

Hierarchical Molecular Interfaces and Solvation Electrostatics

Chandrajit Bajaj^{1,2}, Shun-Chuan Albert Chen¹, Guoliang Xu³, Qin Zhang², and Wenqi Zhao²

¹Department of Computer Sciences, The University of Texas at Austin, Austin, TX, USA

²Institute of Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA

³Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing, China

ABSTRACT

Electrostatic interactions play a significant role in determining the binding affinity of molecules and drugs. While significant effort has been devoted to the accurate computation of biomolecular electrostatics based on an all-atomic solution of the Poisson-Boltzmann (PB) equation for smaller proteins and nucleic acids, relatively little has been done to optimize the efficiency of electrostatic energetics and force computations of macromolecules at varying resolutions (also called coarse-graining). We have developed an efficient and comprehensive framework for computing coarse-grained PB electrostatic potentials, polarization energetics and forces for smooth multi-resolution representations of almost all molecular structures, available in the PDB. Important aspects of our framework include the use of variational methods for generating C^2 -smooth and multi-resolution molecular surfaces (as dielectric interfaces), a parameterization and discretization of the PB equation using an algebraic spline boundary element method, and the rapid estimation of the electrostatic energetics and forces using a kernel independent fast multipole method. We present details of our implementation, as well as several performance results on a number of examples.

1. INTRODUCTION

Atomistic simulation of bio-molecules is known to play a critical role in various biological activities such as drug design or molecular trajectory simulation. Given the complexity of large-scale data computation, developing an accurate and effective approach for the simulation has drawn great attention from recent computational biology studies. In particular, one important technique of atomistic simulations of bio-molecules can be carried out based on the numerical solutions of solvation energy and forces. Considerable research effort has been devoted to calculating solvation energy and forces in the past two decades. One important

solution, called the implicit solvent model, treats the solvent as a featureless dielectric material. The effects of the solvent are modeled in terms of dielectric and ionic physical properties instead of the micro elements such as ions used in explicit solvent models. Poisson-Boltzmann (PB) equation are widely used to obtain good electrostatic approximations in implicit solvent models.

The interface between the atomic-level solute and continuum solvent is key to an implicit solvent model. This interface is also called the solvent-excluded surface (SES) or simply molecular surface [13]. Since the molecular surface also acts as a dielectric interface for electrostatic and polarization energy and force computations, the molecular surface should be at least C^1 smooth and not too inflated or deflated. In our computational framework, we use the C^2 smooth level set of a tri-cubic B-spline function to model a minimal area molecular surface [5].

Because a macromolecule is composed of thousands to millions of atoms, it makes the simulation computation costly. A lot of important efforts have been devoted to develop lower resolution coarse-grained (CG) models for proteins with reasonable accuracy. One of the earliest and simplest models is the Gō model which represents the polypeptide chain as a chain of C_α atoms [20]. In [3] this model has been improved by adding one more bead on each side chain (SC). In the C_α -SC-Pep model [12], an additional interaction center (Pep) is added on the backbone in the middle of the C-N peptide bond which strongly improves the orientation-dependent potentials. In [8] extended side chains (such as Arg, Lys, etc.) are represented by two beads in order to have CG beads of about the same size. A four-bead model is given in [18] in which each residue is explicitly represented by three heavy atoms on the backbone and one bead on the side chain. A multi-resolution CG model is developed in [2] which allows different resolution in different parts of the molecule and therefore fixes the deficiency of assigning each CG bead to the same number of atoms. In this short paper, we briefly present a hierarchical CG clustering scheme which can flexibly generate a CG molecular model for electrostatic potential, polarization energetics and forces. Details can be found in [7].

After the molecular model and the molecular interface of a molecule are defined and constructed, the boundary element method (BEM), one of the most common numerical methods, is applied for solving molecular electrostatics problems in this paper. BEM relies on the molecular interface

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIAM/ACM Joint Conference on Geometric and Physical Modeling '09 San Francisco, California USA

Copyright 2009 ACM X-XXXXXX-XX-X/XX/XX ...\$10.00.

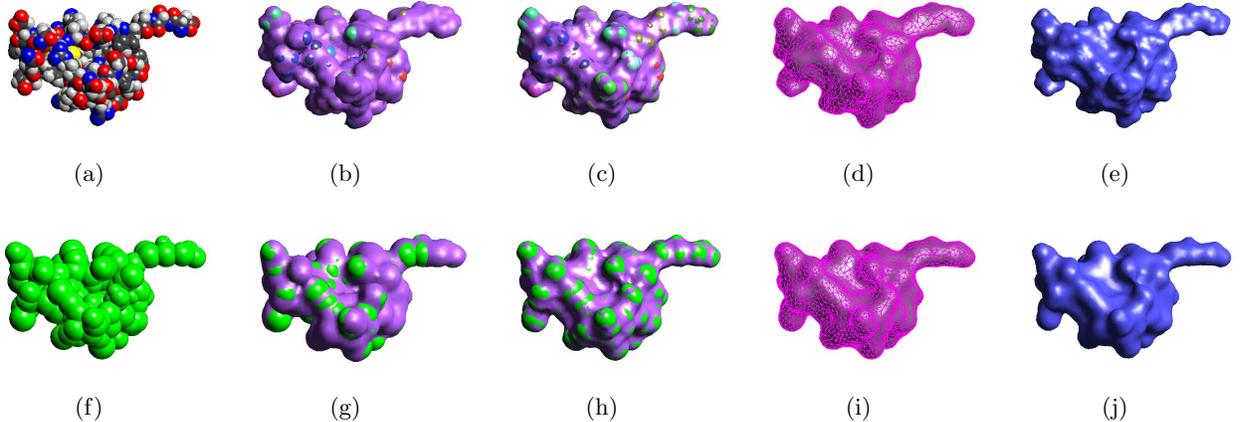


Figure 1: Molecular model of a protein(1CGI); (a),(f) The van der Waals surface of the all-atomic (AA) and coarse-grained (CG) structures (852 atoms and 157 beads) ; (b),(g) The surface generated using Gauss density function from AA and CG structures; (c),(h) The solvent excluded surfaces (SES) from AA and CG structures which are much close to the surface; (d),(i) The decimated triangulation of SESs; (e),(j) The piecewise algebraic surfaces patches generated from the decimated triangulation of SESs. The visual error between SESs generated from AA and CG models is small.

to discretize the implicit continuum electrostatics problems. Since Zauhar and Morgan introduced a BEM paper on continuum electrostatic of biological systems [23], in the past two decades, scientists have made contributions to improve and extend BEM solutions and performances and proved that BEM is one of the most accurate and efficient solutions for molecular electrostatics problems [9, 11]. Based on the electrostatic solution of BEM, electrostatic force can also be evaluated, and thereby different numerical treatments of electrostatic force are also described [14, 10, 15]. In this paper, we will introduce a algebraic spline boundary element method for numerically solving PB electrostatic problem.

The rest of the paper will be organized by the steps of pipeline. In section 2, the details of construction of our hierarchical molecular models are given, followed by C^2 smooth spline discretization of molecular interfaces in section 3. In section 4, an implicit continuum model and boundary element solution for PB electrostatics are introduced. In section 5, we present our numerical implementation and analyze experimental results in details. We conclude the paper in section 6.

2. PROTEIN MOLECULAR MODELS

The 3-D structure data of the protein molecules are obtained from the RCSB protein data bank (PDB). We construct a CG model in three steps. First we build a hierarchical clustering of the atoms according to the hierarchy of the protein structure. In this hierarchy, from top to bottom, they are the tertiary structures, secondary structures, residues, backbone and side chains, functional groups, and atoms. All the atoms in one of the groups can be represented as one CG bead. Since at the top levels, too much detail of the protein is lost, the coarsest CG model we start at in our current work is to group the atoms in an amino acid into one bead.

In the second step, we compute the new locations and

sizes of the CG beads. Our goal is to let the new molecular surface of the CG model be as close to the surface of the AA model as possible, where the molecular surface is modeled as the level set of the Gaussian density function. For the purpose of accuracy, we individually find the center (\vec{x}'_k) and the radius (r'_k) for each CG bead such that the Gaussian density function $\rho(\vec{x}) = e^{-C_i(\|\vec{x}-\vec{x}'_k\|^2 - (r'_k)^2)}$, where C_i is a Gaussian decay rate, agrees with the density function $G_i(\vec{x}) = \sum_{k=i_1}^{i_{M_i}} e^{-C_k(\|\vec{x}-\vec{x}'_k\|^2 - r_k^2)}$, where atoms $i_1 \dots i_{M_i}$ are grouped into bead i . This is done by solving the least squares problem

$$\min \frac{1}{2} \sum_{j=1}^n [\rho(\vec{x}_j) - G_i(\vec{x}_j)]^2,$$

where \vec{x}_j are sample points on the level set $\{\vec{x} : G_i(\vec{x}) = 1\}$.

In the third step, we assign charges to the CG beads such that the electrostatic solvation energy of the CG model reproduces that of the atomic model. We use the Generalized Born energy function [19] as the objective function and the optimization is subject to the constraint that the total charge of the molecule does not change. This constrained nonlinear optimization problem is solved by using the Levenberg-Marquardt algorithm [16]. The details of the CG model generation are described in [7]. In Figure 1, we show an example of the AA and CG model and the corresponding molecular surface. In the CG model, each amino acid is split into five groups and therefore five beads, two for the backbone and three for the side chain. The molecular surface in (e) and the molecular surface in (j) are similar and the Hausdorff distance between them is 1.881 Å. The CG model also well preserves the topology of the AA model.

2.1 variational C^2 -smooth Molecular interface

In this section, we sketch the method to produce smooth molecular surfaces. For details one is referred to [5]. The

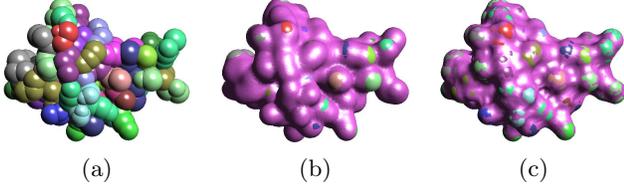


Figure 2: Molecular interface for a protein 1BUH; (a) the Van der Waal (union of balls) model; (b) the initial molecular surface computed using Gaussian density function; (c) the molecular surface computed using geometric flow evolution.

basic idea can be summarized as three steps. Step one, for given a molecule which consists of a sequence of atoms with centers $\{\vec{x}_k\}_{k=1}^{n_c}$ and radii $\{r_k\}_{k=1}^{n_c}$, we construc an initial surface such that $\{\vec{x} \in \mathbb{R}^3 | g(\vec{x}) = 0\}$, where

$$g(\vec{x}) = 1 - \sum_{k=1}^{n_c} e^{-C_k(\|\vec{x} - \vec{x}_k\|^2 - \tilde{r}_k^2)}, \quad \tilde{r}_k = r_k + r_b,$$

where r_b is the water probe sphere radius. The Gaussian decay rate $C_k > 0$, is determined so that $g(\vec{x}) = 0$ is an approximation of SES within a given tolerance [5].

Step two, for the obtained initial surface Γ^0 , we define an energy functional on the surface Γ as

$$E(\Gamma) = \int_{\Gamma} g(\vec{x})^2 d\vec{x} + \tau \int_{\Gamma} d\vec{x}, \quad (1)$$

where $\tau \geq 0$ is a regularization coefficient and the second term is to minimize the surface area avoiding both big inflation and deflation of the surface. We pursue the surface which minimizes energy functional (1) by variational calculus, which computes the first order variation to obtain a partial differential equation (PDE) [21]. The PDE is solved as an evolutionary equation by adding a time marching parameter t other than a stationary equation where the evolutionary equation is expressed as level set formulation as

$$\begin{aligned} \frac{\partial \Phi}{\partial t} &= (g^2 + \tau) \operatorname{div} \left(\frac{\nabla \Phi}{\|\nabla \Phi\|} \right) \|\nabla \Phi\| + 2g(\nabla g)^T \nabla \Phi \\ &= H(\Phi) + L(\nabla \Phi), \end{aligned} \quad (2)$$

where

$$H(\Phi) = (g^2 + \tau) \operatorname{div} \left(\frac{\nabla \Phi}{\|\nabla \Phi\|} \right) \|\nabla \Phi\|, \quad L(\nabla \Phi) = 2g(\nabla g)^T \nabla \Phi.$$

Note here the moving surface Γ is expressed as the level set of function Φ , that is, $\Gamma = \{\vec{x} \in \mathbb{R}^3 | \Phi(\vec{x}) = 0\}$. This evolution equation is solved by our higher-order level-set methods [5]. The first order term $L(\nabla \Phi)$ is computed using an upwind scheme over a finer grid, and the higher order term $H(\Phi)$ is computed using a spline presentation defined on a coarser grid. Step three, if Φ is a signed distance function and a steady solution of equation (2), the iso-surface $\Phi = -r_b$ is an approximation of molecular surface.

2.2 Algebraic-spline parametrization

The molecular surface constructed in the last section is discretized into a triangulation when it is applied to the solvation energy computation. An algebraic spline model

(ASMS) which provides a dual implicit and parametric representation of the molecular surface is generated based on this triangulation [24]. For each element $[v_i, v_j, v_k]$ in the triangulation, points on the algebraic spline are defined as $p = b_1 v_i(\lambda) + b_2 v_j(\lambda) + b_3 v_k(\lambda)$, with (b_1, b_2, b_3) being the barycentric coordinates of the points in $v_i v_j v_k$ and λ satisfying

$$F(\lambda) = \sum_{i+j+k=3} b_{ijk}(\lambda) B_{ijk}^3(b_1, b_2, b_3) = 0,$$

where B_{ijk}^3 are the Bernstein-Bezier (BB) basis of degree 3. The coefficients $b_{ijk}(\lambda)$ are not trivial. They are defined so that $F(\lambda)$ is C^1 continuous across the edges of the triangle. We give a thorough explanation in [24], so we will not repeat the definition here.

3. SOLVATION ELECTROSTATIC ENERGY AND FORCE COMPUTATION

In this section, we briefly introduce the main steps the numerical treatments of solving solvation electrostatic problem for proteins. The details of the derivation and implementation are described in the manuscript we are writing [4].

Based on our definition of different protein models for proteins, the continuum model of proteins in the solvent is defined and used for numerical computation of solvation electrostatic computation. We separated the open domain (\mathbb{R}^3) into interior (Ω) and exterior regions ($\mathbb{R}^3 - \Omega$) by the molecular interface. The dielectric coefficient $\epsilon(\vec{x})$ and ion strength $\iota(\vec{x})$ at position \vec{x} depends on which region \vec{x} belongs to,

$$\epsilon(\vec{x}) = \begin{cases} \epsilon_i, & \vec{x} \in \Omega, \\ \epsilon_e, & \vec{x} \in \mathbb{R}^3 - \Omega. \end{cases} \quad \iota(\vec{x}) = \begin{cases} 0, & \vec{x} \in \Omega, \\ \iota, & \vec{x} \in \mathbb{R}^3 - \Omega. \end{cases}$$

Because the simulation environment is now in a dielectric medium, we describe the electrostatic behavior of the protein-solvent system by the linearized PB equation.

$$\vec{\nabla}(\epsilon(\vec{x}) \vec{\nabla} \phi(\vec{x})) + 4\pi \sum_{k=1}^{n_c} q_k \delta(\vec{x} - \vec{x}_k) = \bar{\kappa}(\vec{x})^2 \phi(\vec{x}),$$

where $\phi(\vec{x})$ is the electrostatic potential at x . q_k and \vec{x}_k are the charge and the position of atom k , $k = 1, \dots, n_c$. The modified Debye-Huckel parameter $\bar{\kappa}^2(\vec{x})$ is $\bar{\kappa}^2(\vec{x}) = \frac{8\pi e_c^2 \iota}{k_B T}$, where e_c is the charge of an electron. k_B is the Boltzmann's constant. T is the absolute temperature. ι is the ionic strengths.

In Figure 3, we show the scenario and notations of the PB boundary element formulations. The molecular surface Γ is discretized into triangular elements Γ_i , $i = 1, \dots, N$ and \vec{x}^i represents a point on an element Γ_i as \vec{y}^j on Γ_j . Their normal vectors are written as \vec{n}_x^i and \vec{n}_y^j . In order to obtain the electrostatic potential $\phi(\vec{x})$, we solve PB equation by formulating it into the derivative boundary integral equation by a BEM technique [1].

$$\begin{aligned} & \left[\begin{array}{cc} \frac{\epsilon_i + \epsilon_e}{\epsilon_i} I + \frac{\partial G_{\kappa}}{\partial \vec{n}_y} - \frac{\epsilon_e}{\epsilon_i} \frac{\partial G_{\Omega}}{\partial \vec{n}_y} & G_{\Omega} - G_{\kappa} \\ \frac{\partial^2 G_{\Omega}}{\partial \vec{n}_y \partial \vec{n}_x} - \frac{\partial^2 G_{\kappa}}{\partial \vec{n}_y \partial \vec{n}_x} & \frac{\epsilon_e + \epsilon_i}{2\epsilon_e} I + \frac{\partial G_{\kappa}}{\partial \vec{n}_x} - \frac{\epsilon_i}{\epsilon_e} \frac{\partial G_{\kappa}}{\partial \vec{n}_x} \end{array} \right] \begin{bmatrix} \phi \\ \frac{\partial \phi}{\partial \vec{n}} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^{n_c} \frac{q_k}{\epsilon_i} G_{\Omega, k} \\ \sum_{k=1}^{n_c} \frac{q_k}{\epsilon_i} \frac{\partial G_{\Omega, k}}{\partial \vec{n}_x} \end{bmatrix}, \end{aligned}$$

where ϕ_j and $\frac{\partial \phi_j}{\partial \vec{n}_j}$ are the j^{th} unknown electrostatic potential and its normal derivative at the point \vec{y}^j on the element Γ_j .

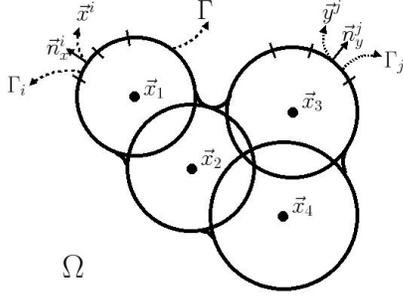


Figure 3: The example of boundary element decomposition of a four-atom molecule: \vec{x}_k are the center of k^{th} atom, \vec{x}^i and \vec{y}^j are the points on the element Γ_i and Γ_j of the surface Γ and \vec{n}_x^i and \vec{n}_y^j are their normals.

Each term in the matrix is an operator including I as an identity operator that $I_{ij}\phi_j = \phi_j$ and the surface integral operators given in the following expression,

$$\begin{aligned} \left(\frac{\partial G_l}{\partial \vec{n}_y^j} \right)_{ij} \phi_j &= \int_{\Gamma_j} \frac{\partial G_l}{\partial \vec{n}_y^j}(\vec{x}^i, \vec{y}^j) \phi(\vec{y}^j) d\vec{y}^j \\ (G_l)_{ij} \frac{\partial \phi_j}{\partial \vec{n}_y^j} &= \int_{\Gamma_j} G_l(\vec{x}^i, \vec{y}^j) \frac{\partial \phi(\vec{y}^j)}{\partial \vec{n}_y^j} d\vec{y}^j \\ G_{0,k} &= G_0(\vec{x}^i, \vec{x}_k) \end{aligned}$$

where \vec{x}^i and \vec{y}^j are points on the element Γ_i and Γ_j . $l = 0$ or κ such that G_0 and G_κ are the Green's kernels, or fundamental solutions, for linearized PB equations

$$\begin{aligned} G_0(\vec{x}, \vec{y}) &= \frac{1}{4\pi|\vec{x}-\vec{y}|} & \frac{\partial G_0(\vec{x}, \vec{y})}{\partial \vec{n}_y} &= \frac{-(\vec{x}-\vec{y}) \cdot \vec{n}_y}{4\pi|\vec{x}-\vec{y}|^3} \\ G_\kappa(\vec{x}, \vec{y}) &= \frac{e^{-\kappa|\vec{x}-\vec{y}|}}{4\pi|\vec{x}-\vec{y}|} & \frac{\partial G_\kappa(\vec{x}, \vec{y})}{\partial \vec{n}_y} &= \frac{-e^{-\kappa|\vec{x}-\vec{y}|}(1.0+\kappa|\vec{x}-\vec{y}|)(\vec{x}-\vec{y}) \cdot \vec{n}_y}{4\pi|\vec{x}-\vec{y}|^3} \end{aligned}$$

and \vec{n}_y is the surface normal on the point \vec{y} .

Numerical aspects of the formulation has been historically studied and discussed. Both equations gives rise to dense and nonsymmetrical coefficient matrices with numerous singular and hypersingular surface integrals for a linear system to be solved numerically by a conjugate gradient-based iterative solver (GMRES) using ASMS instead of linear model [24] and the matrix-vector product is computed using kernel-independent fast multipole method (KIFMM) developed by L. Ying [22].

After PB electrostatic potential is computed, we compute PB electrostatic solvation energy as follows,

$$G_{pol} = \frac{1}{2} \sum_{k=1}^{n_c} q_k \phi_{rf}(\vec{x}_k),$$

where the reaction field electrostatic potential $\phi_{rf}(\vec{x}) = \phi_{sol}(\vec{x}) - \phi_{vac}(\vec{x})$ is computed at the atomic center [17].

The electrostatic force formulation F_{pol} is composed of three terms; the reaction field force F_{RF} , the dielectric boundary force F_{DB} , and the ionic boundary force F_{IB} [14]. Each term is described in the following formulations and is computed using numerical quadrature of integrals over ASMS [6].

$$\begin{aligned} F_{pol} &= F_{RF} + F_{DB} + F_{IB}, \\ F_{RF} &= - \int_{\Omega} \rho_c(\vec{x}) E(\vec{x}) d\vec{x} = -4\pi \sum_{k=1}^{n_c} q_k E(\vec{x}_k), \\ F_{DB} &= \int_{\Omega} \frac{1}{2} E(\vec{x})^2 \nabla \epsilon(\vec{x}) d\vec{x}, \\ F_{IB} &= \int_{\Omega} k_B T \sum_i [c_i (e^{-z_i \phi(\vec{x})/k_B T} - 1) \nabla \lambda(\vec{x})] d\vec{x}, \end{aligned}$$

where $\rho_c(\vec{x}) = 4\pi \sum_{k=1}^{n_c} q_k \delta(\vec{x} - \vec{x}_k)$, the electric field $E(\vec{x}) = -\nabla \phi(\vec{x}) = (-\frac{\partial}{\partial x} \phi(\vec{x}), -\frac{\partial}{\partial y} \phi(\vec{x}), -\frac{\partial}{\partial z} \phi(\vec{x}))$ is the gradient of the electrostatic potential and $\lambda(\vec{x})$ is ionic boundary function which is 1 outside Γ and 0 inside Γ .

4. EXPERIMENTAL RESULTS

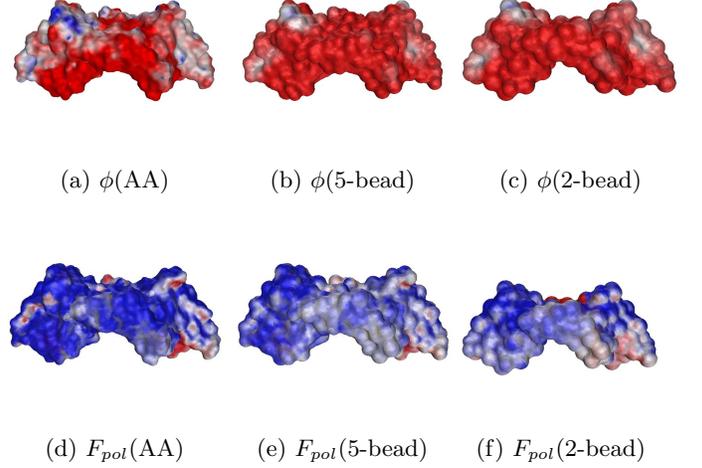


Figure 4: The PB potential, and the inner product of unit normal and PB forces on hierarchical molecular surfaces for a protein (PDB id: 1BJ1). The visual difference of PB potential between AA model (a) and both CG models (b),(c) is small where the color goes from red ($\phi \geq -3.8 k_b T/e_c$) to blue ($\phi \leq +3.8 k_b T/e_c$). The visual difference of PB forces between AA (d) and 5-bead CG models (e) is small but not 2-bead CG model (f) where the color goes from blue ($F_{pol} \cdot \vec{n} \geq 7.6 \text{ kcal/mol} \cdot \text{\AA}$) to red ($F_{pol} \cdot \vec{n} \leq -7.6 \text{ kcal/mol} \cdot \text{\AA}$).

We gathered a set of proteins from RCSB protein data bank (PDB) for evaluating our solution. Here, we construct three different levels of molecular model, including AA model, 5-bead CG model (5 beads per residue; three for side chain and two for backbone), and 2-bead CG model (2 beads per residue; one for side chain and one for backbone), for these proteins. These molecular models are in hierarchy. Based on these different molecular models, we generate their molecular surfaces and compute their PB energy, potential and forces using our PB solver "PB-CFMM" which can be downloaded from our website <http://cvcweb.ices.utexas.edu>. We set the temperature T to be 298.15 K , the interior and exterior dielectric constants ϵ_i and ϵ_e to be 1 and 80 and the ion concentration ι to be 0.05 M . All experiments are done on a linux machine with Dual Core AMD Opteron processor 280 with 4 GB memory.

First, we compute electrostatic potential of proteins using AA, 5-bead and 2-bead CG models. We also take one of those proteins (PDB id: 1CGI) as an example to see the details of PB electrostatic potential results. Figure 5 (a) shows that the reaction field electrostatic potential of each CG beads of the protein computed using the 5-bead CG

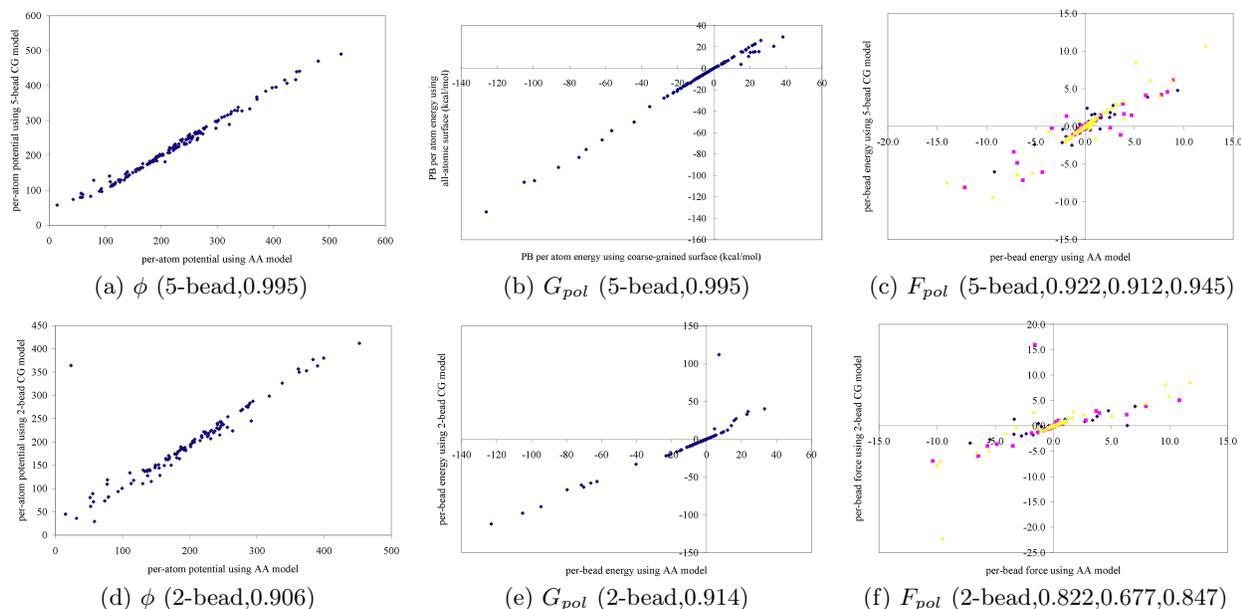


Figure 5: The relation of per-bead PB potential (k_bT/e_c), energy ($kcal/mol$) and forces ($kcal/mol \cdot \text{\AA}$) between AA molecular model and CG molecular models for the protein (PDB id: 1CGI); (a) and (d) are the relations between PB potential computed using AA model and 5-bead and 2-bead CG models; (b) and (e) are the relations between PB energy computed using AA model and 5-bead and 2-bead CG models; (c) and (f) are the relations between PB forces computed using AA model and 5-bead and 2-bead CG models where blue, pink, yellow dots indicate x, y, z-dimensional values respectively of solvation forces.

molecular model is highly related to that computed using the AA molecular model. The correlation between them is 0.9950. The same experiment is applied for 2-bead CG model and its correlation is 0.9059 in Figure 5 (c).

With a good approximation of the CG molecular surface, BEM can provide highly accurate surface electrostatic potential. In Figure 4 (a), (b) and (c), the color of the surface represents electrostatic potential on the molecular surface of a protein (PDB id: 1BJ1), going from red ($-3.8 k_bT/e_c$) to blue ($+3.8 k_bT/e_c$) and white is neutral potential. From our observation, the distribution of electrostatic potential on the molecular surfaces in different levels are highly related. The parts of the AA molecular surface with highly positive or negative electrostatic potential will hold in CG cases.

The relation of PB electrostatic energy between AA model and different levels of CG models of proteins are also shown in Table 1. The correlation of PB electrostatic energy between 5-bead CG and AA models is up to 0.9992. This result indicates that the evaluation accuracy of the PB electrostatic energy using 5-bead CG model is consistently satisfactory, while that of using coarser 2-bead CG model is not. The errors of PB energy computation of 2-bead CG model are from 0.016 to 0.385 in this set of proteins.

Now, we take one of those proteins (PDB id: 1CGI) as an example to see the details of PB energetic results. Figure 5 (b) shows that the electrostatic solvation free energy of each CG bead of the protein computed using 5-bead CG model is highly related to the energy computed using AA model. The correlation of per-bead electrostatic free energy between AA model and CG model is 0.9950. Figure 5 (e) shows the same comparison between 2-bead CG model and AA model. The correlation is 0.9135. Here, 5-bead CG model performs

better than 2-bead CG model.

In Figure 5 (c), we show the relation of per-bead electrostatic forces computed by AA model and 5-bead CG model for a protein (PDB id: 1CGI). Blue, pink and yellow dots on the chart are the value of forces at x, y, z dimensions. The correlations between them are 0.9223, 0.9117 and 0.9448 at x, y, z dimensions. The same experiment is done for 2-bead CG model and shown in Figure 5 (f). The correlations at x, y, z dimensions are 0.8224, 0.6773 and 0.8472. 5-bead CG model is a reasonably good approximate model for electrostatic force computation but doesn't give such high correlation as the per-atom energy in Figure 5 (b). The PB force computation looks more sensitive to the resolution of molecular models than the PB energy or potential computation.

In Figure 4 (d), (e) and (f), we show the electrostatic forces on the molecular surface of one protein (PDB id: 1BJ1). The color of the molecular surface represents the inner product of the electrostatic force and the surface normal at the surface point. The outward force gives the positive inner product and negative otherwise. The color goes from blue ($\geq 7.6 kcal/mol \cdot \text{\AA}$) to red ($\leq -7.6 kcal/mol \cdot \text{\AA}$). The distribution of inward and outward forces is almost the same in AA and 5-bead CG models but not 2-bead CG models.

All these results provide very good tradeoffs between accuracy and efficiency. For example, 5-bead CG model with radius and charge optimization is shown to be a great approximation for PB energetics.

5. CONCLUSION

In this paper, we introduce a complete framework from producing different hierarchical-level molecular models, generating molecular surface to computing electrostatic proper-

PDB id	# atoms/beads	# A-patches	G_{pol}	relative error of G_{pol}	PB-CFMM time
1AK4 (l)	2260/401/290	12829/10294/9438	-1638.17/-1560.44/-1790.65	-/0.047/0.093	682.85/446.40/417.88
1AK4 (r)	2503/440/330	11730/9562/8568	-1907.00/-1862.45/-1689.96	-/0.023/0.114	661.41/535.04/470.18
1AVX	2662/477/344	12468/10341/9547	-3349.88/-3209.52/-3220.84	-/0.041/0.039	614.96/437.93/402.25
1AY7	2875/517/370	13493/10602/9694	-3657.04/-3591.34/-3061.09	-/0.018/0.163	973.38/454.36/424.88
1AY7 (l)	1434/251/178	9049/7106/6461	-1601.60/-1576.64/-1576.64	-/0.016/0.016	438.89/303.32/204.92
1AY7 (r)	1441/266/192	10594/8548/5208	-1768.38/-1721.51/-1267.87	-/0.027/0.283	630.15/435.05/386.96
1B6C	1663/290/214	10062/8051/7411	-1342.30/-1300.59/-896.64	-/0.031/0.332	484.07/425.44/326.63
1BJ1	2986/544/376	13022/11167/10230	-3812.96/-3712.58/-3031.71	-/0.026/0.205	587.75/586.55/501.30
1BUH	1190/205/140	8996/6983/6188	-1456.83/-1510.51/-1494.61	-/0.037/0.026	675.80/438.34/244.29
1CGI	852/157/112	7277/6312/5497	-971.77/-931.69/-597.23	-/0.041/0.385	374.49/258.66/243.36

Table 1: The experimental results for a set of proteins. In each column, from left to right are the results of the AA model/the 5-bead CG model/the 2-bead CG model; column 1 is PDB id of the protein where (l) and (r) indicate the ligand and receptor of the complex protein; column 2 is number of atoms of AA model and number of beads of 5-bead CG and 2-bead CG model; column 3 is the number of A-patches used in the molecular surfaces for AA, 5-bead CG and 2-bead CG models; column 4 is the electrostatic solvation energy G_{pol} (kcal/mol); column 5 is the relative error of electrostatic free energy computed using 5-bead CG and 2-bead CG models relative to G_{pol} for the AA model; column 6 is computation time in seconds of our curved boundary element PB method (PB-CFMM).

ties. At each step of this framework, we identify the sources of errors and produce and analyze the empirical results of PB electrostatic energy, potential and force by adjusting resolution of the molecular model.

This paper is an initial study to realize the opportunity to control the resolution of the molecular model with acceptable error tolerance in molecular simulation. It is always difficult to investigate an optimized tradeoff between computational cost and accuracy. We believe that it is still possible to find out a more reliable way to control the resolution of the molecular model according to different computational purposes. We are going to develop a following technique to adjust the resolution at each different parts of a molecule for electrostatic force computation.

6. ACKNOWLEDGEMENTS

This research was supported in part by NSF grant CNS-0540033 and NIH contracts R01-EB00487, R01-GM074258, R01-GM07308. Sincere thanks to members of CVC for their help in maintaining the software environment in particular for the use of TexMol (<http://cvcweb.ices.utexas.edu/software>). Authors also thank to Prof. Lexing Ying of the Dept. of Mathematics for the use of KiFMM (kernel independent fast multipole method). Curved boundary element PB implementation based on KiFMM is called PB-CFMM and is callable from TexMol.

7. REFERENCES

- [1] A.H.Juffer, E. Botta, B. van Keulen, A. van der Ploeg, and H. Berensen. *J Chemical Physics*, 97:144–171, 1991.
- [2] A. Arkhipov, P. Freddolino, K. Imada, K. Namba, and K. Schulten. *Biophys. J.*, 91:4589–4597, 2006.
- [3] I. Bahar and R. Jernigan. *J. Mol. Biol.*, 266:195–214, 1997.
- [4] C. Bajaj and S.-C. A. Chen. *SIAM Journal on Scientific Computing*, submitted, 2009.
- [5] C. Bajaj, G. Xu, and Q. Zhang. *Journal of Computer Science and Technology*, 23(6):1026–1036, 2008.
- [6] C. Bajaj and W. Zhao. *SIAM Journal on Scientific Computing*, under review, 2009.
- [7] C. Bajaj, W. Zhao, S.-C. A. Chen, G. Xu, and Q. Zhang. *ICES Report*, 2009.
- [8] N. Basdevant, D. Borgis, and T. Ha-Duong. *J. Phys. Chem. B*, 111:9390–9399, 2007.
- [9] R. Bharadwaj, A. Windemuth, S. Sridharan, B. Honig, and A. Nicholls. *J. Comput Chem.*, 16:898, 1995.
- [10] A. J. Bordner and G. A. Huber. *J. Comput Chem.*, 24:353–367, 2003.
- [11] A. H. Boschitsch, M. O. Fenley, and H.-X. Zhou. *J. Phys. Chem.*, 106:2741–2754, 2002.
- [12] N.-V. Buchete, J. Straub, and D. Thirumalai. *Protein Sci.*, 13:862–874, 2004.
- [13] M. Connolly. *J. Appl. Cryst.*, 16:548–558, 1983.
- [14] M. K. Gilson, M. E. Davis, B. A. Luty, and J. McCammon. *J. Phys. Chem.*, 97:3591–600, 1993.
- [15] B. Z. Lu and J. A. McCammon. *J. Chem. Theory Comput.*, 3:1134–1142, 2007.
- [16] D. Marquardt. *J. Soc. Indust. Appl. Math.*, 11:431–441, 1963.
- [17] K. A. Sharp and B. Honig. *J. Phys. Chem.*, 94:7684–7692, 1990.
- [18] A. Smith and C. Hall. *Proteins: Structure, function, and genetics*, 44:344–360, 2001.
- [19] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.
- [20] Y. Ueeda, H. Taketomi, and N. Gō. *Biopolymers*, 17:1531–1548, 1978.
- [21] G. Xu and Q. Zhang. *Mathematica Numerica Sinica*, 28(4):337–356, 2006.
- [22] L. Ying, G. Biros, and D. Zorin. *Journal of Computational Physics*, 196(2):591–626, 2004.
- [23] R. Zauhar and R. Morgan. *J. Mol. Biol.*, 186:815–820, 1985.
- [24] W. Zhao, G. Xu, and C. Bajaj. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2009.