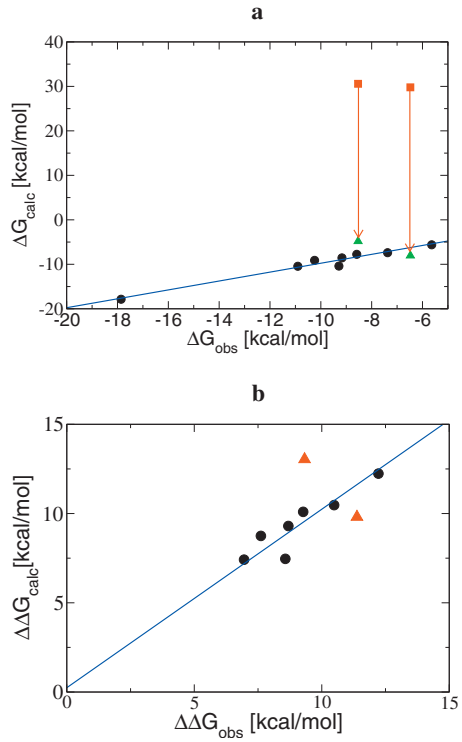

FIG. 4. The binding free energy changes of BPTI-trypsin complexes after applying the pK_a shifts of $P_1$1 Asp and P_1 Glu (a). The red square boxes represent the negative charged acidic P_1 variants and the green up-triangles are for the protonated P_1 after using PROPKA 2.0 (Ref. 15). The red arrows indicate the binding affinity shifts from positive to negative one. The Y-axis is the experimental binding free energy of 10 P_1 variants from Krowarsch *et al.* (Ref. 32). After considering the pK_a shifts for the acidic P_1 residues, the correlation of $\Delta\Delta G$ between observed and calculated data is shown in (b). The linear fit correlation coefficient from all mutants is 0.912. The linear fit excluding the mutations involving water mediated hydrogen bonds (two acidic P_1 mutants indicated by red triangles) yields 0.982.

III. RESULTS AND DISCUSSION

A. Binding energy calculations of BPTI-trypsin complexes

The binding free energies of BPTI-trypsin complexes are calculated according to our model. First, the binding free energy, ΔG, is calculated and the change of binding affinity from the mutation of P_1 residue, $\Delta\Delta G = \Delta G_{\text{bind}}(\text{mutant}) - \Delta G_{\text{bind}}(\text{wild-type})$, is obtained. The correlation between calculated and experimental data from Krowarsch *et al.*[32] of the binding free energy, ΔG, is shown in Fig. 4(a); the relation of changes in the binding free energy with a single mutation is shown in Fig. 4(b) and the values are also listed in Table II. In Fig. 4(a), there are two mutants data which give the positive binding affinity (repulsion) instead of small negative affinity as the experiment shows.

The calculation of the binding free energy of BPTI-trypsin complexes shows the positive binding energy for the acidic P_1 Asp and P_1 Glu variants in BPTI-trypsin complexes. Considering the binding arrangement of the P_1-S_1 site in the BPTI-trypsin complex, the electrostatic repulsion between S_1 Glu and acidic P_1 makes the binding too unfavorable when both residues are negatively charged. The experimental result still indicates favorable binding affinities of P_1 Asp and P_1 Glu mutants, $\Delta G = -6.478$ and -8.534, re-

TABLE II. Comparison of the binding free energy between the experimental data and calculated ones ΔG (kcal/mol). The first set is the results of P_1 mutants of BPTI-trypsin complexes, the second and third sets are the results of P_1 mutants of barnase-barstar complexes and SGPB-OMTKY3 complexes, respectively. The PDB codes used here are based on the PDB code of the wild-type template for each complex set: the first four letter code is the experimental PDB code of the wild-type and fifth code is the one-letter code of the mutated amino acid. The wild-type itself is shown in bold font. Kcal/mol unit is used for all energy terms. The calculated binding free energies in each complex set are shifted by setting the calculated binding free energy of the wild-type equal to the experimental binding free energy as our model only includes the electrostatic and van der Waals energy contributions.

PDB	ΔG_{obs}	ΔG_{calc}
3BTK	−17.86	−17.86
3BTKD	−6.48	−8.05
3BTKE	−8.53	−4.82
3BTKF	−10.91	−10.44
3BTKG	−5.64	−5.62
3BTKH	−9.17	−8.56
3BTKM	−10.25	−9.11
3BTKQ	−8.59	−7.76
3BTKT	−7.37	−7.39
3BTKW	−9.29	−10.40
1B27	−19.00	−19.00
1B27A	−16.70	−16.36
1B27C	−16.50	−16.45
1B27F	−16.80	−16.64
1B27Q	−17.60	−17.66
1B27S	−16.00	−16.42
1B27W	−17.40	−17.11
1B27Y	−16.60	−16.68
3SGB	−14.51	−14.51
3SGBA	−11.55	−12.01
3SGBC	−14.52	−13.88
3SGBD	−8.90	−7.29
3SGBE	−8.59	−8.18
3SGBF	−13.15	−12.11
3SGBH	−12.81	−10.73
3SGBI	−10.07	−11.27
3SGBK	−11.36	−9.87
3SGBM	−14.08	−14.42
3SGBR	−11.17	−9.95
3SGBS	−10.39	−10.63
3SGBT	−11.34	−9.04
3SGBV	−11.50	−12.76
3SGBW	−12.66	−11.66

spectively. Thus the local environmental effect on pK_a, hence the net charges on the residue, may be important in certain situations.

The pK_a shift of P_1 Glu mutant in OMTKY3-SGPB complex from 4.46 (unbound) to 8.74 (bound) was measured by Qasim *et al.*[42] Brandsdal *et al.*[43] calculated the pK_a of P_1 Glu mutant in OMTKY3-SGPB complex to be 13.1. They also calculated the pK_a shift of P_1 Glu mutant of BPTI-trypsin complex upon binding from 4.3 to 14.3. Even though their calculations of pK_a shifts are overestimated, the idea that acidic P_1 Glu in BPTI-trypsin complex may be protonated, thus negative charge no longer exists in the complex under pH=8.3 condition, may be possible. To obtain reliable data of local pK_a in unbounded state and bounded state, we

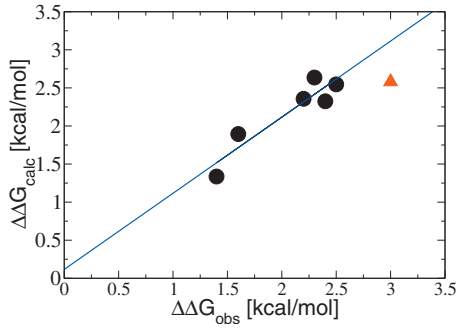

FIG. 5. Calculated vs observed changes of the binding free energies from P_1 mutants of barnase-barstar complexes. The linear fit from all mutant sets yields 0.890. A residue (E73Ser) reported to form hydrogen bonds with water molecules in the complex is indicated by red up-triangles in barnase-barstar complexes. The linear fit excluding this mutation involving water mediated hydrogen bonds yields 0.932 for the barnase-barstar complexes.

used the PROPKA 2.0 (Ref. 15) since this program is known to be the most accurate one to predict the pK_a values of amino acids based on extensive comparisons with a large set of experimentally determined pK_a.[44] For our calculations it gives pK_a values 8.7 and 8.8 for P_1 Asp and P_1 Glu mutants in BPTI-trypsin complexes and also gives pK_a value 8.8 for P_1 Glu mutant in OMTKY3-SGPB complex which is essentially the same as the experimental one 8.74 from Qasim *et al.*[42] With this shifted pK_a due to local environments, the calculated binding free energies are much improved as compared with the experimental one shown in Fig. 4(a). The red arrows indicate the binding free energy changes from the positive one (using generic pK_a) to the negative one (using pK_a that accounts for the local environments).

After considering the pK_a shifts for P_1 Asp and P_1 Glu in BPTI-trypsin complexes, the correlation of changes in the binding affinities between the observed and calculated data is improved to have a correlation coefficient 0.912. Without the two acidic P_1 data the correlation coefficient is 0.982. Some additional molecular level effects, such as the stabilization effect of water molecules in the interface of the two proteins, might play a role. In our model, water molecules are not explicitly represented, thus, the effect of the hydrogen bonds between water molecules and the side chains of interfacial amino acids are not considered in the binding energy calculations. For example, Helland *et al.*[37] observed that water molecules Sol653 and Sol654 participate in forming the hydrogen bonds with the carboxylate group of P_1 Asp and P_1 Glu and the interfacial interaction between P_1-S_1 is stabilized by the bridge-forming water molecules. Nevertheless, our coarse-grained model at the residue level can clearly capture most of the important contributions of the binding energy except some rare situations where local environments at atomic level need to be taken into account.

B. Binding energy calculations of barnase-barstar complexes

The residual model is also applied to a set of barnase-barstar complexes. As comparisons, the experimental data set from Schreiber *et al.*[33] for barnase-barstar is used to correlate with our calculations. In Fig. 5, a linear fit yields the correlation coefficient 0.890 for barnase-barstar complex set. The

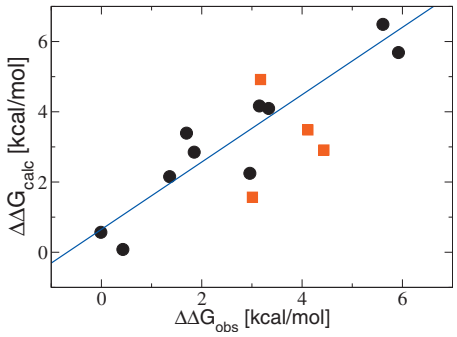

FIG. 6. Calculated vs observed changes in the binding free energy brought by P_1 mutants of OMTKY3-SGPB complexes. The linear fit from all mutant sets yields 0.828. Deleterious effects of β-branched residues are indicated by red rectangles in the figure of OMTKY3-SGPB complexes. Excluding the mutants indicated by red rectangles the linear fit coefficient raises to 0.945.

calculated binding free energies, ΔG, for this set and the changes in binding free energies, $\Delta\Delta G$, from the single mutations on the active site are listed in Table II.

Excluding a mutant which may involve additional hydrogen bonds, the linear fit correlation improves from 0.890 to 0.932. The mutants from the wild-type Glu73 in barnase-barstar complexes show the loss of hydrogen bonds and insertion of additional water molecules reducing the loss in binding energy. Especially for the Ser73E mutant in our calculation (the red triangle in Fig. 5) the loss of the hydrogen bond and insertion of a water molecule causing destabilization may be more severe than other. Lo Conte *et al.*[45] and Bahadur *et al.*[46] analyzed the interface of protein-protein complexes with the interfacial atomic structures and classified the ratios of water molecule participation. According to this analysis, all the interfacial residues in the barnase-barstrar complex are buried with water molecules. Again, given the simplicity of our model without explicit water modeling the correlation between the observed and the calculated binding energies is quite good for this system.

C. Binding energy calculations of OMTKY3-SGPB complexes

The experimental pK_a shift of OMTKY3 P_1 Glu bound to SGPB is 8.7 (see Qasim *et al.*),[42] using the PROPKA 2.0 (Ref. 15) pK_a values 8.7 and 8.8 for OMTKY3 P_1 Asp and P_1 Glu are obtained. With these shifted pK_a, the binding free energies, ΔG, of protein complex set are calculated and changes in binding free energies, $\Delta\Delta G$, from the single mutations on the active site are listed in Table II. The correlation of $\Delta\Delta G$ data of our calculations with the observed data from Lu *et al.*[34] yields the linear fit correlation coefficient 0.828.

After taking into account the pK_a shifts of the acidic P_1 mutants, there are four additional exceptional data points in the correlation fitting in Fig. 6. If all exceptional data points are excluded, the correlation between our calculations and the experimental results of $\Delta\Delta G$ yields an improved linear fit from 0.828 to 0.945.

The SGPB protein prefers hydrophobic P_1 side chain which are not branched at β-carbon.[34] The wild-type Leu18I fits into the S_1 pocket of SGPB binding site (Fig. 4 in Bateman *et al.*).[40] This pocket has narrow top entrance and broadening cavity toward the bottom. This narrow top structure causes that the β-branched residues cannot fit into the pocket. Thus, the β-branched side chains are not complementary to the shape of the S_1 binding site. The observed χ_1 angles of these residues in S_1 pocket are approximately 40° (Ile18I, 33°; Val18I, 47°; Thr18I, 39°; Ser18I, −46° or 40°) that are rotated ≅180° away from their actual orientations (for Val mutation, see Fig. 4 from Bateman *et al.*).[40] The alternate conformations for Ser18I O^{γ} are also observed (Fig. 6 in Bateman *et al.*).[40] When the β-branched residues involve binding the bottom of the pocket is left relatively empty to avoid the steric crashes in contrast with the wild-type Leu18I whose side chain tightly fits into the bottom. Finally the empty cavity which is rare in protein-protein recognition site[47] causes the complementary action involving close packing of the atoms between the two protein molecules. The destabilization of a protein complex with respect to the cavity made by the mutation with β-branched residue is directly proportional to the cavity size. This uneven empty cavity followed by the closed packing from the β-branched residue mutation finally alters the geometric structure of the interface and this effect is not described in our residual model. That is why the binding free energies of Ile18I, Val18I, Thr18I, and Ser18I are more widely spread in Fig. 6. On the other hand, our model can still account for the major changes of the binding energies due to single mutations for this system besides some effects due to atomic details.

IV. CONCLUDING REMARKS

Three sets of protein-protein binding complexes, BPTI-trypsin, barnase-barstar, and OMTKY3-SGPB, are studied using our residue level protein-protein interaction model.[11,12] These complex sets involve changes in binding affinities of mutations on positively charged, negatively charged and neutral residues on the interfacial surfaces. Using the Poisson–Boltzmann linear integral equation solver implemented with the fast multipole method to calculate the electrostatic and the van der Waals interaction free energy, reasonable agreements with the binding affinities of these complexes from experiments demonstrate the utility of such a coarse-grained model to capture the most important contributions of protein binding.

At the same time additional effects due to atomic details during binding have to be considered to yield accurate binding affinities. For example, for P_1 Asp and P_1 Glu mutants in BPTI-trypsin complexes, the calculated pK_a based on the PROPKA 2.0 (Ref. 15) to describe the neutral behaviors of acidic residues greatly improve the correlations with experimental data. Considering the limited accuracy of calculations of residual pK_a, there will be a possible improvement of binding free energy calculations by using the experimental pK_a especially for the residues may have large charge changes upon binding.

ACKNOWLEDGMENTS

The authors (B.K. and X.S.) are grateful to the financial support from an NSF Grant No. CHE-0809431.

APPENDIX: THE FORMULATIONS OF THE FAST MULTIPOLE METHOD

The linear equations in Eqs. (4) and (5) for the electrostatic energy calculation and Eqs. (21) and (22) for the van der Waals energy calculation have the following form:

$$(I-L)A=B, \tag{A1}$$

where I is the identity matrix with the size of N^2, A, and B are single column vectors with the size of N, the number of amino acid residues, for the electrostatic energy calculation and are $N\times N$ matrix for the reaction field of the van der Waals energy calculation. Rewriting this equation with more details yields the following form:

$$\begin{pmatrix} \varphi_0 & 0 \\ 0 & \varphi_1 \end{pmatrix} - \begin{pmatrix} L_1 & L_2 \\ L_3 & L_4 \end{pmatrix} \begin{pmatrix} \varphi_0 \\ \varphi_1 \end{pmatrix} = \begin{pmatrix} F_0 \\ F_1 \end{pmatrix}, \tag{A2}$$

where the matrix elements, L_1, L_2, L_3, and L_4, are defined in Eqs. (6)–(9). All matrix elements consist of terms involving summation of Columbic interaction and screened Columbic interaction. The key point for solving this linear system with efficiency is to accelerate the matrix-vector multiplications during iterations in the iterative linear equation solver, such as GMRES. This can be done by introducing the FMM. implementation of FMM is described as follows.

The multipole moment expansion for the Columbic interaction is given by[48]

$$\int_{S_y} \frac{\partial F(\mathbf{x}-\mathbf{y})}{\partial n_y}\phi(\mathbf{y})dS_y = \frac{1}{4\pi}\sum_{n=0}^{p}\sum_{m=-n}^{n} \overline{S_{n,m}}(\vec{O\mathbf{x}})M_{n,m}(O), \tag{A3}$$

where Green's function $F(\mathbf{x}-\mathbf{y})=1/|\mathrm{x}-\mathbf{y}|$ and the multipole moment coefficients is

$$M_{n,m}(O) = \int_{S_y} \frac{\partial R_{n,m}(\vec{O\mathbf{y}})}{\partial n_y}\phi(\mathbf{y})dS_y \tag{A4}$$

and $R_{n,m}$ and $S_{n,m}$ are the solid harmonics defined as

$$\overline{S_{n,m}}(\vec{O\mathbf{x}}) = (n-m)!P_n^m(\cos\,\theta)e^{-im\phi}\frac{1}{r^{n+1}}, \tag{A5}$$

$$R_{n,m}(\vec{O\mathbf{y}}) = \frac{1}{(n+m)!}P_n^m(\cos\,\alpha)e^{im\beta}\rho^n. \tag{A6}$$

The multipole moment expansion for the screened Columbic interaction can be written as

$$\int_{S_y} \frac{e^{-\kappa|\mathrm{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}\phi(\mathbf{y})dS_y = \frac{2\kappa}{\pi}\sum_{n=0}^{p}(2n+1)k_n(\kappa r)\sum_{m=-n}^{n}\overline{S_{n,m}}(\theta,\phi)M_{n,m}(\kappa,O), \tag{A7}$$

where the multipole coefficients

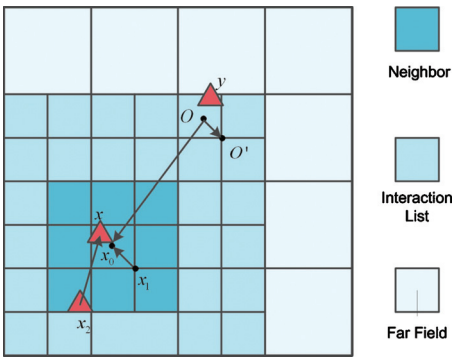

FIG. 7. A schematic illustration showing the hierarchical rectangular boxes in two dimensional space for convenience. The largest box represents the highest level, level-zero, and the smallest boxes are in the finest level, level-three in this picture. The lightly shaded level-two boxes are in the far field list from the target point **x**. The light blue boxes are in the interaction list (up to 189 boxes in three dimension) which translates the multipole expansion to the local expansion. The arrows O to O' and $\mathbf{x}_1$ to $\mathbf{x}_0$ indicate multipole to multipole and local to local translation, respectively. Finally the dark blue boxes (up to 27 boxes in three dimension) are the neighbor boxes. The interaction between the neighbors including the self-interaction can be calculated by the direct BEM solver (Ref. 22).

$$M_{n,m}(\kappa,O) = \int_{S_y} i_n(\kappa\rho)R_{n,m}(\alpha,\beta)\phi(y)dS_y \tag{A8}$$

and $i_n(\kappa\rho)$ and $k_n(\kappa r)$ are modified spherical Bessel and modified spherical Hankel functions are defined in terms of Bessel function.[49]

$$I_\nu(r) = i^{-\nu}J_\nu(ir), \tag{A9}$$

$$K_\nu(r) = \frac{\pi}{2\,\sin\,\nu\pi}[I_{-\nu}(r)-I_\nu(r)], \tag{A10}$$

$$i_n(r) = \sqrt{\frac{\pi}{2r}}I_{n+1/2}(r), \tag{A11}$$

$$k_n(r) = \sqrt{\frac{\pi}{2r}}K_{n+1/2}(r). \tag{A12}$$

and $R_{n,m}(\alpha,\beta)$ and $\overline{S_{n,m}}(\theta,\phi)$ are the spherical harmonics are defined as

$$S_{n,m}(\theta,\phi) = R_{n,m}(\theta,\phi) = \sqrt{\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\,\theta)e^{im\phi}, \tag{A13}$$

where the upper bar represents the complex conjugate of the harmonics. The integrals in Eqs. (A3) and (A7) can be evaluated with the local expansion coefficients as follows:

$$\int_{S_y} \frac{\partial F(\mathbf{x}-\mathbf{y})}{\partial n_y}\phi(\mathbf{y})dS_y = \frac{1}{4\pi}\sum_{n=0}^{p}\sum_{m=-n}^{n}\frac{\partial R_{n,m}(\vec{\mathbf{x}_0\mathbf{x}})}{\partial n_x}L_{n,m}(\mathbf{x}_0), \tag{A14}$$

$$\int_{S_y} \frac{e^{-\kappa|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}\phi(\mathbf{y})dS_y = \frac{2\kappa}{\pi}\sum_{n=0}^{p}(2n+1)i_n(\kappa r)\sum_{m=-n}^{n}\overline{S_{n,m}}(\theta,\phi)L_{n,m}(\kappa,\mathbf{x}_0). \tag{A15}$$

The expression of the local expansion coefficient for the Columbic interaction can be written as follows:[48]

$$L_{n,m}(\mathbf{x}_0) = \sum_{n'=0}^{p}\sum_{m'=-n'}^{n'}(-1)^{n'}\overline{S_{n+n',m+m'}}(\overrightarrow{O\mathbf{x}_0})M_{n',m'}(O). \tag{A16}$$

The above procedure is called, "the multiple to local translation (simply M2L translation)." The equation for the M2L translation of screened Coulombic interaction can be derived using the properties[48] of the translational equalities in the spherical Bessel and Hankel functions derived by Epton and Dembart[50] and applying them to the modified spherical Bessel and Hankel functions. The final expression of the M2L translation is given by

$$L_n^m(\kappa,\mathbf{x}_0) = \sum_{n'=0}^{p}\sum_{m'=-n'}^{n'}\sum_{\substack{l=|n-n'|\\ n'+n-l:\text{even}}}^{n+n'}(2n'+1) \times W_{n',n,m',m,l}k_l(\kappa,\overrightarrow{O\mathbf{x}_0}) \times\overline{S_{l,-m-m'}}(\overrightarrow{O\mathbf{x}_0})M_{n',m'}(\kappa,O), \tag{A17}$$

where $W_{n',n,m',m,l}$ is written by the following equation with the Wigner-3j symbol.[51]

$$W_{n,n',m,m',l} = (2l+1)i^{n'-n+l}\begin{pmatrix} n & n' & l \\ 0 & 0 & 0 \end{pmatrix} \times\begin{pmatrix} n & n' & l \\ m & m' & -m-m' \end{pmatrix}. \tag{A18}$$

The oct-tree structure source code developed by Song[52] is used to find its "interaction list," which has a key role to connect the multipole expansion to the local expansion coefficients, the M2L translation. At the finest level, the interaction between the elements in the nearest neighbor, called the near field interaction, can be calculated by the direct boundary element solver with the collocation method from Atkinson and co-worker[26] and the interaction from the far field elements, the multipole moment expansion coefficients are translated to the higher level expansion, called "the multipole to multipole translation (M2M)." Once the M2L translations are computed in the higher level of tree structure, they should be translated to the lower level local expansions, finally the local expansions in the finest level in order to evaluate the integrals and the matrix-vector multiplications. This process is called "local to local translation (L2L)." The equations for M2M and L2L translations for Columbic and screened Columbic interactions are given below:

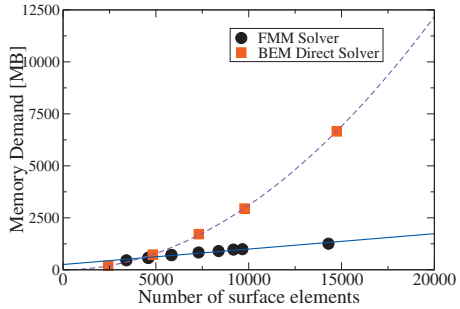

FIG. 8. The comparison of memory requirements between BEM direct solver and FMM solver for the calculation of electrostatic energy contribution to the binding of BPTI-trypsin complex. The red square boxes represent the memory usage from the BEM direct solver and the dashed line is a curve fitting with power 1.973. The black spheres represent the FMM solver data and the blue solid line is a curve fitting with power 1.00 with maximum level of tree=6.

$$M_{n,m}(O') = \sum_{n'=0}^{n}\sum_{m'=-n'}^{n'}R_{n',m'}(\overrightarrow{O'O})M_{n-n',m-m'}(O), \tag{A19}$$

$$M_n^m(\kappa,O') = \sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'}\sum_{\substack{l=|n-n'|\\ n'+n-l:\text{even}}}^{n+n'}(2n'+1) \times(-1)^{m'}W_{n,n',m,m',l}i_l(\kappa,\overrightarrow{O'O}) \times S_{l,-m-m'}(\kappa,\overrightarrow{O'O})M_{n',-m'}(\kappa,O), \tag{A20}$$

$$L_{n,m}(\mathbf{x}_0) = \sum_{n'=n}^{\infty}\sum_{m=-n'}^{n'}R_{n'-n,m'-m}(\overrightarrow{\mathbf{x}_1\mathbf{x}_0})L_{n',m'}(\mathbf{x}_1), \tag{A21}$$

$$L_n^m(\kappa,\mathbf{x}_0) = \sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'}\sum_{\substack{l=|n-n'|\\ n'+n-l:\text{even}}}^{n+n'}(2n'+1) \times(-1)^{m}W_{n',n,m',-m,l}i_l(\kappa,\overrightarrow{\mathbf{x}_1\mathbf{x}_0}) \times S_{l,m-m'}(\overrightarrow{\mathbf{x}_1\mathbf{x}_0})L_{n',m'}(\kappa,\mathbf{x}_1). \tag{A22}$$

The restarted generalized minimal residual method[27] is used to solve the linear equations and we modified the code computing the matrix-vector product to make the interface to the FMM. The comparison between the direct BEM solver and our FMM solver in computational cost is shown on Fig. 8. According to this figure our FMM code has a linear dependence on the number of element N at the maximum level $\log N$, finally it follows the order $O(N \log N)$ algorithm. Figure 8 shows the comparison of the memory demand between the BEM direct solver and the FMM-BEM solver for the calculations of electrostatic energy contribution to the binding of BPTI-trypsin complexes. With the implementation of FMM in our model, our electrostatic free energy solver only takes about 1 Gbyte of memory, significantly smaller than the BEM direct solver which occupies more than 7 Gbyte using about 15 000 surface elements. The BEM solver fol-

lows $O(N^2)$ algorithm whereas the FMM solver is an $O(N \log N)$ algorithm. This trend can also be applied to the computational time. The reduced computational cost in terms of memory and time by the implementation of the FMM algorithm; our model can be used to calculate the binding free energy of mutations of protein-protein complexes efficiently.

[1] D. B. Target, http://targetdb.pdb.org/statistics/index.html, 2010.
[2] R. F. Service, Science **307**, 1554 (2005).
[3] C. J. Gerdts, V. Tereshko, M. K. Yadav, I. Dementieva, F. Collart, A. Joachimiak, R. C. Stevens, P. Kuhn, A. Kossiakoff, and R. F. Ismagilov, Angew. Chem., Int. Ed. **45**, 8156 (2006).
[4] M. J. Anderson, C. L. Hansen, and S. R. Quake, Proc. Natl. Acad. Sci. U.S.A. **103**, 16746 (2006).
[5] A. Elcock, D. Sept, and J. A. McCammon, J. Phys. Chem. B **105**, 1504 (2001).
[6] X. Song, the protein data bank entries are 6lyt,1lzt,1lkr,0lzt,1lys and the crystallization conditions are from www.bmcd.nist.gov:8080/bmcd, unpublished results, 2002.
[7] M. Crosio, J. Janin, and M. Jullien, J. Mol. Biol. **228**, 243 (1992).
[8] O. Carugo and P. Argos, Protein Sci. **6**, 2261 (1997).
[9] J. Janin and F. Rodier, Proteins **23**, 580 (1995).
[10] S. Dasgupta, G. Iyer, S. Bryant, C. Lawrence, and J. Bell, Proteins **28**, 494 (1997).
[11] X. Song, Mol. Simul. **29**, 643 (2003).
[12] X. Song and X. Zhao, J. Chem. Phys. **120**, 2005 (2004).
[13] A. A. Zamyatnin, Annu. Rev. Biophys. Bioeng. **13**, 145 (1984).
[14] M. Sanner, http://www.scripps.edu/sanner/html/msms_home.html.
[15] C. B. Delphine, M. R. David, and H. J. Jan, Proteins: Struct., Funct., Bioinf. **73**, 765 (2008).
[16] X. Song, J. Chem. Phys. **116**, 9359 (2002).
[17] S. Millefiori, A. Alparone, A. Millefiori, and A. Vanella, Biophys. Chem. **132**, 139 (2008).
[18] F. Dong and H.-X. Zhou, Proteins: Struct., Funct., Bioinf. **65**, 87 (2006).
[19] K. Brock, K. Talley, K. Coley, P. Kundrotas, and E. Alexov, Biophys. J. **93**, 3340 (2007).
[20] L. Greengard, *The Rapid Evaluation of Potential Fields in Particle Systems*, ACM Distinguished Dissertations (MIT Press, Cambridge, MA, 1988).
[21] L. Greengard and V. Rokhlin, J. Comput. Phys. **135**, 280 (1997).
[22] B. Lu, X. Cheng, and J. A. McCammon, J. Comput. Phys. **226**, 1348 (2007).
[23] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, Biophys. J. **72**, 1047 (1997).
[24] B. Yoon and A. Lenhoff, J. Comput. Chem. **11**, 1080 (1990).
[25] 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).
[26] K. Atkinson and W. Han, *Numerical Solution of Fredholm Integral Equations of the Second Kind*, Texts Applied in Mathematics, 3rd ed. (Springer, New York, 2009).
[27] R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods* (SIAM, Philadelphia, 1994).
[28] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999).
[29] J. Israelachvili, *Intermolecular and Surface Forces* (Academic, San Diego, 1991).
[30] V. Parsegian, *Physical Chemistry: Enriching Topic from Colloid and Surface Science* (Theorex, La Jolla, CA, 1975).
[31] M. Karplus and R. N. Porter, *Atoms and Molecules; an Introduction for Students of Physical Chemistry* (Benjamin, New York, 1970).
[32] D. Krowarsch, M. Dadlez, O. Buczek, I. Krokoszynska, A. O. Smalas, and J. Otlewski, J. Mol. Biol. **289**, 175 (1999).
[33] G. Schreiber, C. Frisch, and A. R. Fersht, J. Mol. Biol. **270**, 111 (1997).
[34] W. Lu, I. Apostol, M. A. Qasim, N. Warne, R. Wynn, W. L. Zhang, S. Anderson, Y. W. Chiang, E. Ogin, I. Rothberg, K. Ryan, and M. Laskowski, J. Mol. Biol. **266**, 441 (1997).
[35] N. Guex and M. C. Peitsch, *Swiss-Model and the Swiss-PDB Viewer: An Environment for Comparative Protein Modeling*, 1997, online at http://spdbv.vital-it.ch
[36] B. Brooks, R. Bruccoleri, D. Olafson, D. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[37] R. Helland, J. Otlewski, O. Sundheim, M. Dadlez, and A. O. Smal, J. Mol. Biol. **287**, 923 (1999).
[38] C. K. Vaughan, A. M. Buckle, and A. R. Fersht, J. Mol. Biol. **286**, 1487 (1999).
[39] R. Read, M. Fujinaga, A. Sielecki, and M. James, Biochemistry **22**, 4420 (1983).
[40] K. S. Bateman, M. N. G. James, S. Anderson, W. Lu, M. A. Qasim, and L. J. Michael, Protein Sci. **9**, 83 (2000).
[41] H. Kui, N. G. J. Michael, L. Wuyuan, J. M. Laskowski, and A. Stephen, Protein Sci. **4**, 1985 (1995).
[42] M. A. Qasim, M. R. Ranjbar, R. Wynn, S. Anderson, and M. Laskowski, J. Biol. Chem. **270**, 27419 (1995).
[43] B. O. Brandsdal, A. O. Smals, and J. Aqvist, BMC Biochemistry **64**, 740 (2006).
[44] M. Davies, C. Toseland, D. Moss, and D. Flower, BMC Biochemistry **7**, 18 (2006).
[45] L. Lo Conte, C. Chothia, and J. Janin, J. Mol. Biol. **285**, 2177 (1999).
[46] R. Bahadur and M. Zacharias, Cell. Mol. Life Sci. **65**, 1059 (2008).
[47] J. Janin and C. Chothia, J. Biol. Chem. **265**, 16027 (1990).
[48] K.-I. Yoshida, Ph.D. thesis, Department of Global Environment Engineering, Kyoto University, 2001.
[49] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1964).
[50] M. A. Epton and B. Dembart, SIAM J. Sci. Comput. (USA) **16**, 865 (1995).
[51] A. Messiah, *Quantum Mechanics* (North Holland, Amsterdam, Netherlands, 1962), Vol. 2.
[52] W. C. Chew, J.-M. M. Jin, and J. Song, *Fast and Efficient Algorithms in Computational Electromagnetics* (Artech House, Boston, MA, 2001).