Improvements to the APBS biomolecular solvation software suite

Elizabeth Jurrus,[†] Keith Star,[†] Juan Brandi,[†] Lisa E. Felberg,[‡] Jiahua Chen,[¶] Leighton Wilson,[§] Karina Liles,[‖] Kyle Monson,[†] Dave Engel,[⊥] Minju Chun,[†] Peter Li,[†] David Brookes,[‖] Weihua Chen,[‖] ?? Todd Dolinsky ??,[‖] ?? Robert Krasny ??,[§] Teresa Head-Gordon,[#] ?? Robert Konecny ??,[@] ?? Jens Erik Nielsen ??,[△] ?? Marilyn Gunner ??,[∇] ?? Michael Holst ??,[@] ?? J. Andrew McCammon ??,[@] and Nathan A. Baker[*,⊥,††]

[†]*Computational and Statistical Analytics Division; Pacific Northwest National Laboratory; Richland, WA 99352, USA*

[‡]*University of California Berkeley*

[¶]*Southern Methodists University*

[§]*University of Michigan*

[‖]...

[⊥]*Advanced Computing, Mathematics, and Data Division; Pacific Northwest National Laboratory; Richland, WA 99352, USA*

[#]*University of California Berkeley;*

[@]*University of California San Diego*

[△]*Novozymes*

[∇]*City University of New York*

[††]*Division of Applied Mathematics; Brown University; Providence, RI 02912, USA*

E-mail: nathan.baker@pnnl.gov

Phone: +1-509-375-3997

Abstract

The Adaptive Poisson-Boltzmann Solver (APBS) software was developed to solve the equations of continuum electrostatics for large biomolecular assemblages.[1] The understanding of electrostatic interactions is essential for the study of biomolecular processes, which is used for many applications, including drug discovery. APBS addresses three key technology challenges for understanding solvation and electrostatics in biomedical applications: accurate and efficient theories and models for biomolecular solvation and electrostatics, robust and scalable software for applying those theories to biomolecular systems, and mechanisms for sharing and analyzing biomolecular electrostatics data in the scientific community. To address new research applications and advancing computational capabilities, APBS and and its suite of accompanying software, have been continually updated. In this paper, we discuss the models and capabilites that have recently been implemented within the APBS software package, include: a Poisson-Boltzmann analytical and a semi-analytical solver, an optimized boundary element solver, a geometry-based geometric flow solvation model, a graph theory based algorithm for determining pK_a values, and an improved web-based visualization tool for viewing electrostatics.

⋆⋆ **author list and order subject to change!!!!** ⋆⋆

Contents

1 Introduction

1.1 Why solvation matters

Robust models of electrostatic interactions are key for the understanding of intermolecular interactions, such as the solvation properites of biomolecules and the effects of solvation at the biomolecule level.[2–5] The structure and binding of biomolecules are controlled by these intermolecular interactions. Such interactions include several components: contributions from linear, angular, and torsional forces in covalent bonds, van der Waals forces, as well as electrostatics. Therefore, it is essential to understand these interactions for any biomedical problem studied at the molecular scale.

1.2 Brief review of solvation model types

In the modeling of electrostatics and solvation, models tend to utilize either explicit or implicit methods.[4,6,7] Explicit methods treat the solvent in atomic detail where implicit methods generally replace the explicit solvent with a dielectric continuum. Because explicit methods solve systems directly, they tend to offer the highest levels of detail. However, they generally require extensive sampling to converge properties of interest. Implicit methods, on the other hand, trade detail and some accuracy for methods that tend to eliminate the sampling, utilizing pre-equilibration. The focus of this paper and the APBS software is on implicit methods. The discussion of explicit models has been published elsewhere.[4]

A variety of implicit solvent methods have been developed for the description of polar solvation,[4] with Poisson-Boltzmann (PB) methods being among the most popular of these. The PB equation[8–10] provides a global solution for the electrostatic potential (ϕ) and field within and around a biomolecule, making them uniquely suited to visualization and other structural analyses, diffusion simulations, and a number of other methods that require global electrostatic properties. The coefficients of this equation are directly related to the molecular structure of the system under consideration. The PB theory is approximate and, as a result, has several well-known limitations

which can affect its accuracy, particularly for strongly charged systems or high salt concentrations.[8,11] Despite these limitations, PB methods are still very important for biomolecular structural analysis, modeling, and simulation.

1.3 Workflows for biomolecular solvation and electrostatics modeling

In order to run the APBS software, a PQR file is needed. A PQR file can be created by modifying a protein data bank (PDB) file where the temperature and occupancy columns are replaced by the per-atom charge (Q) and radius (R). In order to help with this conversion (PDB to PQR), the PDB2PQR software was developed.[12] The user begins an experiment by entering a protein data bank ID (e.g., 1FAS) into the PDB2PQR software. PDB2PQR automatically sets up, executes, and optimizes the structure for Poisson-Boltzmann electrostatics calculations, outputting a PQR file. Besides this conversion, PDB2PQR can also add a limited number of missing heavy atoms to biomolecular structures, determine side-chain pK_a values, place missing hydrogens, optimize the protein for favorable hydrogen bonding, and assign charge and radius parameters from a variety of force fields.[12,13] The flow of this analysis is shown in Figure 1. Once complete, the electrostatics are calculated using the APBS solver and a cube file is created containing the visualization parameters for our molecular rendering tool.

1.4 Software for biomolecular solvation and electrostatics

APBS and PDB2PQR are software packages designed to help analyze the solvation properties of small and macro-molecules such as proteins, nucleic acids, and other complex systems. APBS is unique software which solves the equations of continuum electrostatics for large biomolecular assemblages. This software was designed from the ground up using modern design principles to ensure its ability to interface with other computational packages and evolve as methods and applications change over time. The PDB2PQR software automates many of the common tasks of preparing structures for continuum solvation calculations as well as many other types of biomolecular structure modeling, analysis, and simulation.

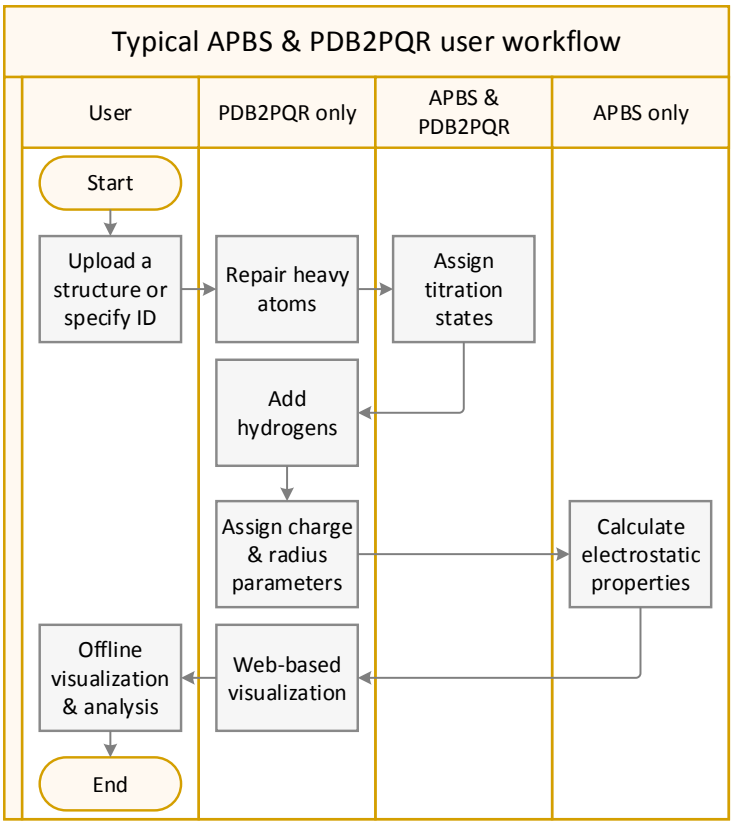


Figure 1: Flow diagram for the APBS-PDB2PQR process.

The capabilities in these software have recently been enhanced with the addition of more powerful and efficient methods. Specifically, the following methods have been implemented:

- A Poisson-Boltzmann analytical solver and a semi-analytical solver
- Optimized boundary element solver for the linearized Poisson-Boltzmann equation
- A geometric flow solver
- Graph cut algorithm for determining pK_a values
- Improved web-based visualization tool for viewing electrostatics

These new additions to the software will allow more diverse analyses while also focusing on computational efficiency.

Along with APBS, there exist many different flavors of models and software to solve the electrostatics problem. The most significant (based on user base and citations) of these include CHARMM,[14] AMBER,[15] DelPhi,[16] Jaguar,[17] Zap,[18] and MIBPB.[19] The CHARMM and AMBER packages are multipurpose software designed for a wide range of modeling tasks. DelPhi and MIBPB are the most similar packages to APBS. Like APBS, DelPhi and MIBPB are command-line programs, also available as web servers, focused on biomolecular electrostatics calculations. DelPhi has a significant and established user base but is only available (for free) under a closed-source license that prohibits modifications. MIBPB is a newer software package also available for free under a standard open-source license similar to APBS. Both packages currently have better performance than APBS: DelPhi is slightly faster than APBS for small-to-moderate size systems, while MIBPB is both faster and more accurate than APBS for most molecular systems.

2 Functionality

Before discussing the new capabilities within PDB2PQR and APBS, we first discuss the background and previous capabilities of our software system. The full functionality of the APBS-PDB2PQR system was illustrated in Figure 1. In this section, we discuss the individual pieces that

make up the system. Previously available methods are briefly discussed, while the new methods that have been implemented are discussed in detail to provide readers and previous users enough detail to select appropriate methods when using the software.

2.1 Repairing missing atoms

Within PDB2PQR, the PDB file is examined to see if there are missing heavy (non-hydrogen) atoms. Missing heavy atoms can be rebuilt using standard amino acid topologies in conjunction with existing atomic coordinates to determine new positions for the missing heavy atoms. To assure that the new atoms are not within the Van der Waals radii of other nearby atoms, there is a *debump* option. For this option, the sidechain χ angles are varyied until the steric conflict is resolved.

2.2 Optimizing titration states

The pK_a values are important determinants of biomolecular (particularly enzymatic) function and can be used to assess functional activity and identify active sites. The APBS-PDB2PQR system contains several methods for this analysis.

2.2.1 Empirical methods

An empirical model (PROPKA[20]) which uses a heuristic method to compute the pK_a perturbations due to desolvation, hydrogen bonding and charge-charge interactions is included in PDB2PQR. For this model, an empirical relationship is used for structure-based protein pK_a prediction and rationalization, dealing with an empirical relationships between protein pK_a shifts and structures. It has been shown that this empirical method is able to accurately predict most protein pK_a values in a matter of seconds.[21]

2.2.2 Poisson-based methods

PDB2PQR also contains a Metropolis Monte Carlo method for calculating titration curves and pK_a values (PDB2PKA). In this method, a random titration state is selected from a random residue and

the energy difference from the previous time step is calculated. Computational performance and sampling issues can be a major problem with Monte Carlo methods when searching over the 2^N titration states.

2.2.3 Graph Cut

A new polynomial time algorithm for the optimization of discrete states in macromolecular systems[22] has been implemented into PDB2PKA. This method utilizes graph theory in order to transform the interaction energy graph into a flow network. The interaction energy graph is a graph in which the vertices represent amino acids and the edges represent the interactions of the amino acids, which are weighted by their respective energies. Because this algorithm is much faster than other implemented models, it may be possible to discover the ionization state of much larger proteins.

In this approach, we assign a titration state to a rigid protein where the backbone and amino acid location are fixed. We also assign specific titration states and not pK_a or titration probabilities. The goal in the titration state assignment is to find a state P in the set of all protonation states that minimizes the protein's energy at a given pH, volume, and temperature (T). This free energy can be approximated as a pairwise-decomposable function between titration states as shown in Eq. 1.

$$G(P) = \sum_{i=1}^{N} \gamma_i \ln(10) kT \left(\mathrm{pH} - pK_{a,i}^{int}\right) + \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j U^{int}\left(P_i, P_j\right) \tag{1}$$

where γ_i is based on the charge of the amino acid: 1 if charged otherwise 0, $pK_{a,i}^{int}$ is the intrinsic pK_a of the amino acid i and $U^{int}(P_i, P_j)$ is the interaction energy between amino acids i and j.

Since it is computationally intensive to compute the fractional protein occupancies, this algorithm follows a basic two-state linkage analysis. In this approach, $G_p(i)$ is defined as the change in energy between the states where residue i is protonated and where it is deprotonated. Thus, the fraction of protonated state is given in Eq. 2.

$$\theta_p(i) = \frac{e^{-\beta G_p(i)}}{1 + a_H e^{-\beta G_p(i)}} \tag{2}$$

where $\beta = (RT)^{-1}$, R is the gas constant and a_H is the activity of the proton. For the case of histidine (HIS), our method relies on finding a minimum energy state for the protein at any given pH value, changing the state of each residue and recording the change in energy.

To apply graph theory to this energy minimization problem, an energy graph is created which is a weighted graph that holds all of the information from the energy function. For each titration state (protonated and deprotonated for the non-HIS) of an amino acid, a vertex will be assigned. Each vertex will be weighted by the energy contribution from the difference between the amino acid intrinsic pK_a and the current pH value, as well as the $\gamma_i^2 U^{int}(P_i, P_j)$ background and desolvation energies, which are combined into $E_i(P_i)$ in Eq. 3.

$$E(L) = \sum_{i=1}^{N} E_i(L_i) + \sum_{(i,j) \in \mathscr{E}} \sum E_{ij}(L_i, L_j) \tag{3}$$

where $L = \langle L_1, L_2, ..., L_N \rangle$, L_i is the label of vertex i, and $\mathscr{E}$ is the list of vertex interactions, such that vertices i and j are said to interact if $(i, j) \in \mathscr{E}$.

In our method, the selection of vertices and edges in the energy graph are represented through a graph-cut in the energy flow network. This graph-cut, defines a given state of the protein by assigning all vertices associated with the state into one set and all others into a second set. The value from this cut can be shown to be exactly the energy of the associated state of the system.

Not all titration sites can be represented by two states (e.g., HIS). In this case, each state will be analyzed separately and then combined. For instance, HIS can be represented by two neutral tautomers and one positively charged state. Each one of the two neutral tautomers are analyzed. If only one of these states are protonated, than it represents the appropriate minimum energy state. If both states are protonated, then HIS is fully protonated and is excluded and must be analyzed using other methods.

Titration probabilities should be calculated as thermodynamic averages over the entire ensem-

ble of states. However, this graph method solves and optimization problem, and thus could lead to errors. Also, residues that violate the modularity condition for titration state interactions do not get labeled and must then be calculated using other methods. This method was tested against the ensemble averaging method of APBS (from PDB2PKA). Despite the limitations, the results showed good agreement (with in 5%) for the majority of proteins tested.

2.3 Add missing hydrogens

Monte Carlo sampling and optimization are used to position hydrogen atoms and assign protonation states to optimize the global hydrogen-bonding network in the structure. By default, newly added hydrogen atoms are then checked for steric conflicts via the debumping procedure discussed previously.

2.4 Assigning charge and radius parameters

The atomic charges and radii are then assigned based on the chosen force field. They are assigned based on translating the atom and residue names found in the fouce field to those found in the input structure file.

2.5 Solving the Poisson-Boltzmann equation with APBS

APBS has traditionally used volume grid-based methods for solving the PB equation. One of these methods is an adaptive finite element solver. This solver uses a "solve-estimate-refine" cycle. Specifically, starting from an initial mesh, it performs the following iteration:

- solve the problem with the current mesh,
- estimate the error in the solution,
- adaptively refine the mesh to reduce the error, and
- repeat until a global error tolerance is reached.

Another solution method within APBS is a multi-grid finite difference method. This method starts by solving the equation on a coarse grid (i.e., few grid points) with large dimensions (i.e., grid lengths). The solution on this coarse grid is then used to set the Dirichlet boundary condition values for a smaller problem domain, and therefore a finer grid surrounding the region of interest. The finer grid spacing in the smaller problem domain often provides greater accuracy in the solution.

These methods use fast iterative techniques that solve a sparse linear system. However, issues have been identified which effect the results and performance of the models. The following issues were identified by Geng and Krasny[23] when solving the PB equation utilizing numerical methods:

- the memory requirements for a three-dimensional grid can be prohibitively large,
- the geometric details of the molecular surface may be obscured on a regular grid,
- the singular atomic charges are smoothed by interpolation onto the grid,
- the interface conditions may not be rigorously enforced on the molecular surface., and
- the far-field boundary condition is often satisfied only approximately on a truncated domain.

To address these issues, new methods have been added to APBS to improve the flexibility for more diverse applications, address issues identified with grid-based methods, increase performance, and to expand the analysis capabilities of APBS. These methods are described in the following sections. For each method, we briefly discuss the algorithms that were implemented, how a user may use them, and show example results for several of the methods.

2.5.1 Poisson-Boltzmann Analytical and Semi-Analytical Methods

As mentioned above, numerical solution methods tend to be computationally intensive. Therefore, analytical methods have been included in the APBS system, which is beneficial for a rapid and accurate solution of the PB equation.

Poisson-Boltzmann Analytical Method The Poisson-Boltzmann Analytical Method (PB-AM), was developed by Lotan and Head-Gordon in 2006.[24] It presents a fully analytical solution to the

linearized PB equation for multiple macromolecules in a screened environment. Each molecule in the system is coarse-grained and represented as a single low dielectric spherical cavity. This domain enables the use of the multipole expansion to represent charge-charge interactions and higher order cavity polarization effects. The solution to this model is represented as a system of linear equations as follows:

$$\mathbf{A} = \Gamma \cdot (\Delta \cdot T \cdot \mathbf{A} + \mathbf{E}) \tag{4}$$

Where $\mathbf{A}$ represents a vector of the effective multipole expansion of the charge distributions of each molecule, $\mathbf{E}$ is a vector of the fixed charge distribution of all molecules, Γ is a dielectric boundary-crossing operator, Δ is a cavity polarization operator and T is an operator that transforms the multipole expansion from the global (lab) coordinates to a local coordinate frame. Once the unknown $\mathbf{A}$'s are solved using the Gauss-Sieidel iterative method, many physical properties, such as interaction energies, forces and torques can be computed. The interaction energy for molecule i, ($\Omega^{(i)}$) is given in Eq. 5.

$$\Omega^{(i)} = \frac{1}{\varepsilon_s} \left\langle \sum_{j \neq i}^{N} T \cdot A^{(j)}, A^{(i)} \right\rangle \tag{5}$$

where ε_s is the dielectric constant of the solvent and $\langle M, N \rangle$ denotes the inner product. When energy is computed, forces follow as:

$$\mathbf{F}^{(i)} = \nabla_i \Omega^{(i)} = \frac{1}{\varepsilon_s} [\langle \nabla_i T \cdot A^{(i)}, A^{(i)} \rangle + \langle T \cdot A^{(i)}, \nabla_i A^{(i)} \rangle] \tag{6}$$

By definition, the torque on a charge in the molecule is the cross product of its position relative to the center of mass of the molecule with the force it experiences. The total torque on the molecule is a linear combination of the torque on all charges of the molecule, as illustrated in Eq. 7.

$$\tau^{(i)} = \frac{1}{\varepsilon_s} \left[{}^{x}H^{(i)}, {}^{y}H^{(i)}, {}^{z}H^{(i)} \right] \times \left[\nabla_i L^{(i)} \right] \tag{7}$$

where ${}^{\alpha}H^{(i)}_{n,m} = \sum_{j=1}^{M_i} \alpha_j^{(i)} \gamma_n^{(i)} q_j^{(i)} (\rho_j^{(i)})^n Y_{n,m}(\vartheta_j^{(i)}, \varphi_j^{(i)})$, $\alpha = x, y, z$, is a coefficient vector for each of the charges in the molecule, M_i is the number of charges in molecule i, $q_J^{(i)}$ is the magnitude of the

j^{th} charge, and $p_j^{(i)} = \left[\rho_j^{(i)}, \vartheta_j^{(i)}, \varphi_j^{(i)}\right]$ is its position in spherical coordinates. For more details on the PB-AM derivation, see Lotan, Head-Gordon.[24]

Poisson Boltzmann Semi-Analytical Method An extension of the fully analytical solution, the Poisson Boltzmann Semi-Analytical method (PB-SAM) incorporates the use of boundary integrals into its formalism to represent a complex molecular domain as a collection of overlapping low dielectric spherical cavities.[25] It presents a semi-analytical solution to the linearized PB equation for multiple macromolecules in a screened environment. The semi-analytical method provides a better representation of the molecular boundary when compared to PB-AM, while maintaining computational efficiency.

Because it is fully analytical, PB-AM can be used for model validation as well as for representing systems that are relatively spherical in nature, such as globular proteins and colloids. PB-SAM, on the other hand has a much more detailed representation of the molecular surface and can therefore be used for many systems that other APBS (numerical) methods are currently used for. Additionally, both methods have been implemented in a Brownian dynamics (BD) scheme[26] and can perform dynamics with full mutual polarization, a novel feature for APBS.

The derivation details of PB-SAM have been reported previously,[25,27] with the main points being summarized in this section. The electrostatic potential (ϕ_r) of the system at any point r is governed by the linearized form of the PB equation:

$$\nabla \cdot [\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \varepsilon(\mathbf{r})\kappa^2\phi(\mathbf{r}) = 4\pi\rho(\mathbf{r}) \tag{8}$$

where κ is the inverse Debye length. For the case of spherical cavities, we can solve Eq. 8 by dividing the system into inner sphere and outer sphere regions, and enforcing a set of boundary conditions that stipulate the continuity of the electrostatic potential and the electrostatic field at the surface of each sphere. The electrostatic potential outside molecule (I) is described by:

$$\phi_{out}^{(i)}(\mathbf{r}) = \sum_{I=1}^{N_{mol}} \left(4\pi \int_{d\Omega^{(I)}} \frac{e^{-\kappa|r-r'|}}{|r-r'|} h^{(I)}(r')dr' \right) \tag{9}$$

where $h(r)$ is an effective surface charge that can be transformed into the unknown multipole expansion $H^{(I,k)}$ with inside molecule I and sphere k. In a similar manner, the interior potential is given as

$$\phi_{in}^{(i)}(\mathbf{r}) = \sum_{\alpha=1}^{N_C^{(I)}} \frac{1}{|r - r_\alpha^{(I)}|} \cdot \frac{q_\alpha^{(I)}}{\varepsilon_{in}} + \frac{1}{4\pi} \int_{d\Omega^{(I)}} \frac{1}{|r - r'|} f^{(I)}(r') dr' \tag{10}$$

where $N_C^{(I)}$ is the number of charges in molecule I, q_α is the magnitude of the α-th charge, $r_\alpha^{(I)} = \left[\rho_\alpha^{(I)}, \theta_\alpha^{(I)}, \phi_\alpha^{(I)}\right]$ is its position in spherical coordinates, and $f(r)$ is a reactive surface charge that can be transformed into the unknown multipole expansion $F^{(I,k)}$. The reactive multipole and the effective multipole, $H^{(I,k)}$, are given as:

$$F_{n,m}^{(I,k)} \equiv \frac{1}{4\pi} \int_{d\Omega^{(I,k)}} f^{(I,k)}(r') \left(\frac{a^{(I,k)}}{r'}\right)^{n+1} \overline{Y_{n,m}^{(I,k)}}(\theta', \phi') dr' \tag{11}$$

$$H_{n,m}^{(I,k)} \equiv \frac{1}{4\pi} \int_{d\Omega^{(I,k)}} h^{(I,k)}(r') \left(\frac{r'}{a^{(I,k)}}\right)^{n} \hat{i}_n(\kappa r') \overline{Y_{n,m}^{(I,k)}}(\theta', \phi') dr' \tag{12}$$

where $Y_{n,m}$ is the spherical harmonics, $\overline{Y_{n,m}}$ is the complete conjugate, and $a^{(I,k)}$ is the radius of sphere k of molecule I. These multipole expansions can be iteratively solved using:

$$F_{n,m}^{(I,k)} = \langle I_{E,n,m}^{(I,k)}, WF^{(I,k)} \rangle \tag{13}$$

$$H_{n,m}^{(I,k)} = \langle I_{E,n,m}^{(I,k)}, WH^{(I,k)} \rangle \tag{14}$$

where $WF^{(I,k)}$ and $WH^{(I,k)}$ are scaled multipoles computed from fixed charges and polarization charges from other spheres. $I_{E,n,m}^{(I,k)}$ is a matrix of the surface integrals over the exposed surface:

$$I_{E,n,m}^{(I,k)} \equiv \frac{1}{4\pi} \int_{\phi_E} \int_{\theta_E} Y_{l,s}^{(I,k)}(\theta', \phi') \overline{Y_{n,m}^{(I,k)}}(\theta', \phi') sin\theta' d\theta' d\phi' \tag{15}$$

Using the above formalism, physical properties of the system, such as interaction energy, forces and torques may be computed. The interaction energy of each molecule, ($\Omega^{(i)}$), is the product of the molecule's total charge distribution (from fixed and polarization charges) with the potential

due to external sources. This is computed as the inner product between the molecule's multipole expansion, ($H^{(I,k)}$), and the multipole expansions of the other molecules in the system, ($LHN^{(I,k)}$) as follows:

$$\Omega^{(i)} = \frac{1}{\varepsilon_s} \sum_{k}^{N_k^{(I)}} \langle LHN^{(I,k)}, H^{(I,k)} \rangle \tag{16}$$

which allows us to define the force which is computed as the gradient of the interaction energy with respect to the position of the center of molecule I:

$$F^{(I)} = -\nabla\Omega^{(I)} = -\frac{1}{\varepsilon_s} \sum_{k}^{N_k^{(I)}} f_{I,k} = -\frac{1}{\varepsilon_s} \sum_{k}^{N_k^{(I)}} (\langle \nabla LHN^{(I,k)}, H^{(I,k)} \rangle + \langle LHN^{(I,k)}, \nabla H^{(I,k)} \rangle) \tag{17}$$

As in the analytical method, the torque on a charge in the molecule is the cross product of its position relative to the center of mass of the molecule with the force it experiences. For a charge at position, P, about the center of mass $c^{(I)}$ for molecule I, the torque is given by the cross product of its position, $r_P^{(I,k)}$, with respect to the center of mass and the force on that charge, f_P. We can re-express $r_P^{(I,k)}$ as the sum of vectors from the center of molecule I to the center of sphere k ($c^{(I,k)}$) and from the center of sphere k to point P ($r_P^{(I,k)}$). The total torque on molecule I is then given by Eq. 18.

$$\tau^{(I)} = \sum_{k}^{N_k^{(I)}} c^{(I,k)} \times f_{I,k} + \sum_{k}^{N_k^{(I)}} \sum_{P \in k} r_P^{(I,k)} \times f_P \tag{18}$$

where $f_{I,k}$ is given in Eq. 17 and

$$f_P = -\frac{1}{\varepsilon_s} \sum_{k}^{N_k^{(I)}} (\langle \nabla_I LHN^{(I,k)}, H_P^{(I,k)} \rangle + \langle LHN^{(I,k)}, \nabla_I H_P^{(I,k)} \rangle) \tag{19}$$

where

$$H_{P,n,m}^{(I,k)} = h(\theta_p, \phi_p) Y_{n,m}^{(I,k)}(\theta_p, \phi_p) \tag{20}$$

$$\nabla_j H_{P,\alpha,n,m}^{(I,k)} = \left[\nabla_j h(\theta_p, \phi_p)\right]_\alpha Y_{n,m}^{(I,k)}(\theta_p, \phi_p) \tag{21}$$

where $\alpha = x, y, z$. For the derivation of the PB-SAM solver please see the previous publications[25,27]

.

Usage in APBS PB-AM and PB-SAM have been fully integrated into APBS, and is invoked using the keyword `pbam-auto` or `pbsam-auto` in the ELEC section of an APBS input file.[1] From APBS, both programs can be used to compute the electrostatic potential at any point in space, report energies, forces, and torques of a system of macromolecules, and simulate a system using a BD scheme.[26] For an example of the key simulation commands used in a run of PB-AM or PB-SAM, see Table 1. The `term` keyword, included in Table 1, allows the user to indicate conditions for the termination of each BD trajectory. This may include the contact of two molecules or the diffusion of a specified molecule beyond a specified point in space, like the conditions defined by Northrup, Allison and McCammon for the NAM simulation protocol.[28] In addition to parameters specific to each type of run (BD, electrostatic potential printing and physical calculations), the user can modify the coarseness of the molecular description by modifying the coarse-graining parameter `tolsp`, which indicates how far beyond the solvent-excluded surface the coarse-grain molecular boundary is allowed to extend. The larger the value of `tolsp`, the coarser the description of the molecule, e.g. less coarse grain spheres used to represent it. This will make computational evaluation of the system more rapid, but less accurate. Conversely, the smaller the `tolsp` value chosen, the more coarse grain spheres needed to represent the system, which means a more accurate but also more costly representation of the system. The `tolsp` parameter is generally in the range of 1 to 5 Å. The commands (keywords) not included in Table 1 are used to specify system conditions, such as temperature and salt concentration.

Examples of electrostatic potential (ESP) visualization results from PB-AM are given in Figure 2. In Figure 2 (A), 2-D cross sections of electrostatic potential surrounding two spheres of opposite charge are shown. The 3-D surface potential visualization given in Figure 2 (B) describes the potential on the surface of a coarse-grain barstar molecule (a ribonuclease inhibitor). Figure 2 (C) is an example of three-dimensional isosurface visualization in VMD.[29] The molecule system which was analyzed was the periplasmic face of the Omp32 Porin trimer, a transmembrane protein.

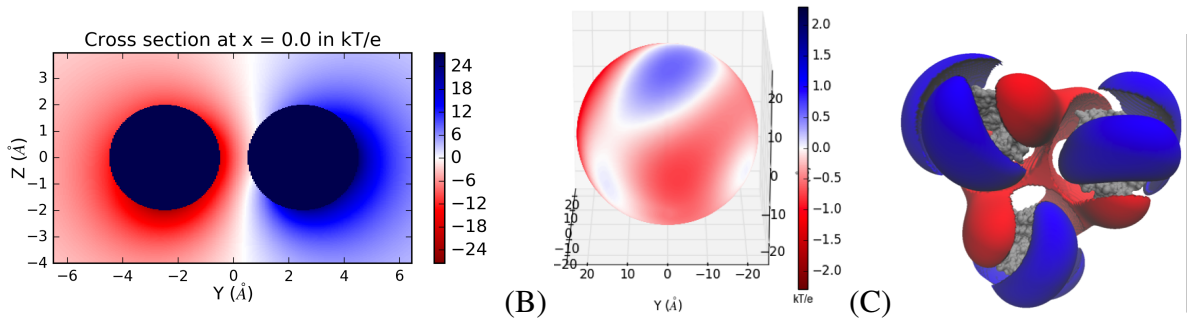
[1] www.poissonboltzmann.org/docs/apbs-overview

Table 1: Major keywords used in PB-AM and PB-SAM. For more keywords, details and examples of implementation, please see the APBS website.

Keyword	Parameters	Description
runname	<name>	Desired name to be used for outputs of each run.
pbc	<boxlength>	Size of periodic box.
runtype dynamics		Perform a Brownian Dynamics run.
ntraj	<trajct>	Number of trajectories to run.
term	<type> <val> <mols>	Attributes of a termination condition: <type> can be time, x<=, y<=, z<=, r<=, (or the >= equivalents), <val> is the value of the condition, <mols> is the molecular index that this condition applies to (time requires 0, all else, 1).
xyz	<idx> <fpath>	Molecule index and file name (path) for the xyz file. A starting configuration is needed for each trajectory for all the molecules, so there should be <trajct> xyz lines for each molecule.
PB-SAM Keywords		
msms		Will run the MSMS program for the coarse-graining process.
tolsp	<tol>	For the coarse-graining process, a tolerance in (Å) to indicate how far from the molecule surface the coarse grain surface is allowed to extend.

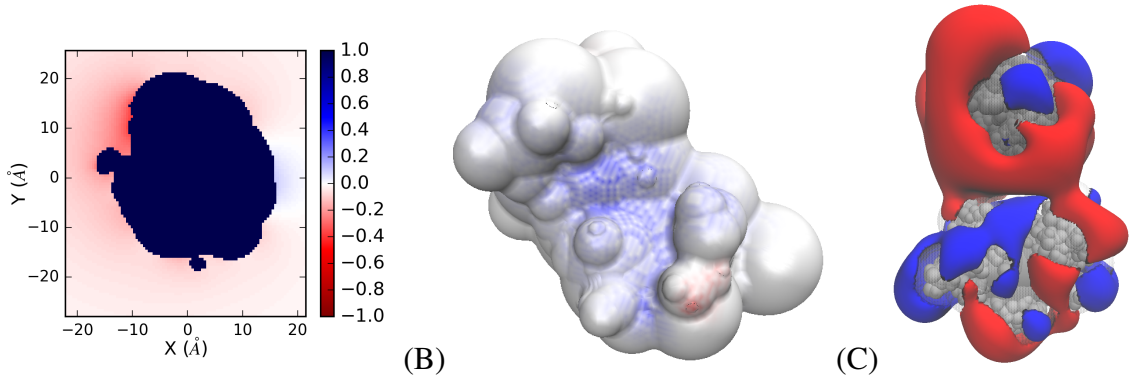
The periplasmic face (channel exit) has a positive surface (blue) at the channel exit which may further enhance anion transport through the channel.

ESP visualization results are given in Figure 3 for the PB-SAM model. They include the visualizations of barnase and barstar molecules, a ribonuclease and its inhibitor, whose association is driven by electrostatic interactions. Figure 3 (A) is a two-dimensional cross section visualization of a barnase molecule. The molecule interior is given in dark blue and the color bar indicates the spatial variation of the potential with the molecule centered at the origin. Figure 3 (B) is a visualization of the potential of a barstar molecule at its association interface. The positive potential distribution at the surface interacts favorably with the negative potential distribution on the barnase surface. Figure 3 (C) visualizes the electrostatic potential surrounding the barnase-barstar interaction complex.



(A) (B) (C)

Figure 2: ESP results derived from APBS PB-AM (0.0M salt, 7 pH, dielectric 2 (protein), and 78 (solution)): (A) 2D cross section of two collections of point charges (each with net charge ±6), (B) potential on the spherical coarse-grain surface of a barstar molecule (k_BT/e), (C) VMD visualization of the transmembrane trimer Porin (molecules are given in grey and isosurfaces are drawn at 1.0 (blue) and -1.0 (red) k_BT/e).



(A) (B) (C)

Figure 3: ESP results derived from APBS PB-SAM (0.0M salt, 7 pH, dielectric 2 (protein), and 78 (solution)): (A) potential in a 2D plane surrounding a barstar molecule (k_BT/e), (B) potential on the coarse-grain surface of the barnase molecule (blue region is the location of barstar association, Max = 1.0 k_BT/e, min = -1.0 k_BT/e), (C) VMD visualization of the ESP around the association of barnase and barstar (molecules are given in grey and isosurfaces are drawn at 1.0 (blue) and -1.0 (red) k_BT/e).

2.5.2 Boundary Element Method

In order to address some of the issues with the numerical methods in APBS, a treecode accelerated boundary integral PB solver (TABI-PB) developed by Geng and Krasny[23] has recently been implement. This method provides a boundary element approach to the linearized PB equation. In this method, two coupled integral equations defined on the solute-solvent boundary define a mathematical relationship between the electrostatic surface potential ϕ_1 and its normal derivative with a set of integral kernels consisting of Coulomb and screened Coulomb potentials with their normal derivatives. The method requires a surface triangulation, generated by a program such as MSMS[30] or NanoShaper,[31] on which to discretize the integral equations. The resulting linear system is then solved with GMRES iteration.[32] To reduce the computational cost of this dense system from

$O(N^2)$ to $O(N\log N)$, a Cartesian particle-cluster treecode is employed to compute matrix-vector products.

The matrix-vector products involve evaluation of the integral kernels over the surface elements. These evaluations effectively take the form of an N-body potential: a sum over a set of N positions of products between a kernel and a "charge" at each position. In this case, the locations of the N particles are the centroids of the surface triangularization elements. To rapidly evaluate the N-body potential at the N particle locations, the treecode subdivides the particles into a tree-like hierarchical structure of clusters. At each location, the potential contribution from nearby particles is computed by direct sum, while for far particle clusters, a Taylor approximation about the center of the cluster is used to evaluate the contribution. The Taylor coefficients are calculated through recurrence relations.

The TABI-PB method utilizes a well-posed boundary integral PB formulation to ensure rapid convergence. In addition, a fast treecode algorithm for the screened Coulomb potential[33] is applied to speed up the matrix-vector products in each GMRES iteration.[34] The TABI-PB solver serves as an effective alternative to the multi-grid finite difference solver when the surface potential accuracy or system memory capacity becomes a concern.

The coupled second kind integral equations employed by TABI-PB for calculating the surface potential ϕ_1 and its normal derivative,[34] are as follows:

$$\begin{aligned}
\frac{1}{2}\left(1+\varepsilon\right)\phi_1\left(\mathbf{x}\right) &= \int_\Gamma \left[K_1\left(\mathbf{x},\mathbf{y}\right)\frac{\partial\phi_1\left(\mathbf{y}\right)}{\partial\nu} + K_2\left(\mathbf{x},\mathbf{y}\right)\phi_1\left(\mathbf{y}\right)\right] dS_\mathbf{y} + S_1(\mathbf{x}),\ \mathbf{x}\in\Gamma, \\
\frac{1}{2}\left(1+\frac{1}{\varepsilon}\right)\frac{\partial\phi_1\left(\mathbf{x}\right)}{\partial\nu} &= \int_\Gamma \left[K_3\left(\mathbf{x},\mathbf{y}\right)\frac{\partial\phi_1\left(\mathbf{y}\right)}{\partial\nu} + K_4\left(\mathbf{x},\mathbf{y}\right)\phi_1\left(\mathbf{y}\right)\right] dS_\mathbf{y} + S_2(\mathbf{x}),\ \mathbf{x}\in\Gamma
\end{aligned} \tag{22}$$

where $\varepsilon = \varepsilon_m/\varepsilon_s$, the ratio of the dielectric constant in the solute region and the dielectric constant in the solvent region. The integral kernels K_1, K_2, K_3, K_4 are defined in Eq. 23.

$$\begin{aligned}
K_1\left(\mathbf{x},\mathbf{y}\right) &= G_0\left(\mathbf{x},\mathbf{y}\right) - G_\kappa\left(\mathbf{x},\mathbf{y}\right),\ K_2\left(\mathbf{x},\mathbf{y}\right) = \varepsilon\frac{\partial G_\kappa\left(\mathbf{x},\mathbf{y}\right)}{\partial\nu_\mathbf{y}} - \frac{\partial G_0\left(\mathbf{x},\mathbf{y}\right)}{\partial\nu_\mathbf{y}} \\
K_3\left(\mathbf{x},\mathbf{y}\right) &= \varepsilon\frac{\partial G_0\left(\mathbf{x},\mathbf{y}\right)}{\partial\nu_\mathbf{x}} - \frac{1}{\varepsilon}\frac{\partial G_\kappa\left(\mathbf{x},\mathbf{y}\right)}{\partial\nu_\mathbf{x}},\ K_4\left(\mathbf{x},\mathbf{y}\right) = \varepsilon\frac{\partial^2 G_\kappa\left(\mathbf{x},\mathbf{y}\right)}{\partial\nu_\mathbf{x}\partial\nu_\mathbf{y}} - \frac{1}{\varepsilon}\frac{\partial^2 G_0\left(\mathbf{x},\mathbf{y}\right)}{\partial\nu_\mathbf{x}\partial\nu_\mathbf{x}}
\end{aligned} \tag{23}$$

where G_0 and G_κ are the Coulomb and screened Coulomb potentials defined as:

$$G_0(\mathbf{x},\mathbf{y}) = \frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}, \;\; G_\kappa(\mathbf{x},\mathbf{y}) = \frac{e^{-\kappa|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} \tag{24}$$

The normal derivatives of the potential kernels G are:

$$\begin{aligned}
\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial \nu_\mathbf{y}} &= \sum_{n=1}^{3} \nu_n(\mathbf{y})\, \partial_{y_n} G(\mathbf{x},\mathbf{y}), \\
\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial \nu_\mathbf{x}} &= -\sum_{n=1}^{3} \nu_n(\mathbf{y})\, \partial_{x_n} G(\mathbf{x},\mathbf{y}), \\
\frac{\partial^2 G(\mathbf{x},\mathbf{y})}{\partial \nu_\mathbf{x} \partial \nu_\mathbf{y}} &= -\sum_{m=1}^{3}\sum_{n=1}^{3} \nu_m(\mathbf{x})\, \nu_n(\mathbf{y})\, \partial_{x_n}\partial_{y_n} G(\mathbf{x},\mathbf{y})
\end{aligned} \tag{25}$$

for the three spatial components n of the normal direction. Additionally, the source terms S_1 and S_2 in Eq. 22 are:

$$S_1(\mathbf{x}) = \frac{1}{\varepsilon_m}\sum_{k=1}^{N_c} q_k G_0(\mathbf{x},\mathbf{y}_k), \;\; S_2(\mathbf{x}) = \frac{1}{\varepsilon_m}\sum_{k=1}^{N_c} q_k \frac{\partial G_0(\mathbf{x},\mathbf{y}_k)}{\partial \nu_\mathbf{x}} \tag{26}$$

where N_c is the number of atoms in the solute molecule, and q_k is the charge of the kth atom. Note that S_1 is a linear superposition of the point charge electrostatic potentials, and S_2 is a linear superposition of the normal derivatives of the potentials.

Given a surface triangularization with N elements, where $\mathbf{x}_i$ and A_i are the centroid and area, respectively, of the ith triangle, the integral equations are discretized as:

$$\begin{aligned}
\frac{1}{2}(1+\varepsilon)\,\phi_1(\mathbf{x}_i) &= \sum_{\substack{j=1 \\ j\neq i}}^{N} \left[K_1(\mathbf{x}_i,\mathbf{x}_j) \frac{\partial \phi_1(\mathbf{x}_j)}{\partial \nu} + K_2(\mathbf{x}_i,\mathbf{x}_j)\,\phi_1(\mathbf{x}_j) \right] A_j + S_1(\mathbf{x}_i), \\
\frac{1}{2}\left(1+\frac{1}{\varepsilon}\right)\frac{\partial \phi_1(\mathbf{x}_i)}{\partial \nu} &= \sum_{\substack{j=1 \\ j\neq i}}^{N} \left[K_3(\mathbf{x}_i,\mathbf{x}_j) \frac{\partial \phi_1(\mathbf{x}_j)}{\partial \nu} + K_4(\mathbf{x}_i,\mathbf{x}_j)\,\phi_1(\mathbf{x}_j) \right] A_J + S_2(\mathbf{x}_I)
\end{aligned} \tag{27}$$

The omission of the $j = i$ term in the summation avoids the singularity of the kernels at that point. Note that the right hand sides of these equations consist of sums of products of kernels and the

surface potential or its normal derivative. These are the analogues to the N-body potential in the treecode, with the surface potential or its normal derivatives playing the role of the charges. In the discretized form, the total electrostatic energy of solvation is given by Eq. 28.

$$E_{\text{sol}} = \frac{1}{2} \sum_{k=1}^{N_c} q_k \sum_{j=1}^{N} \left[K_1 \left(\mathbf{x}_k, \mathbf{x}_j \right) \frac{\partial \phi_1 \left(\mathbf{x}_j \right)}{\partial \nu} + K_2 \left(\mathbf{x}_k, \mathbf{x}_j \right) \phi_1 \left(\mathbf{x}_j \right) \right] A_j \tag{28}$$

where q_k is the charge on the kth atom of the solute molecule, and $\mathbf{x}_k$ is its position.

Several unique features of the boundary integral formulation are worth mentioning. Because the integral equations are defined on the molecular boundary, the singular charges are handled analytically and do not introduce the same issues present in grid-based schemes. The integral equations also rigorously enforce the interface conditions on the surface, and the boundary condition at infinity is exactly satisfied. Thus, the boundary integral formulation can potentially be superior to other methods for investigating electrostatic potential on the boundary. However, because the method is only well defined for the linearized PB equation, it may perform poorly for high salt concentration in the solvent.

Usage in APBS The APBS user can invoke TABI-PB with the `bem-manual` flag in the ELEC section of the input file. TABI-PB will produce output including the potential and normal derivative of potential for every element and vertex of the triangularization, as well as the electrostatic solvation energy. The user can additionally specify the order of the Taylor expansions used in the far field potential contribution approximations, the maximum number of particles allowable in a leaf (clusters in the last level of the tree), and the multipole acceptance criterion, which determines the distance at which potential contributions from particles in a cluster are evaluated directly or through the Taylor expansion (see Table 2).

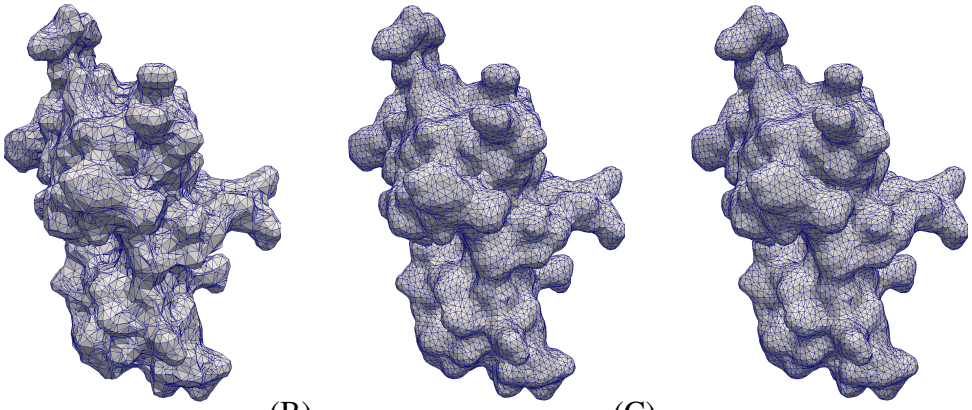
In general, a lower multipole acceptance criterion (θ) will result in a more accurate but more expensive computation; a lower θ causes more direct summations and forces the particle-cluster interaction to descend to a finer cluster level. Similarly, a higher Taylor expansion order will also result in a more accurate but more expensive computation. A typical first-pass choice for these

parameters would be a multipole acceptance criterion of 0.8, a Taylor expansion order of 3, and a maximum leaf size of 500. More investigation is needed to determine optimal parameter choices.

Table 2: Keywords used in TABI-PB. For more details and examples of implementation, please see the APBS website.

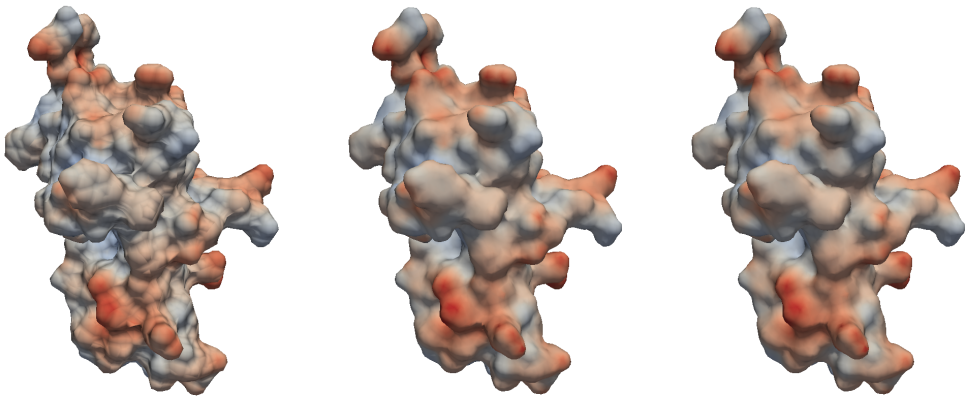
Keyword	Parameters	Description
tree_order	`<order>`	Integer indicating the order of the Taylor expansion for determining treecode coefficients.
tree_n0	`<max_number>`	Maximum number of particles allowable in a treecode leaf.
mac	`<theta>`	Multipole acceptance criterion, specifies the distance ratio at which the Taylor expansion is used.
mesh	`<flag>`	Software to be used for meshing the molecular surface: 0=MSMS, 1=NanoShaper's SES implementation, and 2=NanoShaper's Skin implementation.
outdata	`<flag>`	Type of output data file generated: 0=APBS format and 1=compatible with ParaView (including VTK data file).

The user can also choose to specify whether TABI-PB uses MSMS, the SES implementation in NanoShaper, or the Skin surface implementation of NanoShaper to generate a surface triangularization, as shown in Figure 4. NanoShaper, an open-source program for surface triangularization, was recently added to TABI-PB, and preliminary results suggest superior performance to MSMS. Examples of surface potential on 1a63 are given in Figure 5 by using MSMS and NanoShaper.



(A) (B) (C)

Figure 4: PDB ID 1a63 surfaces generated with (A) MSMS, number of triangles=20228; (B) NanoShaper SES, number of triangles=20744; (C) NanoShaper Skin, number of triangles=21084.



(A) (B) (C)

Figure 5: 1a63 visualization of surface potential for (A) MSMS, max=5.06 kcal/mol, min=-5.15 kcal/mol; (B) NanoShaper SES, max=4.44 kcal/mol, min=-7.95 kcal/mol; (C) NanoShaper Skin, max=4.25 kcal/mol, min=-20.03 kcal/mol.

2.5.3 Geometric Flow

To increase the accuracy of our implicit solvent modeling, we have recently implemented a differential geometry based geometric flow solvation model. In this model, polar and nonpolar free energies are coupled through a characteristic function. This function describes a smooth dielectric interface profile across the solvent-solute boundary.[35]

For our modeling, a generalized form of the Poisson equation for computing the electrostatic is used as shown in Eq. 29, where ε_m and ε_s are the dielectric coefficients of the solute and solvent, respectively, and ρ_m is the charge distribution of the fixed solute molecule. For this model, the solutions for the electrostatic potential (ϕ) and the characteristic function (S) are obtained by minimizing the free energy functional. The dielectric function $\varepsilon(S)$ takes on the value ε_m in the solute region (S=1) and the value ε_s in the solvent region (S=0).

$$-\nabla \cdot (\varepsilon(S)\nabla\phi) = S\rho_m \tag{29}$$

For this model, we have utilized an Eulerian formulation of the geometric flow problem. This means that S varies smoothly across the solute-solvent interfaces and leads to a non-linear partial differential equation for the characteristic function, S, as shown in Eq. 30.

$$-\nabla\cdot\left(\gamma\frac{\nabla S}{\parallel\nabla S\parallel}\right)+p-\rho_0U^{att}+\rho_m\phi-\frac{1}{2}\varepsilon_m\mid\nabla\phi\mid^2+\frac{1}{2}\varepsilon_s\mid\nabla\phi\mid^2=0 \tag{30}$$

where γ is the microscopic surface tension, p is the hydrodynamic pressure, and U^{att} is the attractive potential of the van der Waals dispersion interaction between the solute and the solvent.

The solution to this non-linear PDE can be solved by utilizing the parabolic PDE shown in Eq. 31. This equation is known as the generalized geometric flow equation and it is coupled with the Poisson equation through the characteristic function S.[36]

$$\frac{\partial S}{\partial t}=\gamma\parallel\nabla S\parallel\left[\nabla\cdot\left(\frac{\nabla S}{\parallel\nabla S\parallel}\right)+\frac{V}{\gamma}\right], \tag{31}$$

where V is known as the generalized flow potential as described in Eq. 32.

$$V=-p+\rho_0U^{att}-\rho_m\phi+\frac{\varepsilon_m}{2}\mid\nabla\phi\mid^2-\frac{\varepsilon_s}{2}\mid\nabla\phi\mid^2. \tag{32}$$

These equations were solved using a second-order central finite difference scheme, which utilizes a bi-conjugate gradient stabilized solver.

The sensitivity and accuracy of an implicit solvent method typically depends on the values selected for the force-field parameters. However, the geometric flow method is only dependent on the parameters of the implicit solvent model (surface tension, dielectric coefficient, solvent pressure, etc.).

In a sensitivity analysis performed by Thomas et. al.,[35] changing the distance between the axis boundary and the surface of each molecule and the grid spacing for the solver showed little effect on the results (3-5%). However, changes in the internal dielectric (ε_m) resulted in much larger sensitivity (i.e., changes in the results).

Usage in APBS The differential geometry-based geometric flow solvation model is used to describe a smooth dielectric interface profile across the solvent-solute boundary in a thermodynamically self-consistent fashion. The main parameters of the model are the solute/solvent dielec-

tric coefficients, solvent pressure on the solute, microscopic surface tension, solvent density, and molecular force-field parameters.

This model is accessed under the ELEC section of the input file. Results using the OPLS-AA force-field calculations, with and without van der Waals dispersion (vdW), are shown in Figure 6 and were discussed by Thomas and friends.[35]

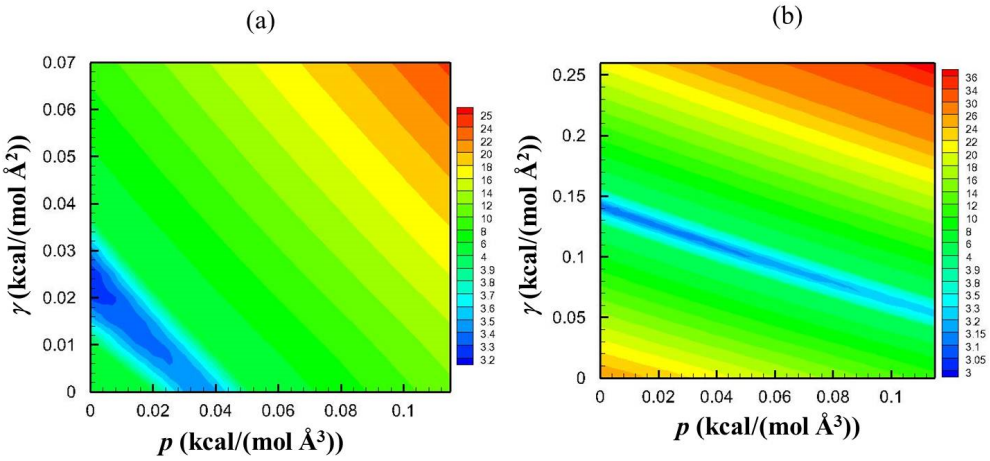


Figure 6: Contour plot of the error between experimental and calculated solvation free energies based on OPLS-AA calculations: (a) without vdW; (b) with vdW.

2.6 Visualizing and using the results

APBS was designed to facilitate use with other programs. This section outlines some of the programs with which APBS is known to work with.

2.6.1 Thick-client visualization

Graphical User Interfaces

PyMOL - molecular visualization and animation package which provides an interface to APBS. The APBS plugin to PyMOL permits isocontour and surface map visualization of APBS results.[2]

VMD - molecular visualization and animation package which provides an interface to APBS. It permits visualization of APBS results as isocontours, electric field lines, or on biomolecular

[2] www.pymol.org

surfaces. VMD also has graphical plugin to setup APBS calculations and execute them either locally or remotely via BioCoRE.[29]

PMV - Python-based molecular visualization package which provides an interface to APBS. It not only permits visualization of APBS results but it also integrates setup and executation of APBS calculations.[3]

Chimera - molecular visualization package which provides an interface to APBS.The APBS plugin integrated in Chimera allows the user to run a protein through both PDB2PQR and APBS to optimize the protein and view the electrostatic potential.[37]

Visualization Software

Electrostatic potentials are commonly visualized in the context of biomolecular structure to better understand functional aspects of biological systems. Note that the graphical user interfaces discussed above can also be used to visualize APBS output. The following molecular graphics software can display potentials and other data output from APBS:

Off-Target Pipeline - implements APBS to estimate ligand binding energies and compare electrostatic potential distributions within binding cavities.[38]

Dino3D - molecular graphics program which can read UHBD-formatted electrostatic data. APBS can write multigrid results in UHBD format and therefore can be used with Dino3D.[4]

OpenDX - general data visualization package which can read APBS output using the scripts provided in tools/visualization/opendx. However, as there is no straightforward way to visualize the potential in the context of the atomic structure, OpenDX should not be a first choice for APBS visualization.[5]

2.6.2 Web-based visualization

Computational chemists and biologists rely on molecular visualization tools to view model proteins and predict unobserved protein simulations.[39] These tools allow scientists to efficiently visualize

[3] www.mgltools.scripps.edu/packages/pmv
[4] www.dino3d.org
[5] www.opendx.org

and explore proteins. Thus allowing for a better understanding of relationships and functions of proteins.

Visualization tools for evaluating simulation results are critical for the discovery of new insights. Molecular viewers are available as stand-alone desktop programs or web-based applications. Desktop users are often burdened with obtaining software licenses or downloads when using desktop viewers such as VMD and web-based users are troubled with java security settings, permissions, or browser plug-ins when using web-based viewers such as Jmol.[40] While most viewers offer the same features such as rendering styles and coloring methods, lack of accessibility and poor performance can become a hindrance for scientists as they conduct research.

To help with these issues, we have integrated 3Dmol[41] into the PDB2PQR software and APBS pipeline. 3Dmol is a molecular viewer that offers the performance of a desktop application and convenience of a web-based viewer which broadens accessibility for all users. In order to integrate 3Dmol, enhancements were need for our tools, including extending our output file formats and creating a customized user interface. In the end, we have enabled a new light weight capability for visualization of complex electrostatic properties.

To allow users web-based access to PDB2PQR and APBS, we utilize a web server interface hosted at the National Biomedical Computation Resource (NBCR)[6] Center which uses the Opal2 toolkit to interface with the NBCR cluster. Opal2 handles the scheduling and execution of jobs and stores the results for later retrieval.[42] The NBCR server is designed to facilitate the setup and execution of continuum electrostatics calculation from PDB data, particularly by non-experts. The web service is driven by a modular Python-based collection of routines, which provides considerable flexibility to the software and permits non-interactive, high-throughput usage.

We use the new capabilities in 3Dmol to render common molecular data representations for visualization and analysis. Data from the APBS output file is used to generate surfaces, apply color schemes, and display different molecular styles such as cartoons and spheres. An example of the 3Dmol interface is shown in Figure 7. The graphical user interface (GUI) has three surface

[6] www.nbrc.ucsd.edu

options; solvent accessible, solvent excluded, and van der Waals.

Using the GUI, the user can show or hide the surface or display it as translucent or opaque. The user can apply a minimum and maximum surface potential value. The default minimum and maximum isovalues are -5 kT/e and 5 kT/e, respectfully. The user can select a line, stick, cross, sphere, or cartoon model and apply a red-white-blue or red-green-blue color scheme to the surface. Additionally, 3Dmol renders proteins using different styles and color schemes, detailed in Table 3. Figure 8 shows the visual representation of some selected features.

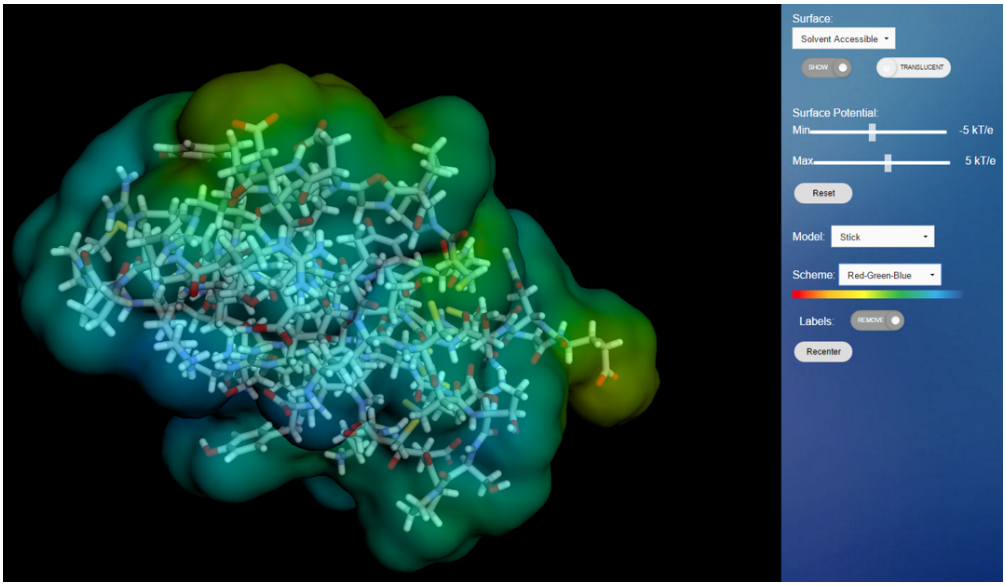


Figure 7: 3Dmol interface displaying a rendering of FASCICULIN (1FAS) protein with translucent, solvent accessible surface using a stick model and red-green-blue color scheme.

Table 3: Selected 3Dmol features.

Feature	Description
Solvent accessible surface	describes the surface area of a biomolecule that is accessible to a solvent
Red-White-Blue	describes the color scheme, red to white to blue, for charges
Stick model	draws bonds as capped cylinders
Cross model	draws atoms as crossed lines (i.e., stars)
Sphere model	draws atoms as spheres

3Dmol is an effective, interactive molecular viewer that is accessible across web browsers and operating systems. With 3Dmol, users can make selections using a simple GUI such as select

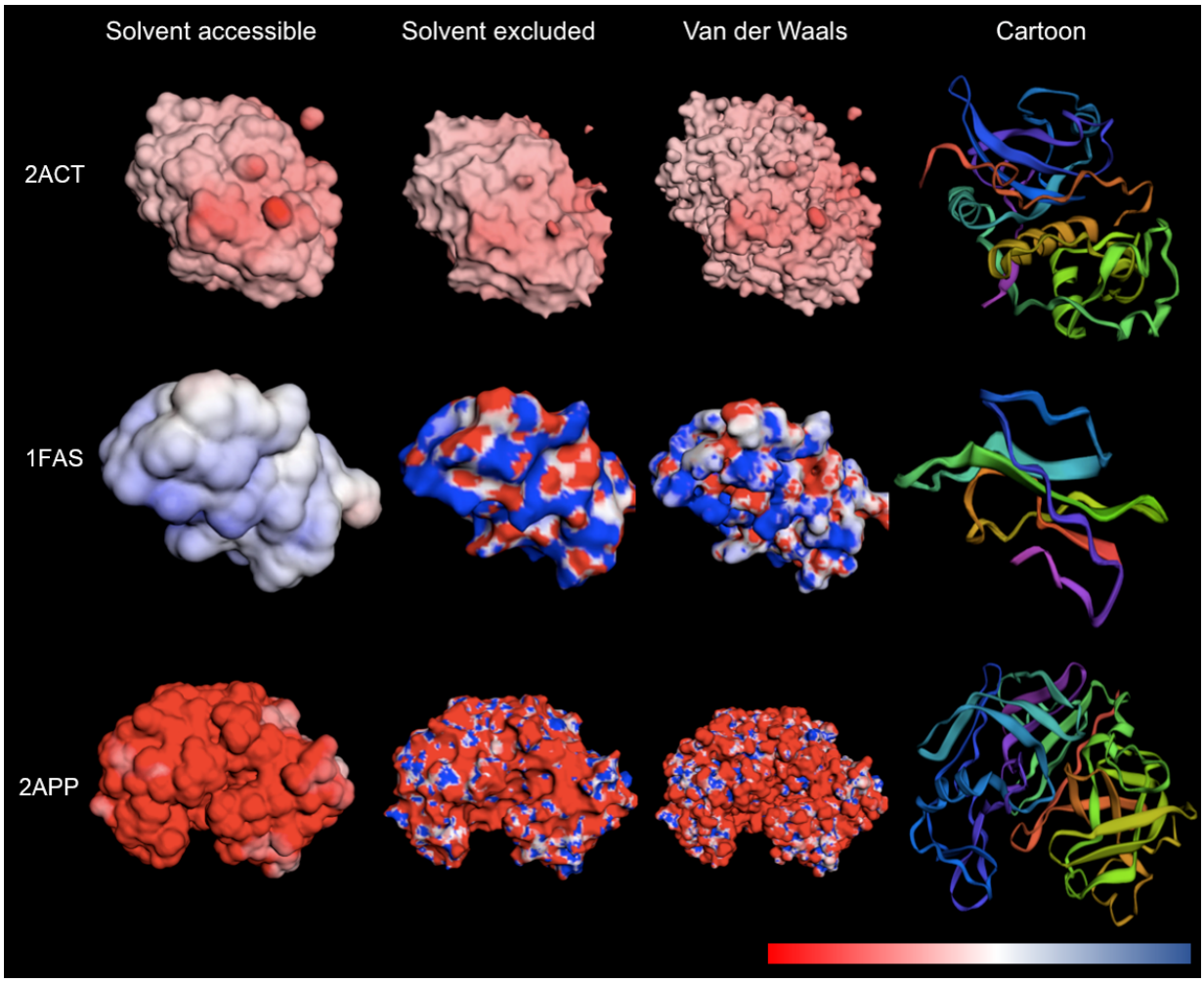


Figure 8: Renderings of three different proteins using renderings of Actinidin (2ACT) (top), Fasciculin (1FAS) (center), and Pepsin, Penicillium (2APP) (bottom). To demonstrate the different visualization options. From left to right, solvent accessible surface, solvent excluded surface, van der Waals surface, and cartoon model are shown all using red-white-blue color scheme (excluding cartoon model).

different surfaces, apply various models, and choose from different color schemes. The features available in 3Dmol are comparable to Jmol[7], a molecular visualization tool written in Java; however, there are also differences between the two tools. While Jmol offers trace, backbone, cartoon, and ball and stick models, 3Dmol offers cartoon, cross, line, sphere, and stick models. Unlike Jmol, 3Dmol loads quickly and does not require any downloads or plug-ins for use.

2.6.3 Other data uses

There are a number of other applications of implicit solvent models to biomedical problems. For example, during the past four years, our APBS software has been used in the post-simulation energetic analyses of molecular dynamics trajectories,[43] understanding protein-nanoparticle interactions,[44–46] understanding nucleic acid-ion interactions,[47,48] biomolecular docking,[49] developing new coarse-grained protein models,[50] setting up membrane protein simulations,[51] etc. APBS also plays a key role in software packages from the Wade group, including PIPSA for protein surface electrostatics analysis[52] and SDA for simulation of protein-protein interactions through Brownian dynamics.[53] Another application area for implicit solvent methods is in the evaluation of biomolecular kinetics where implicit solvent models are generally used to provide solvation forces (or energies) for performing (or analyzing) discrete molecular or continuum diffusion simulations with APBS in both of these areas.[43,53–58]

3 Using the software

APBS was designed as a standalone system and also to facilitate use with other programs. This section outlines execution methods and some of the programs with which APBS is known to work with. All of these methods are described in depth on the APBS-PDB2PQR website.[8]

[7] www.jmol.sourceforge.net
[8] www.poissonboltzmann.org

3.1 Command-line execution

Most users will likely interact with APBS and PDB2PQR through web servers. However, it is also possible to install local versions. These local installations give a command line version of the software that can be customized through a variety of extensions. Compiled from source, a local version can also provide a web server.

3.2 Execution through other programs

APBS can also be executed from other programs. As an example, the following programs have had capabilities to run APBS:

- PDB2PQR
- iAPBS - Robert Konecny (McCammon Group) has developed iAPBS, an interface between APBS and the simulation packages AMBER, CHARMM, and NAMD. The iAPBS package provides a C/C++ and Fortran interface to APBS through a single function call. The interface code can be compiled as a library (libiapbs.a) which can be linked with a Fortran or C/C++ application thus making most of APBS functionality available from within any C/C++/Fortran code. [9]
- TINKER - APBS 1.3 is available with TINKER; TINKER is a molecular modeling software package for molecular mechanics and dynamics, and includes some special features for bioploymers.[10]

3.3 Web-based execution

As discussed earlier, web-based executation of APBS is accomplished using the Opal Toolkit produced by the NBCR. This toolkit allows for the computing load for processor intensive scientific applications to be shifted to a 3rd party and/or generic computing grid. This can be tremendously

[9] www.mccammon.ucsd.edu/iapbs
[10] www.dasherwustl.edu/tinker

advantageous in situations where a large amount of computing power is not locally available, but is required, for the task at hand. In particular, many users have discovered that their local computational resources are insufficient for certain types of APBS calculations on large systems or at extremely high accuracy. This client removes this resource limitation by allowing users to run on clusters at NBCR. Recent developmental versions of APBS add optional support for the offloading of APBS calculations to an Opal service. The APBS Opal support is in the form of a Python script `ApbsClient.py` and is installed by default when following the installation procedure outlined elsewhere. The script has been tested on Python 2.5; newer/older versions of Python may or may be functional. As mentioned above, the basic invocation is the same as the main binary. The only difference is the executable, which is called `ApbsClient.py`, rather than `apbs`. This client should be installed by default when APBS is installed.

3.4 Extending APBS capabilities

There are several challenges when working with large code bases. Careful planning must go into the design, especially if the program is to be extended. Ideally, one would like the ability to link new code, which extends functionality, without having to change the existing base code. However, this is rarely the case. In the best scenario, a few line changes may be required. On the other hand, if proper care has not been taken, large modifications may be unavoidable.

To help with modularity, several programming paradigms have been introduced over the years. Object Oriented Programming (OOP), the paradigm used in the design of APBS, focuses on principal objects. Classes are used to define objects with given characteristics and functionality. This gives rise to a natural hierarchical structure where subclasses share any number of characteristics with the main class. APBS works with atoms, that form residues, that form molecules, etc, which naturally lends itself to the OOP paradigm. The C programming language is known for fast performance but does not natively supports OOP. To overcome this drawback, APBS was written using a subset of ANSI C called Clean OO C style by Mike Holst.[59]

The main APBS file takes inputs from the command line and initializes some of the variables

necessary for I/O and passing data to the Routines manager. The Routines manager, in turn, creates atom objects and invokes whichever solver is to be used. The advantages of using Clean OO C becomes clear when adding new solver methods to APBS. Because of APBS' OOP style, all one must do is make Routines aware of the new method. As long as the new methods adhere to Clean OO C calling conventions, the passing and return of data is seamless and most other methods (e.g. reading and writing to DX files) can be used. New methods which are external or do not adhere to Clean OO C standards can be managed by writing auxiliary methods for translating data from one to the other. These auxiliary methods are given in a header and source file. The external methods are linked at compilation time.

4 Future directions

The focus of APBS and its accompanying software is the modeling of solvation and electrostatic properties of macromolecules in biomedical research. As such, continued research and development of APBS is focused on modeling different possible application scenarios that may arise in the modeling of the solvation and electrostatic properties.

APBS contains many different models and methods, which gives it a unique capability as a software package to be utilized in many different research applications. Methods that have been examined for future implementation into APBS and PDB2PQR include: advanced methods (gPC) for quantifying the influence of uncertainty on simple solvation-related properties (such as solvent-accessible surface area), ensemble averaging methods for titration state predictions, and virtual reality visualization of electrostatic potentials and related properties.

Our vision for APBS is to build the infrastructure that can enable our users to implement their own models and methods so that they can run on a common system. Imagine all the duplication of effort in building UIs, file loaders, parsers, output writers, etc. All that could be eliminated.

Our goal is to have a well designed, well tested and well documented industry grade framework that will make it possible for all of these various bimolecular software packages to work together.

To leverage each other's capabilities, to exceed the sum of the individual software package capabilities.

5 Acknowledgments

The authors gratefully acknowledge NIH grant GM069702 for support of this research. PNNL is operated by Battelle for the U.S. DOE under contract DE-AC05-76RL01830.

References

(1) Baker, N.; Sept, D.; Joseph, S.; Holst, M.; McCammon, J. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.

(2) Davis, M. E.; McCammon, J. A. Electrostatics in biomolecular structure and dynamics. *Chemical Reviews* **1990**, *90*, 509–521.

(3) Perutz, M. F. Electrostatic effects in proteins. *Science* **1978**, *201*, 1187–91.

(4) Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. Biomolecular electrostatics and solvation: a computational perspective. *Q Rev Biophys* **2012**, *45*, 427–91.

(5) Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules: theory and applications. *Annu Rev Biophys Biophys Chem* **1990**, *19*, 301–32.

(6) Roux, B.; Simonson, T. Implicit solvent models. *Biophys Chem* **1999**, *78*, 1–20.

(7) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochim Biophys Acta* **2006**, *1764*, 1647–76.

(8) Fixman, M. The Poisson-Boltzmann equation and its application to polyelectrolytes. *J. Chem. Phys.* **1979**, *70*, 4995–5005.

(9) Grochowski, P.; Trylska, J. Continuum molecular electrostatics, salt effects, and counterion binding–a review of the Poisson-Boltzmann theory and its modifications. *Biopolymers* **2008**, *89*, 93–113.

(10) Lamm, G.; Lipkowitz, K. B.; Larter, R.; Cundari, T. R.; Boyd, D. B. The Poisson-Boltzmann Equation. *Rev. Comput. Chem.* **2003**, *19*, 147–365.

(11) Netz, R. R.; Orland, H. Beyond Poisson-Boltzmann: Fluctuation effects and correlation functions. *Eur. Phys. J. E* **2000**, *1*, 203–214.

(12) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

(13) Dolinsky, T. J.; Czodrowski, P.; Hui, L.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525.

(14) Brooks, B.; Brooks, A., CL adn Mackerell; Nilsson, L.; Petrella, R.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A. et al. CHARMM: The biomolecular simulation program. *J. Comp. Chem.* **2009**, *30*, 1545–1614.

(15) Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comp. Chem.* **2005**, *26*, 1668–1688.

(16) Sarkar, S.; Witham, S.; Zhang, J.; Zhenirovskyy, M.; Rocchia, W.; Alexov, E. DelPhi web server: A comprehensive online suite for electrostatic calculations of biological macromolecules and their complexes. *Comm. Comp. Phys.* **2013**, *13*, 269–284.

(17) Bochevarov, A.; Harder, E.; Hughes, T.; Greenwood, J.; Braden, D.; Philipp, D.; Rinaldo, D.; Halls, M.; Zhang, J.; Friesner, R. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *International J. of Quantum Chem.* **2013**, *113*, 2110–2142.

(18) Grant, J.; Pickup, B.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comp. Chem.* **2001**, *22*, 608–640.

(19) Chen, D.; Chen, Z.; Chen, C.; Geng, W.; Wei, G. MIBPB: A software package for electrostatic analysis. *J. Comp. Chem.* **2011**, *32*, 756–770.

(20) Sondergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comp.* **2011**, *7*, 2284–2295.

(21) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **2005**, *61*, 704Ű721.

(22) Purvine, E.; Monson, K.; Jurrus, E.; Starr, K.; Baker, N. A. Energy Minimization of Discrete Protein Titration State Models Using Graph Theory. *J. Phys. Chem.* **2016**, *120*, 8354–8360.

(23) Geng, W.; Krasny, R. A treecode-accelerated boundary integral Poisson-Boltzmann solver for electrostatics of solvated biomolecules. *J. Comp. Phys.* **2013**, *247*, 62–78.

(24) Lotan, I.; Head-Gordon, T. An Analytical Electrostatic Model for Salt Screened Interactions between Multiple Proteins. *J. Chem. Theory and Comp.* **2006**, *2*, 541–555, PMID: 26626662.

(25) Yap, E. H.; Head-Gordon, T. A New and Efficient Poisson-Boltzmann Solver for Interaction of Multiple Proteins. *J. Chem. Theory Comp.* **2010**, *6*, 2214–2224.

(26) Ermak, D.; McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352–1360.

(27) Yap, E. H.; Head-Gordon, T. Calculating the Bimolecular Rate of Protein-Protein Association with Interacting Crowders. *J. Chem. Theory Comp.* **2013**, *9*, 2481–9.

(28) Northrup, S.; Allison, S.; McCammon, J. Brownian dynamics simulation of diffusion influenced bimolecular reactions. *J. Chem. Phys.* **1984**, *80*, 1517–1524.

(29) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(30) Sanner, M.; Olson, A.; Spehner, J. C. Fast and robust computation of molecular surfaces. Proc 11th ACM Symp Comp Geom. 1995; pp C6–C7.

(31) Decherchi, S.; Rocchia, W. A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLoS ONE* **2013**, *8*, e59744.

(32) Saad, Y.; Schultz, M. GMRES: A generalized minimal residual algorithm for solving non-symmetric linear systems. *J. Sci. Statist. Comput.* **1986**, *7*, 856–869.

(33) Li, P.; Johnston, H.; Krasney, R. A Cartesian treecode for screened coulomb interactions. *J. Comp. Phys.* **2009**, *228*, 3858–3868.

(34) Juffer, A.; Botta, E.; van Keulen, B.; van der Ploeg, A.; Berendsen, H. The electric potential of a macromolecule in a solvent: A fundamental approach. *J. Comp. Phys.* **1991**, *97*, 144–171.

(35) Thomas, D. G.; Chun, J.; Chen, Z.; Wei, G.; Baker, N. A. Parameterization of a geometric flow implicit solvation model. *J. Comp. Chem.* **2013**, *34*, 687–695.

(36) Chen, Z.; Baker, N. A.; Wei, G. Differential geometry based solvation model I: Eulerian formulation. *J. Comp. Phys.* **2010**, *229*, 8231–8258.

(37) Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comp. Chem.* **2004**, *13*, 1605–1612.

(38) Evangelidis, T.; Bourne, P.; Xie, L.; Xie, L. An integrated workflow for proteome-wide off-target identification and polypharmacology drug design. International Conference on Bioinformatics and Biomedicine Workshops. 2009; pp 32–39.

(39) Jones, L. L.; Jordan, K. D.; Stillings, N. A. Molecular visualization in chemistry education: the role of multidisciplinary collaboration. *Chemistry Education Research and Practice* **2005**, *6*, 136–149.

(40) Herraez, A. Biomolecules in the computer: Jmol to the rescue. *Biochemistry and Molecular Biology Education* **2006**, *34*, 255–261.

(41) Rego, N.; Koes, D. 3Dmol. js: molecular visualization with WebGL. *Bioinformatics* **2015**, *31*, 1322–1324.

(42) Krishnan, S.; Clementi, L.; Ren, J.; Papadopoulos, P.; Li, W. Design and evaluation of opal2: A toolkit for scientific software as a service. Services-I, 2009 World Conference on. 2009; pp 709–716.

(43) Dror, R.; Green, H.; Valant, C.; Borhani, D.; Valcourt, J.; Pan, A.; Arlow, D.; Canals, M.; Lane, J.; Rahmani, R. et al. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **2013**, *503*, 295–299.

(44) Treuel, L.; Brandholt, S.; Maffre, P.; Wiegele, S.; Shang, L.; Nienhaus, G. Impact of Protein Modification on the Protein Corona on Nanoparticles and NanoparticleŰCell Interactions. *ACS Nano* **2013**, *8*, 503–513.

(45) Nienhaus, G.; Maffre, P.; Nienhaus, K. Studying the Protein Corona on Nanoparticles by FCS. *Methods in Enzymology* **2013**, 115–137.

(46) De Paoli, S.; Diduch, L.; Tegegn, T.; Orecna, M.; Strader, M.; Karnaukhova, E.; Bonevich, J.; Holada, K.; Simak, J. The effect of protein corona composition on the interaction of carbon nanotubes with human blood platelets. *Biomaterials* **2014**, *35*, 6182–6194.

(47) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding Nucleic AcidŰIon Interactions. *Annual Review of Biochemistry* **2014**, *83*, 813–841.

(48) Giambasu, G. M.; Luchko, T.; Herschlag, D.; York, D. M.; Case, D. A. Ion Counting from Explicit-Solvent Simulations and 3D-RISM. *Biophysical Journal* **2014**, *106*, 883–894.

(49) Roberts, V.; Thompson, E.; Pique, M.; Perez, M.; Ten Eyck, L. DOT2: Macromolecular docking with improved biophysical models. *J. Comp. Chem.* **2013**, *34*, 1743–1758.

(50) Spiga, E.; Alemani, D.; Degiacomi, M.; Cascella, M.; Peraro, M. Brownian dynamics simulation of diffusion influenced bimolecular reactions. *J. Chem. Therory and Comp.* **2013**, *9*, 3515–3526.

(51) Stansfeld, P. J.; Goose, J. E.; Caffrey, M.; Carpenter, E. P.; Parker, J. L.; Newstead, S.; Sansom, M. S. MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **2015**, *23*, 1350–1361.

(52) Richter, S.; Wenzel, A.; Stein, M.; Gabdoulline, R.; Wade, R. webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Research* **2008**, *36*, W276–W280.

(53) Martinez, M.; Bruce, N.; Romanowska, J.; Kokh, D.; Ozboyaci, M.; Yu, X.; Öztürk, M.; Richter, S.; Wade, R. SDA 7: A modular and parallel implementation of the simulation of diffusional association software. *J. Comp. Chem.* **2015**, *36*, 1631–1645.

(54) Cheng, Y.; Suen, J.; Zhang, D.; Bond, S.; Zhang, Y.; Song, Y.; Baker, N.; Bajaj, C.; Holst, M.; McCammon, J. Finite element analysis of the time-dependent Smoluchowski equation for acetylcholinesterase reaction rate calculations. *Biophysical Journal* **2007**, *92*, 3397–3406.

(55) Song, Y.; Zhang, Y.; Bajaj, C.; Baker, N. Continuum diffusion reaction rate calculations of wild-type and mutant mouse acetylcholinesterase: adaptive finite element analysis. *Biophysical Journal* **2004**, *87*, 1558–1566.

(56) Song, Y.; Zhang, Y.; Shen, T.; Bajaj, C.; McCammon, J.; Baker, N. Finite element solution of the steady-state Smoluchowski equation for rate constant calculations. *Biophysical Journal* **2004**, *86*, 2017–2029.

(57) Elcock, A. Molecular Simulations of Diffusion and Association in Multimacromolecular Systems. *Methods in Enzymology* **2004**, 166–198.

(58) Mereghetti, P.; Wade, R. Atomic Detail Brownian Dynamics Simulations of Concentrated Protein Solutions with a Mean Field Treatment of Hydrodynamic Interactions. *J. Phy. Chem.* **2012**, *116*, 8523–8533.

(59) Holst, M. Adaptive numerical treatment of elliptic systems on manifolds. *J. Comp. Math.* **2001**, *15*, 139–191.